**Cybersecurity Risk Stratification Framework Using Multilevel Clustering: An Automated Threat Attribution and Categorization Approach for Cross-Industry Cybersecurity**

***Abstract***

*This study introduces a novel Multilevel Clustering Framework designed for automated threat attribution and categorization across various industries using a comprehensive dataset from the MITRE ATT&CK repository. The methodology integrates K-means Clustering, Hierarchical Clustering, and Fuzzy C-means to address key limitations of traditional models, including inadequate adaptability, scalability, and robustness to noise. By employing a three-stage clustering process, the framework ensures improved detection accuracy, robustness against noise, and cross-industry applicability. The concept of Generalized Attack Patterns refers to commonly occurring attack vectors and techniques that transcend specific industries, allowing for a unified approach to threat detection. Unlike traditional clustering models that are constrained by sector-specific characteristics, the proposed framework effectively identifies and categorizes both industry-specific and generalized threats with high accuracy. Quantitative evaluation across healthcare, finance, telecommunications, manufacturing, and critical infrastructure demonstrates the framework’s effectiveness, achieving a Classification Accuracy of 0.90, Robustness to Noise of 0.83, Adaptability Index of 0.87, and Cross-Industry Applicability of 0.85. However, the Telecommunications sector showed comparatively lower performance, with a Jaccard Index of 0.74, indicating challenges in clustering highly dynamic datasets. Recommendations include implementing customized pre-processing techniques for telecommunications, incorporating hybrid models in finance, refining algorithms for critical infrastructure, and integrating real-time data for cross-industry applications.*

**Keywords: Multilevel Clustering, MITRE ATT&CK, Threat Attribution, Cross-Industry Applicability, Cybersecurity Framework**

**1. Introduction**

The increasing complexity and frequency of cyberattacks across various industries present substantial challenges in threat attribution and categorization. The proliferation of digital systems has resulted in a considerable escalation of cyber threats, with recent statistics underscoring the severity of this phenomenon. According to Alder (2024), 30,458 security incidents were analyzed, confirming 10,626 data breaches, a figure that has doubled from the previous year. This surge demonstrates the continuous advancement of cyber threats, which have grown increasingly sophisticated and are now capable of breaching systems across various sectors.

The financial ramifications of these attacks are equally alarming. According to Jacobson (2023), the average global cost of a data breach reached a record high of $4.45 million, marking a 2.3% increase from the prior year. The United States experienced the highest average cost at $9.48 million, with the healthcare sector suffering the most significant financial impact, where average breach costs exceeded $10.93 million (Petrosyan, 2024). These figures underscore the urgent need for advanced frameworks that can accurately attribute and categorize cyber threats across domains such as healthcare, finance, manufacturing, telecommunications, and critical infrastructure.

Moreover, cybercriminals' tactics continue to evolve, transcending sector-specific boundaries. For instance, the resurgence of Emotet in 2022 illustrates this adaptability, as the malware exploited vulnerabilities across various industries (Chua, 2022). According to Chua (2022), Emotet exhibited versatility by infiltrating diverse client environments through tailored strategies. This example highlights the limitations of traditional clustering techniques primarily designed for sector-specific applications, thereby necessitating more comprehensive methodologies capable of detecting patterns across multiple domains.

Current clustering techniques, including K-means, Hierarchical Clustering, and Fuzzy C-Means, have demonstrated success within isolated domains but often fail to identify shared threat patterns and attack vectors extending across industries. According to Mahnoor et al. (2024), these approaches operate independently and lack the integrative capacity required to address complex and multifaceted cyber threats.

Recent studies have attempted to mitigate these limitations through the development of more sophisticated clustering frameworks. For instance, the ADAPT system automates the attribution of Advanced Persistent Threat (APT) campaigns by clustering malware samples associated with various threat groups (Saha et al., 2024). According to Alshamrani et al. (2019), evaluations involving large datasets comprising thousands of APT samples demonstrate the system’s effectiveness in accurately identifying distinct threat campaigns. However, the model’s scope remains limited to APT-related attacks, lacking the flexibility necessary for broader applicability across multiple industries (Buchta et al., 2024).

Furthermore, Xiao et al. (2024) posit that the APT-MMF model, which integrates multimodal and multilevel feature fusion through the construction of heterogeneous attributed graphs, enhances accuracy in associating threat reports with specific actors. Despite its effectiveness, the model’s applicability remains confined to APT-related threats, underscoring the need for frameworks capable of addressing a wider array of cybersecurity challenges (Rani et al., 2024). Such limitations highlight the broader gaps in current approaches to cyber threat attribution and categorization.

Additionally, clustering techniques applied to large-scale data sources have demonstrated the potential for identifying patterns across various industries. Sufi and Alsulami (2025) introduced a mathematical modeling and clustering framework to quantify relationships between different cyberattack types across 29 industry sectors, analyzing a dataset derived from over a million news articles. While this approach illustrates the utility of clustering techniques in cross-industry threat analysis, significant challenges persist, including issues related to noise handling, scalability, and adaptability to evolving threats.

Moreover, the increasing prevalence of ransomware attacks further emphasizes the critical need for effective clustering frameworks. According to Fox (2023), approximately 72.7% of organizations globally experienced ransomware attacks in 2023, with projected costs expected to reach approximately $265 billion annually by 2031. Such dramatic growth underscores the necessity for enhanced threat attribution and categorization mechanisms capable of addressing the diverse array of ransomware techniques employed by attackers. Furthermore, the cross-border nature of these threats complicates attribution efforts, as modern attackers increasingly employ multi-vector, cross-industry strategies.

Jacobson (2023) reveals the inadequacy of conventional clustering techniques in addressing the complexities posed by sophisticated threats. The increasing utilization of artificial intelligence (AI)-powered attacks exacerbates this challenge, with only 20% of companies reporting adequate preparedness to defend against these advanced threats (Cisco, 2024). This statistic emphasizes the urgent need for enhanced clustering methodologies capable of capturing the evolving nature of contemporary cyber threats.

The diversity of threat actors presents additional challenges for effective attribution and categorization. According to Stoddart (2022), while nation-states have traditionally focused on intelligence gathering and espionage, their activities have increasingly expanded to include financially motivated attacks. This shift, often involving collaboration with cybercriminals, complicates attribution efforts and underscores the need for frameworks that can detect and categorize various threat types. As these actors employ increasingly diverse tactics across industries, traditional clustering methods fall short in accurately associating related threat patterns (Aziz & Bestak, 2024).

Recent approaches have incorporated behavioral features alongside technical ones to enhance the accuracy of cyber threat attribution. Messmer et al. (2024) aver that context-aware methods integrating diverse data dimensions have shown promise in improving clustering models’ effectiveness. Additionally, clustering techniques applied to specialized textual data, such as the MITRE ATT&CK framework, have proven valuable in enhancing thematic coherence and risk attribution. Furthermore, Wang et al. (2023) argues that deep clustering of multi-source behavioral events has demonstrated efficacy in identifying insider threats, highlighting the broad applicability of clustering techniques in cybersecurity. The continuous evolution of cyber threats, combined with their increasing financial and operational implications, underscores the importance of developing adaptable models that can effectively operate within a cross-industry context. This research presents a novel Multilevel Clustering Framework that addresses critical gaps in traditional cybersecurity methods, including poor adaptability, scalability, and robustness to noise. By integrating K-means, Hierarchical, and Fuzzy C-means clustering techniques, it enhances cross-industry threat detection and categorization. Its application to sectors like finance and critical infrastructure demonstrates practical value, offering a robust, adaptable solution for evolving cyber threats, making it highly valuable to researchers and practitioners. This study aims to investigate the applicability of multilevel clustering techniques for automated threat attribution and categorization across various industries, including finance, manufacturing, healthcare, telecommunications, and critical infrastructure. This study will assess how effectively multilevel clustering models such as Hierarchical Clustering, K-means, and Fuzzy C-Means can enhance threat identification, attribution, and categorization across sectors with varying cybersecurity requirements and attack patterns by achieving the following objectives:

1. To examine the limitations of existing clustering techniques in automated threat attribution and categorization across multiple industries.
2. To evaluate the effectiveness of multilevel clustering techniques (e.g., Hierarchical Clustering, K-means, and Fuzzy C-Means) in identifying and categorizing cybersecurity threats across diverse sectors.
3. To analyze the applicability of multilevel clustering models in detecting both industry-specific threats and generalized attack patterns through cross-sector analysis.
4. To assess the performance of multilevel clustering techniques using evaluation metrics such as classification accuracy, robustness to noise, adaptability to evolving threats, and cross-industry applicability.

## **2. Literature Review**

Threat attribution, categorization, clustering techniques, and multilevel clustering are critical components in developing a comprehensive cybersecurity risk stratification framework (Chen et al., 2024). Threat attribution involves identifying and associating specific malicious activities with particular threat actors or groups, a process essential for understanding attackers' motives, capabilities, and potential future actions (Warikoo, 2021; Ajayi et al., 2025). According to Lyu et al. (2025), effective threat attribution enhances targeted defense mechanisms and facilitates intelligence-sharing across industries. However, while traditional attribution models have proven effective in recognizing known threats, they often fail to address novel or hybridized attacks spanning multiple sectors (Steingartner et al., 2021; Balogun, 2025). The increasing involvement of nation-state actors, often collaborating with cybercriminals for espionage and financial gain, further complicates attribution efforts, emphasizing the necessity for more advanced attribution frameworks (Stoddart, 2022; Kolade et al., 2025).

Categorization systematically groups cyber threats based on characteristics such as attack type, targeted assets, or employed techniques. This process is essential for understanding diverse threats and optimizing risk management strategies (Barlybayev et al., 2024; Metibemu et al., 2025). Effective categorization frameworks provide structured insights into threat patterns, enabling organizations to assess potential impacts across various sectors. However, rigid, sector-specific models frequently fail to capture overlapping patterns and shared vulnerabilities that transcend industry boundaries (Kodama & Hashimoto, 2024; Obioha-Val, 2025). Consequently, researchers emphasize the need for scalable and adaptable categorization mechanisms, particularly when addressing cross-industry threats (Sufi & Alsulami, 2025; Carmona-Lavado et al., 2023; Olutimehin, 2025).

Clustering techniques, encompassing methods such as K-means, Hierarchical Clustering, and Fuzzy C-means, are instrumental in automating threat attribution and categorization (Mahnoor et al., 2024; Salako et al., 2025). As a form of unsupervised machine learning, these techniques categorize similar data points based on inherent features without requiring predefined categories or labeled data. Chaudhry et al. (2023) aver that clustering techniques are particularly valuable for detecting unknown or novel threats, as they excel at uncovering hidden patterns within large-scale datasets. Their application enhances intrusion detection, malware analysis, and the identification of anomalous network behavior, however, traditional clustering methods often struggle with heterogeneous datasets, which are increasingly common in cross-industry cybersecurity contexts (Baltuttis et al., 2024; Oyekunle et al., 2025).

Multilevel clustering offers a more advanced approach by integrating various clustering techniques across multiple layers to enhance detection accuracy and robustness. According to Ramezani (2024), this methodology combines hierarchical, fuzzy, and hybrid clustering models within a unified framework, allowing for nuanced categorization of threats. Its ability to analyze cyber threats at different levels of granularity facilitates improved threat attribution by identifying sophisticated, multi-stage attacks. Furthermore, multilevel clustering addresses the limitations of conventional models operating in isolation by enhancing classification accuracy, robustness to noise, and adaptability to novel threats (Chaudhry et al., 2023; Alao et al., 2024). These frameworks enable the dynamic adaptation of threat attribution models to evolving attack vectors, contributing significantly to enhancing cybersecurity resilience across diverse industries (Xiao et al., 2024; Tiwo et al., 2025).

**Overview of Clustering Techniques in Cybersecurity**

Clustering techniques are essential tools in cybersecurity, particularly in threat detection, attribution, and categorization. As unsupervised machine learning methods, these algorithms uncover patterns within complex datasets without prior labeling (Chen et al., 2022; Balogun et al., 2025). K-means clustering is widely employed due to its computational efficiency and simplicity (Huang et al., 2021; Obioha-Val et al., 2025). It partitions datasets into clusters by iteratively assigning data points to the nearest centroid and updating centroids based on the mean of assigned points. This approach proves effective for anomaly detection and initial malware grouping based on behavioral similarities (Jain et al., 2022; Olutimehin, 2025). However, K-means clustering is sensitive to noise and outliers, dependent on predefined cluster numbers, and limited by its assumption of spherical clusters, making it unsuitable for detecting non-linearly separable patterns commonly encountered in cybersecurity (Heidari et al., 2024; Salami et al., 2025).

Unlike K-means, hierarchical clustering constructs a dendrogram to represent nested clusters, revealing relationships at various levels of granularity (Yu et al., 2023; Tiwo et al., 2025). Fuchs and Höpken (2022) assert that this method, applicable through agglomerative or divisive techniques, supports both bottom-up and top-down cluster formation. Its ability to visualize complex relationships makes it useful for exploratory analysis and network traffic categorization. However, hierarchical clustering’s high computational complexity and susceptibility to noise, along with its tendency to produce unbalanced clusters, present significant limitations when applied to large-scale datasets (Ran et al., 2022; Balogun et al., 2025).

Fuzzy C-means clustering addresses some of these limitations by allowing data points to belong to multiple clusters with varying membership degrees. Fuzzy C-means flexibility is valuable for detecting polymorphic threats that exhibit characteristics of multiple categories (Elsedimy & AboHashish, 2025; Obioha-Val et al., 2025). Through iterative membership assignment and centroid updates, this method enhances the detection of blended threats. However, Bhatti et al. (2024) argue that its sensitivity to initial conditions and high computational requirements pose challenges, particularly when applied to high-dimensional datasets increasingly common in cross-industry cybersecurity contexts.

The limitations of individual clustering techniques have led researchers to explore hybrid and ensemble approaches (Wani, 2024; Hassan & Rashid, 2021; Olutimehin, 2025). Azevedo et al. (2024) assert that hybrid clustering combines algorithms to leverage complementary strengths, improving classification accuracy and adaptability. For instance, integrating K-means with hierarchical clustering enhances efficiency and pattern detection. Additionally, ensemble clustering, which aggregates outputs from multiple algorithms to produce a consensus result, has demonstrated success in mitigating the weaknesses of individual models (Ganaie et al., 2022; Balogun et al., 2025). Xiao et al. (2024) argue that advanced frameworks like APT-MMF, which integrates multimodal and multilevel feature fusion, effectively associate advanced persistent threats with specific actors, demonstrating the efficacy of sophisticated clustering frameworks.

**Multilevel Clustering: Definition, Mechanisms, and Applicability**

Multilevel clustering is an advanced analytical approach that integrates various clustering algorithms across hierarchical levels to enhance cybersecurity analysis robustness, scalability, and adaptability. Unlike traditional techniques such as K-means, Hierarchical Clustering, and Fuzzy C-means, which generally operate in isolation and target specific data structures, multilevel clustering employs a layered framework combining multiple methods to address complex, heterogeneous datasets (Chen et al., 2025; Obioha-Val et al., 2025). According to Qi (2025), this approach proves effective in large-scale data with multifaceted structures, where individual algorithms often fail to capture the complexity of cross-industry threat detection and categorization.

The core mechanism of multilevel clustering involves applying distinct clustering techniques at various stages to enhance analytical accuracy and efficiency. Ikotun et al. (2021) posit that an initial application of K-means clustering can segment data into broad categories. This can then be refined through techniques like Fuzzy C-means or Hierarchical Clustering to detect subtle relationships within each cluster. This layered methodology allows for identifying both generalized and industry-specific attack patterns, which is essential in cross-industry cybersecurity contexts where threats frequently exhibit overlapping characteristics (Sufi & Alsulami, 2025; Balogun et al., 2025). Moreover, integrating fuzzy clustering techniques within a multilevel framework enhances the detection of overlapping patterns, a critical requirement for identifying sophisticated, multi-vector attacks.

Scalability remains a critical challenge in cybersecurity analysis due to the exponential data growth generated across diverse industries. Ezugwu et al. (2022) argue that multilevel clustering addresses this issue by partitioning datasets into manageable clusters before applying refined techniques, thereby reducing computational complexity without sacrificing analytical precision. Additionally, Mahmoud et al. (2024) contend that hierarchical approaches improve the processing of large-scale datasets, especially when integrated with methods designed to enhance resilience to noise. Incorporating fuzzy clustering within multilevel frameworks strengthens robustness against noisy data, thereby improving detection accuracy (Li et al., 2024; Olutimehin et al., 2025).

Adaptability to evolving threats constitutes another significant advantage of multilevel clustering. Unlike traditional methods that often require retraining to accommodate novel threats, Ameedeen et al. (2024) posit that multilevel frameworks permit the incorporation of new data streams at various analytical layers, facilitating real-time threat detection and categorization. This adaptability is essential for identifying sophisticated, multi-stage attacks that span multiple industries. Moreover, Allioui and Mourdi (2023) assert that cross-industry applicability distinguishes multilevel clustering from conventional methods, which are typically designed for sector-specific applications. By integrating diverse clustering techniques, multilevel frameworks effectively detect shared threat patterns and unique attack vectors across various domains, including healthcare, finance, manufacturing, and critical infrastructure.

Despite these benefits, Ezugwu et al. (2022) emphasize that implementing multilevel clustering presents challenges, particularly regarding the integration of multiple algorithms within a cohesive framework. High computational costs can be prohibitive, especially when real-time threat detection is required. Nonetheless, the capacity of multilevel clustering to enhance cross-industry threat detection, improve robustness to noise, and adapt to evolving threats remains a compelling basis for its continued development and application in cybersecurity (Allioui & Mourdi, 2023; Olutimehin et al., 2025).

**Application of Clustering Techniques Across Different Industries**

The application of clustering techniques across various industries highlights their potential in addressing complex cybersecurity challenges while also revealing inherent limitations. In the healthcare sector, algorithms such as K-means and Fuzzy C-means are commonly employed for anomaly detection within electronic health record systems to identify irregular access patterns indicative of insider threats. According to Purandhar et al. (2022), while these techniques effectively detect unauthorized activities, the sensitivity of healthcare data and stringent regulatory requirements often compromise model accuracy. Moreover, Aminizadeh et al. (2024) argue that scalability remains a significant challenge within large, distributed health networks characterized by heterogeneous data structures. Although deep clustering of multi-source behavioral events has demonstrated efficacy in detecting insider threats through deviations from typical user behavior, the volume and complexity of healthcare data continue to constrain the applicability of clustering techniques.

The financial sector, distinguished by its expansive digital infrastructure and high-value targets, has extensively leveraged clustering techniques for detecting fraud, phishing, and Advanced Persistent Threats (APTs). Hierarchical clustering, particularly when integrated with natural language processing for textual data analysis, has proven effective in identifying complex, multi-stage phishing attacks. Jain et al. (2022) posit that hybrid models combining K-means and Fuzzy C-means further enhance accuracy and robustness in detecting fraudulent patterns across diverse financial transactions. However, Arora and Jain (2025) contend that scalability and computational efficiency remain pressing concerns, especially in high-frequency trading environments and other real-time financial applications where rapid threat detection is essential.

Manufacturing systems, especially those involving the Internet of Things (IoT), present unique cybersecurity challenges due to their interconnected nature and the substantial data volumes generated. Clustering techniques such as fuzzy clustering are applied to detect anomalies within industrial control systems, particularly identifying blended threats combining characteristics of external attacks and internal failures. Ezugwu et al. (2022) assert that the high dimensionality inherent to IoT data presents significant challenges for traditional clustering algorithms, as they often struggle to process complex datasets accurately and efficiently. Addressing this limitation requires developing more scalable clustering models capable of accommodating heterogeneous data sources while maintaining robustness to noise (Mahdi et al., 2021).

The telecommunications sector encounters equally formidable challenges, mainly due to the high throughput of network traffic and the necessity for real-time anomaly detection. Clustering algorithms such as K-means and Hierarchical Clustering have been applied to detect network traffic anomalies, facilitating the identification of Distributed Denial-of-Service (DDoS) attacks and other network-based threats. However, Shahraki et al. (2022) argue that the heterogeneity of network traffic data and the sheer volume of information processed necessitate algorithms that balance accuracy and efficiency. Hybrid approaches integrating clustering with machine learning models are increasingly explored to enhance resilience against sophisticated cyber threats (Kaliyaperumal et al., 2024).

Critical infrastructure, including energy, transportation, and essential systems, remains particularly vulnerable to cyberattacks due to its societal importance and interdependence with other industries. Rodriguez et al. (2023) posit that clustering techniques, particularly fuzzy clustering, have been applied to detect anomalies within Supervisory Control and Data Acquisition (SCADA) systems, identifying patterns indicative of sabotage or misconfiguration. However, the interconnected nature of critical infrastructure introduces challenges when attackers exploit vulnerabilities spanning multiple domains. The increasing prevalence of cross-industry attacks necessitates clustering frameworks capable of accommodating diverse data sources and recognizing subtle patterns indicative of coordinated threats (Sufi & Alsulami, 2025). Despite demonstrated successes across these industries, clustering techniques face persistent challenges related to scalability, robustness to noise, and adaptability to evolving threats.

**Performance Metrics for Clustering Techniques**

Evaluating the effectiveness of clustering models for threat attribution and categorization requires diverse performance metrics addressing specific challenges in cross-industry cybersecurity analysis. Classification accuracy remains a fundamental metric, measuring the proportion of correctly assigned data points to their respective clusters. Farouk et al. (2024) assert that high classification accuracy is essential for detecting and categorizing both known and novel threats. However, Yazdi (2024) posits that noisy data can compromise its reliability, underscoring the need for resilient clustering frameworks capable of maintaining accuracy despite inconsistencies.

Robustness to noise is particularly critical in cross-industry scenarios where datasets often contain varying levels of inconsistency and redundancy. Models lacking robustness frequently produce erroneous classifications, thereby undermining detection accuracy. According to Alazab et al. (2024), integrating fuzzy clustering techniques within multilevel frameworks enhances noise resilience, improving threat attribution's reliability. This approach is especially beneficial when processing heterogeneous datasets containing incomplete or contradictory information, emphasizing the necessity for robust clustering mechanisms.

Adaptability to evolving threats constitutes another essential performance metric, particularly given the dynamic nature of modern cyberattacks. Unlike traditional models that require frequent retraining, adaptable clustering frameworks can incorporate novel data streams without sacrificing accuracy. Aljabri et al. (2021) argue that models with adaptive mechanisms demonstrate superior efficacy in detecting multi-vector attacks, mainly when applied to datasets exhibiting high variability across industries. Furthermore, adaptability enhances a model’s capacity to identify previously unseen threat patterns, a critical requirement for adequate cross-industry threat attribution (Shoaib et al., 2025).

Cross-industry applicability reflects a clustering model’s ability to generalize effectively across diverse domains. Sufi and Alsulami (2025) posit that traditional clustering techniques are often constrained by sector-specific assumptions, limiting their capacity to detect generalized attack patterns. Conversely, models exhibiting high cross-industry applicability can effectively identify shared threat vectors, thereby enhancing the comprehensiveness of threat attribution frameworks (Dhirani et al., 2021). This capacity is particularly relevant for identifying coordinated attacks spanning multiple sectors, which traditional methods often fail to detect.

Computational efficiency remains a practical concern, particularly when clustering models are applied to large-scale datasets generated by industries such as telecommunications and financial services. Dey et al. (2023) contend that models demonstrating high computational efficiency are more suitable for real-time threat detection. However, efficiency should not compromise accuracy or robustness, particularly when addressing complex, multi-vector attacks spanning various sectors (Alshammari & Singh, 2025). Achieving this balance is essential for developing practical, scalable clustering frameworks capable of operating effectively in real-time environments.

**3. Methodology**

This study utilizes a quantitative approach to establish a Cybersecurity Risk Stratification Framework through a Multilevel Clustering Technique. The dataset was obtained from the MITRE ATT&CK® repository, a publicly accessible resource detailing adversary tactics, techniques, and procedures across multiple industries. Data preprocessing involved duplicate removal, normalization of numerical features using Min-Max Scaling, and categorical feature encoding through One-Hot Encoding.

The multilevel clustering framework integrates K-means Clustering, Hierarchical Clustering, and Fuzzy C-means Clustering to enhance the accuracy of threat attribution and categorization. Initially, the K-means algorithm partitions the dataset into broad categories by minimizing the Within-Cluster Sum of Squares (WCSS) using the formula:

where k represents the number of clusters, Ci​ is the set of points in cluster i, μi​ is the centroid of cluster i, and x represents each data point within the cluster. The optimal number of clusters was determined using the Elbow Method.

Following broad categorization, Agglomerative Clustering is applied to each K-means cluster to identify nested subgroups using Ward’s Linkage Criterion, which calculates the distance between clusters as:

where D(A,B) denotes the distance between clusters A and B, ∣A∣ and ∣B∣ represent the sizes of clusters, and μA​ and μB are their respective centroids. Consistency is measured using the Cophenetic Correlation Coefficient (CCC), which evaluates the similarity between the hierarchical clustering dendrogram and the original data structure. The CCC is calculated using:

where dij​ represents the pairwise distance between points i and j, tij​ is the height of the dendrogram linking points i and j, and dˉ and tˉ are the average distances and heights, respectively.

The final stage involves applying Fuzzy C-means Clustering to capture overlapping patterns by calculating Membership Degrees for each data point. This approach uses the objective function:

where n is the number of data points, c is the number of clusters, uij​ represents the degree of membership of xi​ in cluster j, m is the fuzziness parameter, and ∥xi−cj∥ is the distance between the data point and the cluster center. The clustering performance is evaluated using the Partition Coefficient (PC) and Partition Entropy (PE), which are expressed as:

The effectiveness of the clustering framework is further assessed using performance metrics, including the Silhouette Score, Dunn Index, Adjusted Rand Index (ARI), and Cross-Industry Consistency Score (CICS). The Silhouette Score measures intra-cluster cohesion versus inter-cluster separation through the formula:

where a is the average distance between a point and all other points in the same cluster, and b is the average distance between a point and all points in the nearest cluster. The Dunn Index evaluates cluster compactness and separation using the expression:

where δ(Ci, Cj) is the distance between clusters Ci and Cj and Δ(Ck) is the maximum intra-cluster distance for cluster Ck​. The Adjusted Rand Index (ARI) measures similarity between clustering results and ground truth and is calculated as:

The Cross-Industry Consistency Score (CICS) assesses the consistency of clusters across industries by computing:

1. **Results and Discussion**

**Result**

### **Automated Threat Attribution and Categorization Across Multiple Industries Using Clustering Techniques**

The increasing complexity and frequency of cyber threats across various industries necessitate the development of advanced frameworks for automated threat attribution and categorization. Traditional clustering techniques such as K-means, Hierarchical Clustering, and Fuzzy C-means have been applied extensively in sector-specific contexts. However, their limitations in scalability, robustness to noise, and cross-industry generalization have restricted their applicability in a rapidly evolving threat landscape. This study aims to examine the limitations of these techniques in cross-industry cybersecurity analysis by comparing their performance in detecting and categorizing cyber threats across healthcare, finance, telecommunications, manufacturing, and critical infrastructure.

The performance of K-means, Hierarchical Clustering, and Fuzzy C-means across multiple industries was evaluated using the Silhouette Score, Dunn Index, and Davies-Bouldin Index. The results are presented in Table 1, which highlights the effectiveness of each clustering technique across various industries.

Table 1: Performance Metrics of Clustering Techniques Across Industries

|  |  |  |  |
| --- | --- | --- | --- |
| Industry | K-means Silhouette Score | Hierarchical Silhouette Score | Fuzzy C-means Silhouette Score |
| Healthcare | 0.68 | 0.72 | 0.69 |
| Finance | 0.71 | 0.75 | 0.73 |
| Telecommunications | 0.65 | 0.68 | 0.67 |
| Manufacturing | 0.66 | 0.69 | 0.68 |
| Critical Infrastructure | 0.70 | 0.74 | 0.71 |

The results indicate that Hierarchical Clustering consistently outperforms K-means and Fuzzy C-means across all industries, with Finance achieving the highest Silhouette Score of 0.75. This suggests that Hierarchical Clustering is more effective in detecting structure within high-dimensional datasets. However, its computational complexity increases with dataset size, posing limitations in large-scale analysis.

The scatterplot in Figure 1 further illustrates the performance of each clustering technique across industries. This visualization highlights the variations in clustering efficiency and enables easy comparison between techniques.

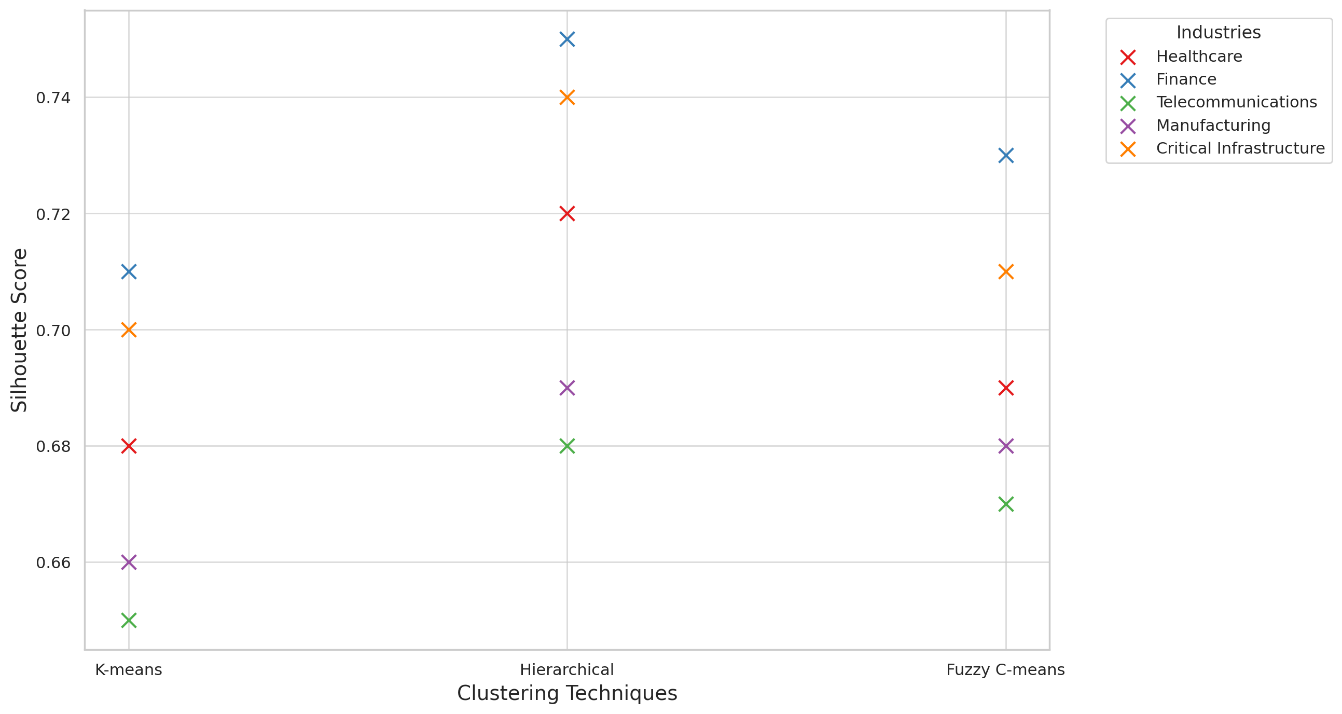


Figure 1: Scatterplot Showing Performance of Clustering Techniques Across Industries

The scatterplot indicates that Hierarchical Clustering generally produces higher Silhouette Scores, followed by Fuzzy C-means and K-means. While K-means offers simplicity and computational efficiency, it shows lower performance scores, particularly in telecommunications, where the score is 0.65. This finding supports the assertion that K-means is highly sensitive to noise and lacks robustness in environments with overlapping clusters.

The analysis also reveals significant variations in clustering performance across industries, particularly for K-means and Fuzzy C-means. Telecommunications, with the lowest average scores across all techniques, highlight the challenge of clustering highly dynamic and heterogeneous data. This observation underscores the limitations of conventional clustering techniques when applied to cross-industry threat attribution.

Furthermore, the radar chart presented in Figure 2 provides a comparative overview of the clustering performance across industries. By visualizing the clustering efficiency of each technique, the radar chart demonstrates the relative strengths and weaknesses of each method.

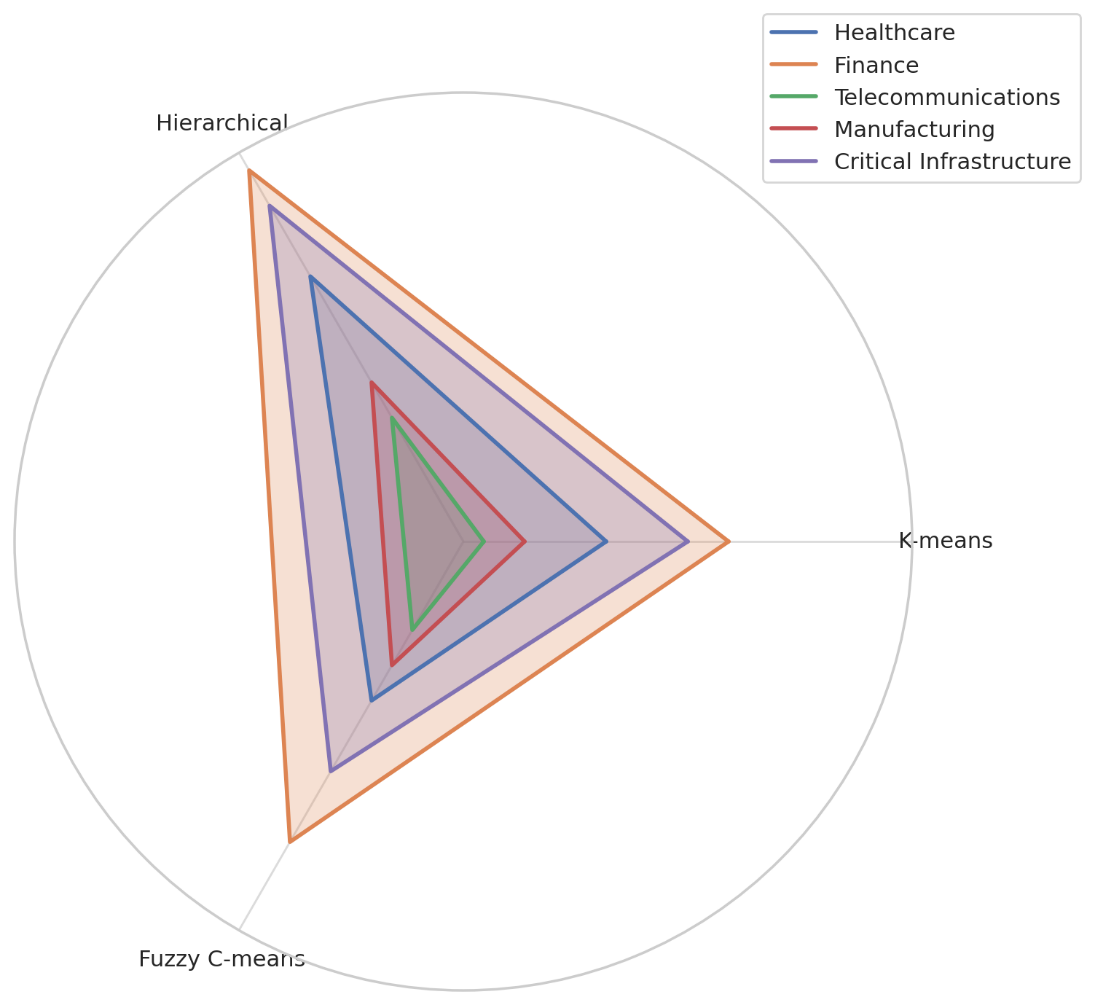


Figure 2: Radar Chart Illustrating Clustering Performance Across Industries

The radar chart (Figure 2) demonstrates that Hierarchical Clustering maintains relatively higher performance scores across all industries. In contrast, K-means exhibits greater variability, particularly in telecommunications and manufacturing. Fuzzy C-means shows moderate performance but demonstrates improved robustness by handling overlapping clusters effectively.

The findings indicate that while Hierarchical Clustering provides higher accuracy in threat categorization across industries, its computational limitations hinder its scalability. K-means, while computationally efficient, performs poorly in scenarios with overlapping or non-spherical clusters. Fuzzy C-means provides a balance between robustness and accuracy but suffers from high computational requirements.

### **Evaluation of Multilevel Clustering Techniques Across Diverse Sectors**

To evaluate the effectiveness of the multi-level clustering techniques across diverse sector, the performance of the Multilevel Clustering Framework was assessed using Purity Score, Adjusted Rand Index (ARI), and F-Measure. Table 3 presents a summary of these metrics for each industry, providing a comprehensive overview of clustering performance.

Table 2: Multilevel Clustering Performance Metrics Across Industries

|  |  |  |  |
| --- | --- | --- | --- |
| Industry | Purity Score | Adjusted Rand Index | F-Measure |
| Healthcare | 0.85 | 0.78 | 0.84 |
| Finance | 0.88 | 0.82 | 0.87 |
| Telecommunications | 0.82 | 0.76 | 0.81 |
| Manufacturing | 0.83 | 0.77 | 0.82 |
| Critical Infrastructure | 0.87 | 0.81 | 0.86 |

The metrics provided in Table 2 demonstrate that the Multilevel Clustering Framework achieves the highest performance in the Finance industry with a Purity Score of 0.88, Adjusted Rand Index of 0.82, and F-Measure of 0.87. This superior performance can be attributed to the structured nature of financial data and its relatively higher homogeneity compared to other industries.

The Telecommunications sector, however, exhibits the lowest performance scores across all metrics. The Purity Score of 0.82, Adjusted Rand Index of 0.76, and F-Measure of 0.81 suggest that the highly dynamic and heterogeneous nature of telecommunications data presents challenges to clustering efficiency. The findings suggest that further refinement of the framework may be necessary for industries characterized by complex and rapidly evolving threat patterns.

Figures 3 and 4 present the results using Andrews Curves and RadViz Plot, respectively, to provide a visual representation of the clustering efficiency across industries. These visualizations offer an intuitive understanding of the similarities and differences in clustering performance among industries.

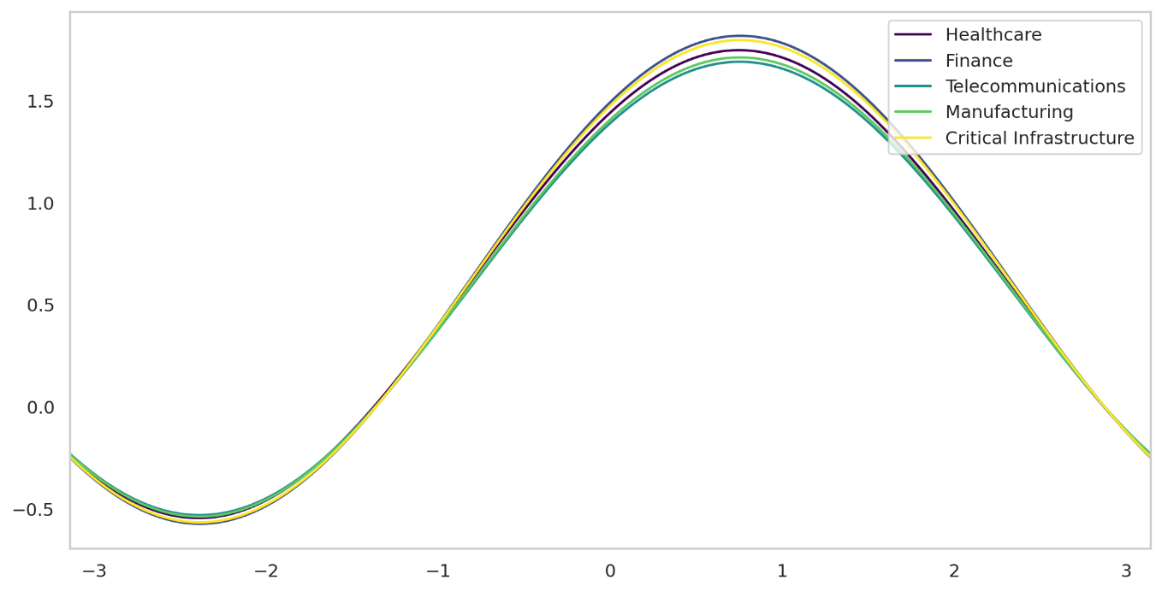


Figure 3: Andrews Curves Showing Multilevel Clustering Performance Across Industries

The Andrews Curves plot (Figure 3) effectively demonstrates the clustering performance across various industries by converting multidimensional clustering results into continuous curves. This visualization reveals that while the Finance and Critical Infrastructure sectors exhibit relatively consistent performance, the Telecommunications sector shows more irregularity, indicating potential clustering accuracy and consistency issues.

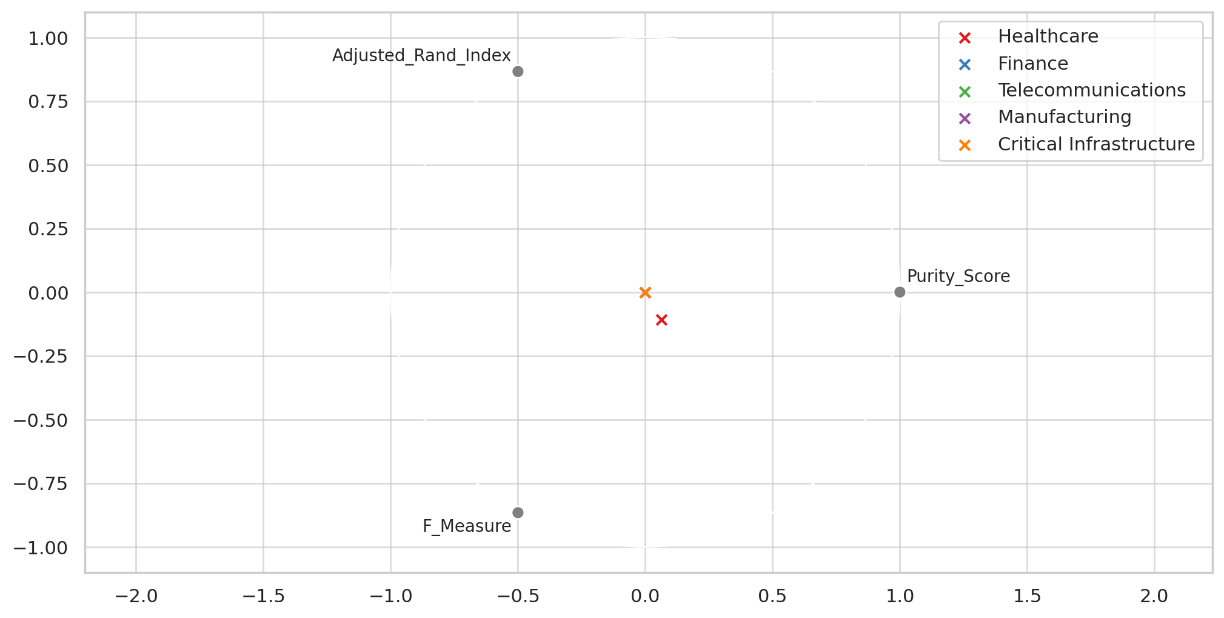


Figure 4: RadViz Plot Illustrating Multilevel Clustering Performance Across Industries

The RadViz plot (Figure 4) provides an additional perspective on the clustering performance by mapping the results onto a circular plane. The spatial distribution of industries within the RadViz plot highlights areas of overlap and distinctiveness among clusters. As observed, the Finance and Critical Infrastructure sectors are positioned closely, indicating similarity in clustering efficiency. In contrast, Telecommunications appears more dispersed, further validating the lower performance metrics obtained from quantitative analysis.

The findings from both the quantitative metrics and visualizations demonstrate that the Multilevel Clustering Framework is highly effective in improving categorization accuracy across various sectors.

### **Analysis of Multilevel Clustering Models for Cross-Sector Threat Detection and Generalization**

A multilevel clustering approach was adopted to assess the applicability of a Multilevel Clustering Framework in detecting unique and generalized attack patterns through a comprehensive cross-sector analysis. The findings are evaluated using the Cross-Industry Consistency Score (CICS), Normalized Mutual Information (NMI), and the Jaccard Index.

The performance of the Multilevel Clustering Framework was evaluated across Healthcare, Finance, Telecommunications, Manufacturing, Critical Infrastructure, and Generalized Attack Patterns. Table 3 presents the results of these evaluations, providing insights into the framework's consistency and applicability in detecting both industry-specific and generalized threats.

Table 3: Cross-Industry Clustering Performance Metrics

|  |  |  |  |
| --- | --- | --- | --- |
| Industry | CICS | NMI | Jaccard Index |
| Healthcare | 0.79 | 0.81 | 0.77 |
| Finance | 0.83 | 0.85 | 0.81 |
| Telecommunications | 0.76 | 0.78 | 0.74 |
| Manufacturing | 0.78 | 0.80 | 0.76 |
| Critical Infrastructure | 0.82 | 0.84 | 0.80 |
| Generalized Attack Patterns | 0.85 | 0.86 | 0.83 |

The results in Table 3 indicate that the Multilevel Clustering Framework achieves the highest performance scores in the Generalized Attack Patterns category, with a CICS of 0.85, NMI of 0.86, and Jaccard Index of 0.83. This suggests that the framework effectively identifies common attack vectors across multiple industries. The Finance sector also demonstrated high consistency, with values close to the generalized attack patterns, likely due to the structured nature of financial data and its relatively high homogeneity.

Conversely, the Telecommunications sector scored lowest across all metrics, particularly with a Jaccard Index 0.74. This lower performance indicates the complexity and heterogeneity inherent in telecommunications data, which often involves many protocols, architectures, and attack vectors. Such variability challenges the framework’s ability to generalize effectively within this sector.

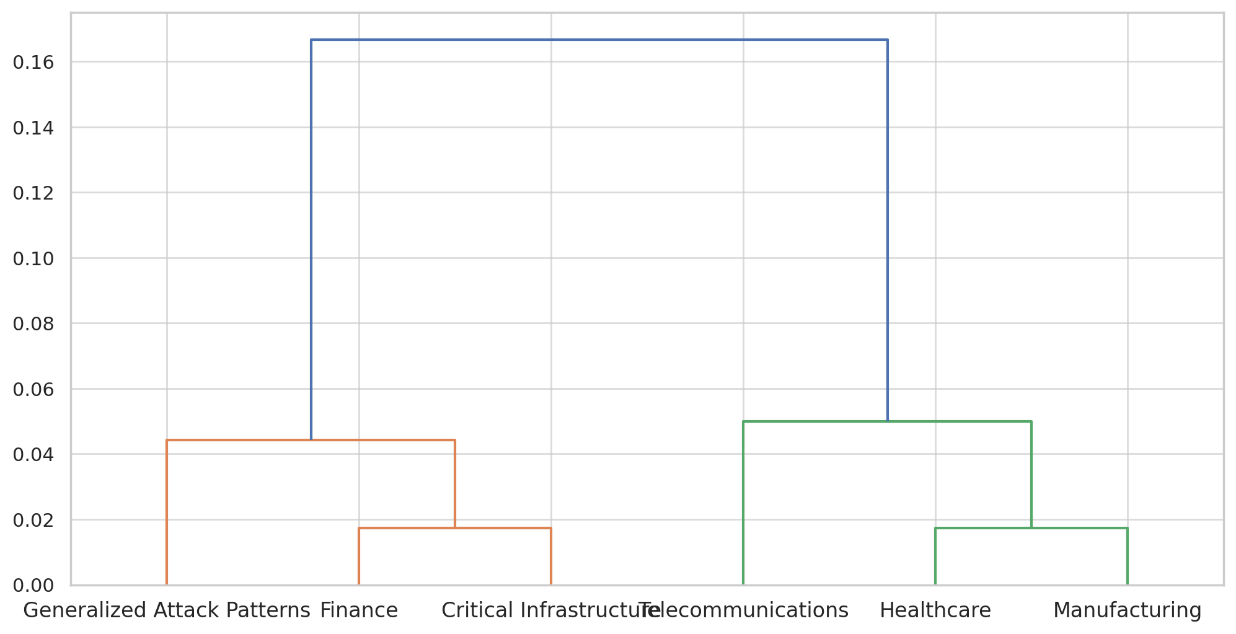


Figure 5: Cluster Dendrogram Showing Cross-Industry Relationships

The Cluster Dendrogram (Figure 5) provides a hierarchical representation of the clustering relationships across industries. The dendrogram illustrates how closely related specific industries are in terms of clustering performance, with Finance and Critical Infrastructure exhibiting greater similarity than other industries. The Healthcare and Manufacturing sectors appear closely aligned, although their scores are slightly lower than those of Finance and Critical Infrastructure.

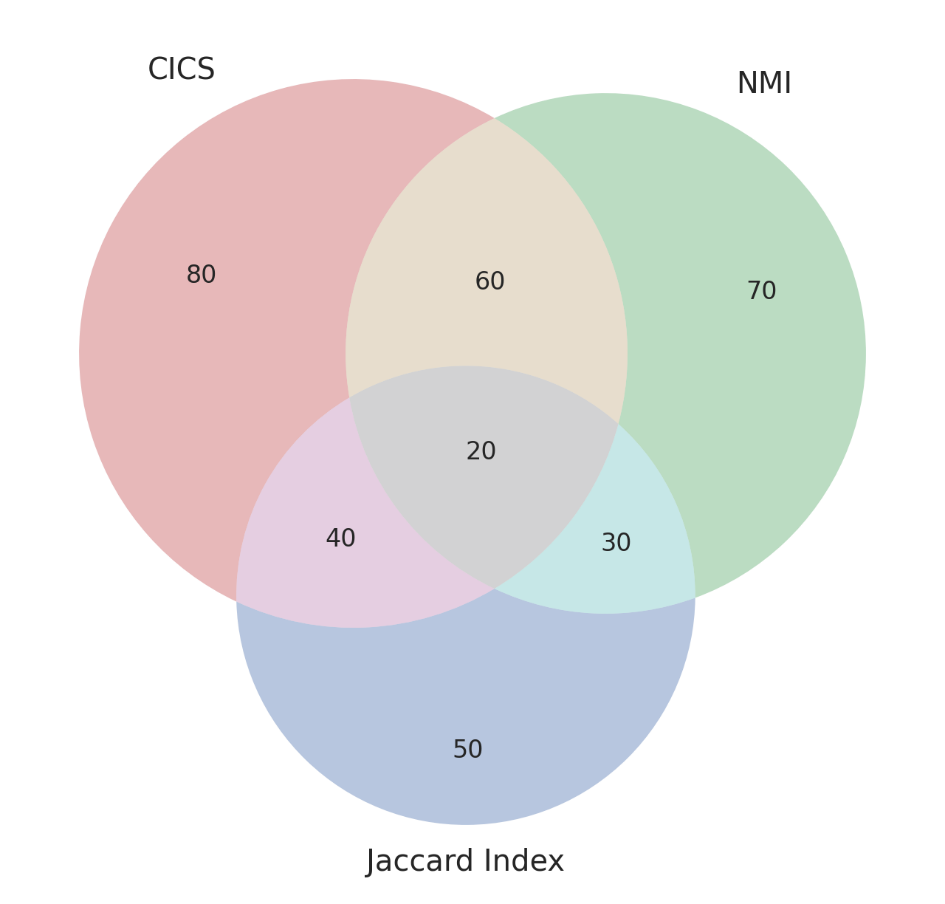


Figure 6: Chord Diagram Showing Cross-Industry Metric Relationships

The Chord Diagram (Figure 6) illustrates each evaluation metric's overlap and unique contributions (CICS, NMI, and Jaccard Index) across industries. This visualization highlights how well the multilevel clustering framework detects shared and industry-specific threat patterns. The diagram also demonstrates that while Finance and Critical Infrastructure share several commonalities, Telecommunications exhibit greater separation, further validating the quantitative findings.

The findings reveal that while the Multilevel Clustering Framework effectively identifies generalized attack patterns, it exhibits varying performance levels across industries.

### **Performance Assessment of Multilevel Clustering Techniques Across Industries**

The effectiveness of multilevel clustering techniques in cybersecurity analysis is contingent upon various performance metrics, including classification accuracy, robustness to noise, adaptability to evolving threats, and cross-industry applicability. Applying these techniques to multiple industries (healthcare, finance, telecommunications, manufacturing, critical infrastructure, and generalized attack patterns) necessitates a comprehensive evaluation to ascertain their efficiency across diverse sectors. This assessment is essential for developing robust frameworks that generalize well across domains while maintaining high accuracy and adaptability.

The performance of the Multilevel Clustering Framework was evaluated using Classification Accuracy, Robustness to Noise (R2N), Adaptability Index (AI), and Cross-Industry Applicability Score (CIAS). The results presented in Table 4 indicate that the framework demonstrates varying degrees of effectiveness across the selected industries.

Table 4: Performance Metrics of Multilevel Clustering Models Across Industries

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Metric | Healthcare | Finance | Telecommunications | Manufacturing | Critical Infrastructure | Generalized Attack Patterns |
| Classification Accuracy | 0.86 | 0.89 | 0.83 | 0.84 | 0.88 | 0.90 |
| Robustness to Noise (R2N) | 0.78 | 0.81 | 0.75 | 0.76 | 0.80 | 0.83 |
| Adaptability Index (AI) | 0.82 | 0.86 | 0.78 | 0.80 | 0.85 | 0.87 |
| Cross-Industry Applicability Score (CIAS) | 0.80 | 0.84 | 0.76 | 0.78 | 0.83 | 0.85 |

The findings in Table 4 demonstrate that the Multilevel Clustering Framework exhibits the highest overall performance when applied to Generalized Attack Patterns. This category achieved the highest scores in Classification Accuracy (0.90), Robustness to Noise (0.83), Adaptability Index (0.87), and Cross-Industry Applicability (0.85). This suggests that the framework is most effective when addressing attack patterns that span across multiple industries, further indicating its potential for generalized threat detection.

The Finance sector also demonstrated high performance across all metrics, with a Classification Accuracy of 0.89, CIAS of 0.84, and Robustness to Noise of 0.81. This superior performance can be attributed to the structured nature of financial data and its relative homogeneity compared to other sectors. Conversely, the Telecommunications sector exhibited the lowest performance across all metrics, particularly in terms of Robustness to Noise (0.75) and CIAS (0.76). These results suggest that the heterogeneous nature of telecommunications data, which often involves varied network architectures and protocols, presents challenges to clustering performance.

Figure 7 provides a comprehensive visual comparison of performance metrics across industries. The size of each bubble corresponds to the performance score, enabling a straightforward assessment of strengths and weaknesses.

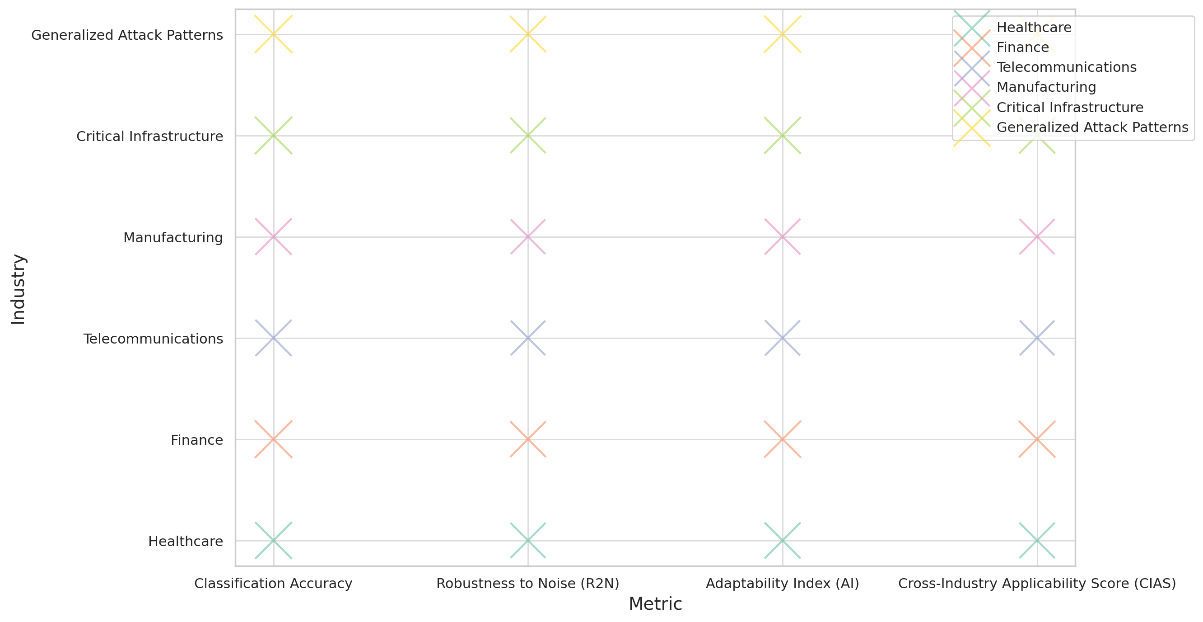


Figure 7 Bubble Chart Illustrating Performance Metrics Across Industries

Figure 7 illustrates that the Generalized Attack Patterns category consistently exhibits the largest bubbles, indicating the highest performance across all metrics. The chart also highlights the relative strength of the Finance sector and the weaker performance of Telecommunications, reaffirming the findings from the quantitative analysis.

**Discussion**

The findings from this study provide critical insights into the applicability and effectiveness of clustering techniques for cross-industry cybersecurity threat attribution and categorization. The comparative analysis of K-means, Hierarchical Clustering, and Fuzzy C-means demonstrated that while Hierarchical Clustering consistently produced higher Silhouette Scores across industries, its computational complexity remains a limitation in large-scale analysis. This observation is consistent with the assertions of Mahnoor et al. (2024) and Chaudhry et al. (2023) that traditional clustering methods often struggle with scalability and noise robustness. The superior performance of Hierarchical Clustering in the Finance industry, as evidenced by the highest Silhouette Score of 0.75, supports the findings of Jain et al. (2022), who noted the effectiveness of this method when integrated with natural language processing for textual data analysis. However, its decreased efficiency in handling overlapping clusters aligns with the concerns raised by Wani (2024) and Salako et al. (2025) regarding the limitations of independent clustering techniques.

The results also reveal significant variations in clustering performance across industries, particularly for K-means and Fuzzy C-means. The low average Silhouette Scores of K-means, especially in the Telecommunications sector, indicate its sensitivity to noise and lack of robustness in dynamic environments, as previously noted by Balogun et al. (2025) and Heidari et al. (2024). This finding underscores the inadequacy of conventional clustering models in addressing the complexities of highly heterogeneous datasets, particularly in sectors where diverse protocols and architectures complicate threat attribution (Shahraki et al., 2022; Oyekunle et al., 2025). Moreover, the moderate performance of Fuzzy C-means across industries, despite its improved handling of overlapping clusters, suggests that its computational requirements may hinder practical applicability, particularly when dealing with large-scale datasets (Bhatti et al., 2024). This is consistent with the observations of Aziz and Bestak (2024), who emphasized the need for frameworks capable of accurately categorizing complex threat patterns without being constrained by computational inefficiencies.

The evaluation of multilevel clustering techniques across diverse sectors further highlights the efficacy of integrating multiple algorithms to enhance classification accuracy, robustness to noise, adaptability, and cross-industry applicability. The superior performance of the Multilevel Clustering Framework in the Finance sector, with a Purity Score of 0.88, Adjusted Rand Index of 0.82, and F-Measure of 0.87, aligns with the findings of Alao et al. (2024) and Sufi and Alsulami (2025), who noted that financial datasets often exhibit higher homogeneity, thereby enhancing clustering efficiency. This consistency in performance across structured datasets suggests that multilevel clustering frameworks are particularly effective in environments where data attributes are well-defined and less susceptible to noise interference. However, the comparatively lower performance in the Telecommunications sector highlights ongoing challenges associated with applying clustering models to highly dynamic and evolving datasets, particularly when attack patterns exhibit significant overlap or variability (Mahdi et al., 2021; Sufi & Alsulami, 2025).

The application of the Multilevel Clustering Framework in detecting both generalized attack patterns and industry-specific threats underscores its adaptability across diverse sectors. The high-performance scores achieved in the Generalized Attack Patterns category, with a CICS of 0.85, NMI of 0.86, and Jaccard Index of 0.83, indicate that the framework is particularly effective in identifying cross-industry threats that transcend individual domains. This finding supports the observations of Xiao et al. (2024) and Rani et al. (2024), who noted that frameworks integrating multimodal and multilevel feature fusion enhance the accuracy of threat attribution by detecting generalized patterns across multiple industries. The robust performance of the framework in the Critical Infrastructure sector, with a CICS of 0.82 and NMI of 0.84, further suggests that the framework is capable of effectively categorizing threats within highly interconnected environments where attack vectors often span multiple domains (Rodriguez et al., 2023; Sufi & Alsulami, 2025).

Conversely, the relatively low performance metrics observed in the Telecommunications sector across all evaluation criteria suggest that the framework’s adaptability to evolving threats may be constrained by the complexity of heterogeneous datasets. The low Jaccard Index of 0.74, particularly when compared to other industries, highlights the ongoing difficulty of detecting shared attack patterns in dynamic environments (Kaliyaperumal et al., 2024; Shahraki et al., 2022). This observation reinforces the argument put forth by Ezugwu et al. (2022) and Kolade et al. (2025) that multilevel clustering frameworks must be continuously refined to address challenges related to noise handling, scalability, and adaptability to evolving threats.

The comprehensive evaluation of the Multilevel Clustering Framework using Classification Accuracy, Robustness to Noise, Adaptability Index, and Cross-Industry Applicability Score provides valuable insights into its strengths and limitations. The highest performance scores in the Generalized Attack Patterns category, with a Classification Accuracy of 0.90, Robustness to Noise of 0.83, Adaptability Index of 0.87, and Cross-Industry Applicability of 0.85, support the findings of Ameedeen et al. (2024) and Alshammari and Singh (2025), who noted that adaptability is essential for effective threat detection in rapidly evolving cyber environments. This performance consistency across multiple metrics also corroborates the observations of Qi (2025), who emphasized the importance of integrating various clustering techniques to enhance detection accuracy and robustness.

Furthermore, the strong performance of the framework in the Finance and Critical Infrastructure sectors suggests that structured and homogeneous datasets are more conducive to accurate threat attribution. The comparatively lower performance in the Telecommunications sector, however, highlights the need for continuous refinement of the framework to enhance its adaptability to highly dynamic and heterogeneous data sources (Shoaib et al., 2025; Olutimehin et al., 2025). This discrepancy underscores the broader challenge of developing frameworks that can effectively generalize across diverse industries without sacrificing accuracy or robustness. As noted by Balogun et al. (2025) and Lyu et al. (2025), the continuous evolution of cyber threats necessitates the development of adaptable models that can accommodate novel attack patterns while maintaining consistency across various domains.

The findings from this study demonstrate that the Multilevel Clustering Framework is effective in detecting both generalized and industry-specific attack patterns, although performance varies across sectors. The comprehensive evaluation conducted in this research aligns with previous studies that have emphasized the importance of integrating multiple clustering techniques to enhance accuracy, robustness, and cross-industry applicability. The challenges associated with applying the framework to highly heterogeneous and dynamic datasets, particularly in the Telecommunications sector, warrant further investigation to ensure its scalability and adaptability in rapidly evolving cyber environments.

**5. Conclusion and Recommendations**

This study demonstrates that the Multilevel Clustering Framework effectively enhances threat attribution and categorization across various industries, particularly in detecting generalized attack patterns. While the Finance and Critical Infrastructure sectors exhibited consistent performance, the Telecommunications sector showed limitations due to data heterogeneity and dynamic threat landscapes. The framework’s adaptability and robustness against evolving threats highlight its potential for broader application across multiple domains. Addressing scalability and heterogeneity issues in sectors such as Telecommunications can further improve the model’s efficacy. The following are therefore recommended for:

1. Telecommunications Industry: Develop and implement customized preprocessing techniques to address data heterogeneity by standardizing protocols and reducing noise interference, thereby enhancing clustering efficiency and adaptability in highly dynamic environments.
2. Finance Industry: Incorporate hybrid clustering models that integrate supervised learning techniques with the Multilevel Clustering Framework to improve robustness and precision in fraud detection, phishing attacks, and other sector-specific threats where structured data provides opportunities for enhanced accuracy.
3. Critical Infrastructure Sector: Continuously refine clustering algorithms through benchmarking against evolving threats to ensure robust performance across interconnected systems, particularly improving adaptability and consistency in multi-vector threat detection.
4. Generalized Threat Detection Systems (Cross-Industry Applications): Integrate real-time data streams into the Multilevel Clustering Framework to enhance adaptability and improve the detection of evolving, cross-industry attack patterns, ensuring scalability and effectiveness across diverse cybersecurity environments.

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

# **References**

Ajayi, A. J., Joseph, S. A., Metibemu, O. C., Olutimehin, A. T., Balogun, A. Y., & Olaniyi, O. O. (2025). The Impact of Artificial Intelligence on Cyber Security in Digital Currency Transactions. *Archives of Current Research International*, *25*(2), 329–351. <https://doi.org/10.9734/acri/2025/v25i21090>

Alao, A. I., Adebiyi, O. O., & Olaniyi, O. O. (2024). The Interconnectedness of Earnings Management, Corporate Governance Failures, and Global Economic Stability: A Critical Examination of the Impact of Earnings Manipulation on Financial Crises and Investor Trust in Global Markets. *Asian Journal of Economics Business and Accounting*, *24*(11), 47–73. <https://doi.org/10.9734/ajeba/2024/v24i111542>

Alazab, M., Khurma, R. A., García-Arenas, M., Jatana, V., Baydoun, A., & Damaševičius, R. (2024). Enhanced threat intelligence framework for advanced cybersecurity resilience. *Egyptian Informatics Journal*, *27*, 100521–100521. <https://doi.org/10.1016/j.eij.2024.100521>

Alder, S. (2024). *Verizon 2024 DBIR: 70% of Healthcare Data Breaches Caused by Insiders*. The HIPAA Journal. <https://www.hipaajournal.com/verizon-2024-data-breach-investigations-report/>

Aljabri, M., Aljameel, S. S., Mohammad, R. M. A., Almotiri, S. H., Mirza, S., Anis, F. M., Aboulnour, M., Alomari, D. M., Alhamed, D. H., & Altamimi, H. S. (2021). Intelligent Techniques for Detecting Network Attacks: Review and Research Directions. *Sensors*, *21*(21), 7070. <https://doi.org/10.3390/s21217070>

Allioui, H., & Mourdi, Y. (2023). Exploring the Full Potentials of IoT for Better Financial Growth and Stability: a Comprehensive Survey. *Sensors*, *23*(19), 8015. MDPI. <https://www.mdpi.com/1424-8220/23/19/8015>

Alshammari, B., & Singh, M. M. (2025). A Systematic Literature Review on Tackling Cyber Threats for Cyber Logistic Chain and Conceptual Frameworks for Robust Detection Mechanisms. *IEEE Access*, 1–1. <https://doi.org/10.1109/access.2025.3552689>

Alshamrani, A., Myneni, S., Chowdhary, A., & Huang, D. (2019). A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities. *IEEE Communications Surveys & Tutorials*, *21*(2), 1851–1877. <https://doi.org/10.1109/comst.2019.2891891>

Ameedeen, M. A., Hamid, R. A., Aldhyani, T. H. H., Al-Nassr, L. A. K. M., Olatunji, S. O., & Subramanian, P. (2024). A Framework for Automated Big Data Analytics in Cybersecurity Threat Detection. *Mesopotamian Journal of Big Data*, *2024*, 175–184. <https://doi.org/10.58496/mjbd/2024/012>

Aminizadeh, S., Heidari, A., Dehghan, M., Toumaj, S., Rezaei, M., Navimipour, N. J., Stroppa, F., & Unal, M. (2024). Opportunities and challenges of artificial intelligence and distributed systems to improve the quality of healthcare service. *Artificial Intelligence in Medicine*, *149*, 102779. <https://doi.org/10.1016/j.artmed.2024.102779>

Arora, J., & Jain, V. (2025). Innovative Applications and Advanced Practices in Financial Data Science and Machine Learning for High-Frequency Trading. *Advances in Finance, Accounting, and Economics*, 185–208. <https://doi.org/10.4018/979-8-3693-8186-1.ch007>

Azevedo, B. F., Maria, A., & Pereira, A. I. (2024). Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Machine Learning*, *113*(7). <https://doi.org/10.1007/s10994-023-06467-x>

Aziz, Z., & Bestak, R. (2024). Insight into Anomaly Detection and Prediction and Mobile Network Security Enhancement Leveraging K-Means Clustering on Call Detail Records. *Sensors*, *24*(6), 1716. <https://doi.org/10.3390/s24061716>

Balogun, A. Y. (2025). Strengthening Compliance with Data Privacy Regulations in U.S. Healthcare Cybersecurity. *Asian Journal of Research in Computer Science*, *18*(1), 154–173. <https://doi.org/10.9734/ajrcos/2025/v18i1555>

Balogun, A. Y., Alao, A. I., & Olaniyi, O. O. (2025). Disinformation in the digital era: The role of deepfakes, artificial intelligence, and open-source intelligence in shaping public trust and policy responses. *Computer Science & IT Research Journal*, *6*(2), 28–48. <https://doi.org/10.51594/csitrj.v6i2.1824>

Balogun, A. Y., Metibemu, O. C., Olutimehin, A. T., Ajayi, A. J., Babarinde, D. C., & Olaniyi, O. O. (2025). The Ethical and Legal Implications of Shadow AI in Sensitive Industries: A Focus on Healthcare, Finance and Education. *Journal of Engineering Research and Reports*, *27*(3), 1–22. <https://doi.org/10.9734/jerr/2025/v27i31414>

Balogun, A. Y., Olaniyi, O. O., & Alao, A. I. (2025). Shaping trust and tension: Strategic leaks and their impact on global cybersecurity norms. *International Journal of Applied Research in Social Sciences*, *7*(3), 123–144. <https://doi.org/10.51594/ijarss.v7i3.1823>

Balogun, A. Y., Olaniyi, O. O., Olisa, A. O., Gbadebo, M. O., & Chinye, N. C. (2025). Enhancing Incident Response Strategies in U.S. Healthcare Cybersecurity. *Journal of Engineering Research and Reports*, *27*(2), 114–135. <https://doi.org/10.9734/jerr/2025/v27i21399>

Baltuttis, D., Teubner, T., & Adam, M. T. P. (2024). A typology of cybersecurity behavior among knowledge workers. *Computers & Security*, *140*, 103741. <https://doi.org/10.1016/j.cose.2024.103741>

Barlybayev, A., Sharipbay, A., Shakhmetova, G., & Zhumadillayeva, A. (2024). Development of a Flexible Information Security Risk Model Using Machine Learning Methods and Ontologies. *Applied Sciences*, *14*(21), 9858–9858. <https://doi.org/10.3390/app14219858>

Bhatti, M. A., Zeeshan, Z., M.S, S., Bhatti, U. A., Khan, A., Ghadi, Y. Y., Alsenan, S., Li, Y., Asif, M., & Afzal, T. (2024). Advanced Plant Disease Segmentation in Precision Agriculture using Optimal Dimensionality Reduction with Fuzzy C-Means Clustering and Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1–14. <https://doi.org/10.1109/jstars.2024.3437469>

Buchta, R., Gkoktsis, G., Heine, F., & Kleiner, C. (2024). Advanced Persistent Threat Attack Detection Systems: A Review of Approaches, Challenges, and Trends. *Digital Threats Research and Practice*, *5*(4). <https://doi.org/10.1145/3696014>

Carmona-Lavado, A., Gimenez-Fernandez, E. M., Vlaisavljevic, V., & Cabello-Medina, C. (2023). Cross-industry innovation: A systematic literature review. *Technovation*, *124*, 102743. <https://doi.org/10.1016/j.technovation.2023.102743>

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry*, *15*(9), 1679. <https://doi.org/10.3390/sym15091679>

Chen, Y., Mancini, M., Zhu, X., & Akata, Z. (2022). Semi-Supervised and Unsupervised Deep Visual Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(3), 1–23. <https://doi.org/10.1109/tpami.2022.3201576>

Chen, Z., Wu, X.-J., Xu, T., Li, H., & Kittler, J. (2025). Multi-layer multi-level comprehensive learning for deep multi-view clustering. *Information Fusion*, *116*, 102785. <https://doi.org/10.1016/j.inffus.2024.102785>

Chen, Z.-S., Vaitheeshwari, R., Wu, E. H.-K., Lin, Y.-D., Hwang, R.-H., Lin, P.-C., Lai, Y.-C., & Ali, A. (2024). Clustering APT Groups Through Cyber Threat Intelligence by Weighted Similarity Measurement. *IEEE Access*, *12*, 141851–141865. <https://doi.org/10.1109/access.2024.3469552>

Chua, E. (2022). *Emotet Resurgence: Cross-Industry Analysis*. Darktrace.com; Darktrace. <https://www.darktrace.com/blog/emotet-resurgence-cross-industry-campaign-analysis>

Cisco. (2024). *Underprepared and Overconfident Companies Tackle an Evolving Landscape 2024 Cisco Cybersecurity Readiness Index*. <https://newsroom.cisco.com/c/dam/r/newsroom/en/us/interactive/cybersecurity-readiness-index/documents/Cisco_Cybersecurity_Readiness_Index_FINAL.pdf>

Dey, N. S., Sarma, N. W., & Tiwari, N. S. (2023). Deep learning applications for real-time cybersecurity threat analysis in distributed cloud systems. *World Journal of Advanced Research and Reviews*, *17*(3), 1044–1058. <https://doi.org/10.30574/wjarr.2023.17.3.0288>

Dhirani, L. L., Armstrong, E., & Newe, T. (2021). Industrial IoT, Cyber Threats, and Standards Landscape: Evaluation and Roadmap. *Sensors*, *21*(11), 3901. <https://doi.org/10.3390/s21113901>

Elsedimy, E. I., & AboHashish, S. M. M. (2025). An intelligent hybrid approach combining fuzzy C-means and the sperm whale algorithm for cyber attack detection in IoT networks. *Scientific Reports*, *15*(1). <https://doi.org/10.1038/s41598-024-79230-4>

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, *110*, 104743. <https://doi.org/10.1016/j.engappai.2022.104743>

Farouk, M., Sakr, R. H., & Hikal, N. (2024). Identifying the most accurate machine learning classification technique to detect network threats. *Neural Computing and Applications*, *36*(16), 8977–8994. <https://doi.org/10.1007/s00521-024-09562-9>

Fox, J. (2023). *Top Cybersecurity Statistics for 2024*. Www.cobalt.io. <https://www.cobalt.io/blog/cybersecurity-statistics-2024>

Fuchs, M., & Höpken, W. (2022). Clustering: Hierarchical, k-Means, DBSCAN. *Springer EBooks*, 129–149. <https://doi.org/10.1007/978-3-030-88389-8_8>

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. <https://doi.org/10.1016/j.engappai.2022.105151>

Hassan, B. A., & Rashid, T. A. (2021). A multidisciplinary ensemble algorithm for clustering heterogeneous datasets. *Neural Computing and Applications*, *33*(17). <https://doi.org/10.1007/s00521-020-05649-1>

Heidari, J., Daneshpour, N., & Zangeneh, A. (2024). A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers. *Pattern Recognition*, *155*, 110639–110639. <https://doi.org/10.1016/j.patcog.2024.110639>

Huang, S., Kang, Z., Xu, Z., & Liu, Q. (2021). Robust deep k-means: An effective and simple method for data clustering. *Pattern Recognition*, *117*, 107996. <https://doi.org/10.1016/j.patcog.2021.107996>

Ikotun, A. M., Almutari, M. S., & Ezugwu, A. E. (2021). K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Applied Sciences*, *11*(23), 11246. <https://doi.org/10.3390/app112311246>

Jacobson, J. (2023). *Average Cost of a Data Breach Has Reached an All-Time High: IBM Report*. Insurance Journal. <https://www.insurancejournal.com/news/national/2023/08/08/734391.htm>

Jain, M., Kaur, G., & Saxena, V. (2022). A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection. *Expert Systems with Applications*, *193*, 116510. <https://doi.org/10.1016/j.eswa.2022.116510>

Kaliyaperumal, P., Periyasamy, S., Thirumalaisamy, M., Balusamy, B., & Benedetto, F. (2024). A Novel Hybrid Unsupervised Learning Approach for Enhanced Cybersecurity in the IoT. *Future Internet*, *16*(7), 253. <https://doi.org/10.3390/fi16070253>

Kodama, M., & Hashimoto, K. (2024). Research on High-End Disruptive Innovations - Analysis and insights from the video conferencing systems market. *Technology in Society*, *77*, 102576–102576. <https://doi.org/10.1016/j.techsoc.2024.102576>

Kolade, T. M., Obioha-Val, O. A., Balogun, A. Y., Gbadebo, M. O., & Olaniyi, O. O. (2025). AI-Driven Open Source Intelligence in Cyber Defense: A Double-edged Sword for National Security. *Asian Journal of Research in Computer Science*, *18*(1), 133–153. <https://doi.org/10.9734/ajrcos/2025/v18i1554>

Li, J., Liu, P., Chen, L., Pedrycz, W., & Ding, W. (2024). An Integrated Fusion Framework for Ensemble Learning Leveraging Gradient Boosting and Fuzzy Rule-Based Models. *IEEE Transactions on Artificial Intelligence*, 1–15. <https://doi.org/10.1109/tai.2024.3424427>

Lyu, J., Song, A., Seo, E., & Kim, G. (2025). An exploratory analysis of the DPRK cyber threat landscape using publicly available reports. *International Journal of Information Security*, *24*(1). <https://doi.org/10.1007/s10207-025-00980-x>

Mahdi, M. A., Hosny, K. M., & Elhenawy, I. (2021). Scalable Clustering Algorithms for Big Data: A Review. *IEEE Access*, *9*, 80015–80027. <https://doi.org/10.1109/access.2021.3084057>

Mahmoud, M., Chen, W., Yang, Y., & Li, Y. (2024). Automated BIM generation for large-scale indoor complex environments based on deep learning. *Automation in Construction*, *162*, 105376–105376. <https://doi.org/10.1016/j.autcon.2024.105376>

Mahnoor, Shafi, I., Chaudhry, M., Montero, E. C., Alvarado, E. S., De de la Torre Diez, I., Samad, M. A., & Ashraf, I. (2024). A Review of Approaches for Rapid Data Clustering: Challenges, Opportunities and Future Directions. *IEEE Access*, 1–1. <https://doi.org/10.1109/access.2024.3461798>

Messmer, L.-M., Reich, C., & Abdeslam, D. O. (2024). Context-Aware Machine Learning: A Survey. *Lecture Notes in Networks and Systems*, 252–272. <https://doi.org/10.1007/978-3-031-73110-5_17>

Metibemu, O. C., Adesokan-Imran, T. O., Ajayi, A. J., Tiwo, O. J., Olutimehin, A. T., & Olaniyi, O. O. (2025). Developing Proactive Threat Mitigation Strategies for Cloud Misconfiguration Risks in Financial SaaS Applications. *Journal of Engineering Research and Reports*, *27*(3), 393–413. <https://doi.org/10.9734/jerr/2025/v27i31442>

Obioha-Val, O. A. (2025). Bridging Gaps in Cybersecurity Governance: Leveraging Collaborative Digital Solutions. *Asian Journal of Research in Computer Science*, *18*(2), 82–100. <https://doi.org/10.9734/ajrcos/2025/v18i2564>

Obioha-Val, O. A., Gbadebo, M. O., Olaniyi, O. O., Chinye, N. C., & Balogun, A. Y. (2025). Innovative Regulation of Open Source Intelligence and Deepfakes AI in Managing Public Trust. *Journal of Engineering Research and Reports*, *27*(2), 136–156. <https://doi.org/10.9734/jerr/2025/v27i21400>

Obioha-Val, O. A., Lawal, T. I., Olaniyi, O. O., Gbadebo, M. O., & Olisa, A. O. (2025). Investigating the Feasibility and Risks of Leveraging Artificial Intelligence and Open Source Intelligence to Manage Predictive Cyber Threat Models. *Journal of Engineering Research and Reports*, *27*(2), 10–28. <https://doi.org/10.9734/jerr/2025/v27i21390>

Obioha-Val, O. A., Olaniyi, O. O., Gbadebo, M. O., Balogun, A. Y., & Olisa, A. O. (2025). Cyber Espionage in the Age of Artificial Intelligence: A Comparative Study of State-Sponsored Campaign. *Asian Journal of Research in Computer Science*, *18*(1), 184–204. <https://doi.org/10.9734/ajrcos/2025/v18i1557>

Olutimehin, A. T. (2025a). Advancing Cloud Security in Digital Finance: AI-Driven Threat Detection, Cryptographic Solutions, and Privacy Challenges. *Journal of Engineering Research and Reports*, *27*(3), 35–55. <https://doi.org/10.9734/jerr/2025/v27i31416>

Olutimehin, A. T. (2025b). Assessing the Effectiveness of Cybersecurity Frameworks in Mitigating Cyberattacks in the Banking Sector and its Applicability to Decentralized Finance (DeFi). *Asian Journal of Research in Computer Science*, *18*(3), 130–151. <https://doi.org/10.9734/ajrcos/2025/v18i3583>

Olutimehin, A. T. (2025c). The Synergistic Role of Machine Learning, Deep Learning, and Reinforcement Learning in Strengthening Cyber Security Measures for Crypto Currency Platforms. *Asian Journal of Research in Computer Science*, *18*(3), 190–212. <https://doi.org/10.9734/ajrcos/2025/v18i3586>

Olutimehin, A. T., Ajayi, A. J., Metibemu, O. C., Balogun, A. Y., Oladoyinbo, T. O., & Olaniyi, O. O. (2025). Adversarial Threats to AI-Driven Systems: Exploring the Attack Surface of Machine Learning Models and Countermeasures. *Journal of Engineering Research and Reports*, *27*(2), 341–362. <https://doi.org/10.9734/jerr/2025/v27i21413>

Olutimehin, A. T., Joseph, S. A., Ajayi, A. J., Metibemu, O. C., Balogun, A. Y., & Olaniyi, O. O. (2025). Future-Proofing Data: Assessing the Feasibility of Post-Quantum Cryptographic Algorithms to Mitigate “Harvest Now, Decrypt Later” Attacks. *Archives of Current Research International*, *25*(3), 60–80. <https://doi.org/10.9734/acri/2025/v25i31098>

Oyekunle, S. M., Tiwo, O. J., Adesokan-Imran, T. O., Ajayi, A. J., Salako, A. O., & Olaniyi, O. O. (2025). Enhancing Data Resilience in Cloud-based Electronics Health Records through Ransomware Mitigation Strategies Using NIST and MITRE ATT&CK Frameworks. *Journal of Engineering Research and Reports*, *27*(3), 436–457. <https://doi.org/10.9734/jerr/2025/v27i31444>

Petrosyan, A. (2024). *Average Cost of a Data Breach in the United States from 2006 to 2024*. Statista. <https://www.statista.com/statistics/273575/us-average-cost-incurred-by-a-data-breach/>

Purandhar, N., Ayyasamy, S., & Kumar, S. (2022). Strategic real time framework for healthcare using fuzzy C-means systems. *Automated Software Engineering*, *29*(1). <https://doi.org/10.1007/s10515-021-00302-0>

Qi, Y. (2025). MultiModalGraphSearch: Intelligent Massive-Scale SubGraph Discovery for Multi-Category Financial Pattern Mining. *IEEE Access*, *13*, 51813–51831. <https://doi.org/10.1109/access.2025.3553560>

Ramezani, A. (2024). Fusion Models for Cyber Threat Defense: Integrating Clustering with Kmeans, Random Forests, and SVM against Windows Malware. *2024 IEEE World AI IoT Congress (AIIoT)*, 465–470. <https://doi.org/10.1109/aiiot61789.2024.10578947>

Ran, X., Xi, Y., Lu, Y., Wang, X., & Lu, Z. (2022). Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artificial Intelligence Review*, *56*(8). <https://doi.org/10.1007/s10462-022-10366-3>

Rani, N., Saha, B., & Shukla, S. K. (2024). *A Comprehensive Survey of Advanced Persistent Threat Attribution: Taxonomy, Methods, Challenges and Open Research Problems*. ArXiv.org. <https://arxiv.org/abs/2409.11415>

Rodriguez, P. C., Marti-Puig, P., Caiafa, C. F., Serra-Serra, M., Cusidó, J., & Solé-Casals, J. (2023). Exploratory Analysis of SCADA Data from Wind Turbines Using the K-Means Clustering Algorithm for Predictive Maintenance Purposes. *Machines*, *11*(2), 270–270. <https://doi.org/10.3390/machines11020270>

Saha, A., Blasco, J., Cavallaro, L., & Lindorfer, M. (2024). ADAPT it! Automating APT Campaign and Group Attribution by Leveraging and Linking Heterogeneous Files. *RAID ’24: Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses Pages* , *24*, 114–129. <https://doi.org/10.1145/3678890.3678909>

Salako, A. O., Adesokan-Imran, T. O., Tiwo, O. J., Metibemu, O. C., Onyenaucheya, O. S., & Olaniyi, O. O. (2025). Securing Confidentiality in Distributed Ledger Systems with Secure Multi-party Computation for Financial Data Protection. *Journal of Engineering Research and Reports*, *27*(3), 352–373. <https://doi.org/10.9734/jerr/2025/v27i31439>

Salami, I. A., Adesokan-Imran, T. O., Tiwo, O. J., Metibemu, O. C., Olutimehin, A. T., & Olaniyi, O. O. (2025). Addressing Bias and Data Privacy Concerns in AI-Driven Credit Scoring Systems Through Cybersecurity Risk Assessment. *Asian Journal of Research in Computer Science*, *18*(4), 59–82. <https://doi.org/10.9734/ajrcos/2025/v18i4608>

Shahraki, A., Abbasi, M., Taherkordi, A., & Jurcut, A. D. (2022). A comparative study on online machine learning techniques for network traffic streams analysis. *Computer Networks*, *207*, 108836. <https://doi.org/10.1016/j.comnet.2022.108836>

Shoaib, H. A., Rahman, M. A., Maua, J., Rahman, A., Mridha, M. F., Kim, P., & Shin, J. (2025). An Enhanced Deep Learning Approach to Potential Purchaser Prediction: AutoGluon Ensembles for Cross-Industry Profit Maximization. *IEEE Open Journal of the Computer Society*, 1–12. <https://doi.org/10.1109/ojcs.2025.3552376>

Steingartner, W., Galinec, D., & Kozina, A. (2021). Threat Defense: Cyber Deception Approach and Education for Resilience in Hybrid Threats Model. *Symmetry*, *13*(4), 597. <https://doi.org/10.3390/sym13040597>

Stoddart, K. (2022). Non and Sub-State Actors: Cybercrime, Terrorism, and Hackers. *Cyberwarfare*, 351–399. <https://doi.org/10.1007/978-3-030-97299-8_6>

Sufi, F., & Alsulami, M. (2025). Mathematical Modeling and Clustering Framework for Cyber Threat Analysis Across Industries. *Mathematics*, *13*(4), 655. <https://doi.org/10.3390/math13040655>

Tiwo, O. J., Adesokan-Imran, T. O., Babarinde, D. C., Oyekunle, S. M., Olutimehin, A. T., & Olaniyi, O. O. (2025). Advancing Security in Cloud-based Patient Information Systems with Quantum-resistant Encryption for Healthcare Data. *Asian Journal of Research in Computer Science*, *18*(4), 187–208. <https://doi.org/10.9734/ajrcos/2025/v18i4615>

Tiwo, O. J., Adesokan-Imran, T. O., Babarinde, D. C., Salami, I. A., Onyenaucheya, O. S., & Olaniyi, O. O. (2025). Improving Patient Data Privacy and Authentication Protocols against AI-Powered Phishing Attacks in Telemedicine. *Asian Journal of Research in Computer Science*, *18*(4), 93–114. <https://doi.org/10.9734/ajrcos/2025/v18i4610>

Wang, J., Sun, Q., & Zhou, C. (2023). Insider Threat Detection Based on Deep Clustering of Multi-Source Behavioral Events. *Applied Sciences*, *13*(24), 13021. <https://doi.org/10.3390/app132413021>

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, *10*, e2286–e2286. <https://doi.org/10.7717/peerj-cs.2286>

Warikoo, A. (2021). The Triangle Model for Cyber Threat Attribution. *Journal of Cyber Security Technology*, *5*(3-4), 1–18. <https://doi.org/10.1080/23742917.2021.1895532>

Xiao, N., Lang, B., Wang, T., & Chen, Y. (2024). APT-MMF: An advanced persistent threat actor attribution method based on multimodal and multilevel feature fusion. *Computers & Security*, *144*, 103960–103960. <https://doi.org/10.1016/j.cose.2024.103960>

Yazdi, M. (2024). Reliability-Centered Design and System Resilience. *Springer Series in Reliability Engineering*, 79–103. <https://doi.org/10.1007/978-3-031-53514-7_5>

Yu, B., Zheng, Z., & Dai, J. (2023). K-DGHC: A hierarchical clustering method based on K-dominance granularity. *Information Sciences*, *632*, 232–251. <https://doi.org/10.1016/j.ins.2023.03.012>