Journal of Human Security | 2025 | Volume 21 | Issue 1 | Pages 1-6
DOI: https://doi.org/10.12924/johs2025.210101
ISSN: 1835–3800

Journal of Human Security

Research Article

Cyber Intelligence in Counterterrorism: AI-Powered Detection of Encrypted Jihadist Threats

Ala Berzinji [1,2,3]

[1]Department of Computer Science, College of Science, University of Sulaimani, Sulaimani, Iraq
[2]Stockholm University, Department of Computer and System Sciences
[3]Cyber Security Department, IQ Group Holding, Sulaymaniyah 46001, Iraq, Email: alabe@dsv.su.se

***Correspondence:** alabe@dsv.su.se

Submitted: 10 January 2025 | Revised: 20 April 2025 | Accepted: 01 June 2025 | Published: 30 June 2025

Abstract: Encryption offers a contradictory challenge of securely enabling communication yet allowing terrorist organizations, and particularly the jihadist organizations, to evade international counterterrorist surveillance systems. Such groups make extensive use of encrypted technologies for coordination of actions, sharing of propaganda, and illegal activities to avoid regular intelligence and cybersecurity standards. This paper discusses the innovative artificial intelligence (AI) and machine learning (ML) techniques for identifying and deciphering encrypted communications employed by the jihadist networks. A real-time analysis framework is proposed that uses deep learning models, unsupervised learning methods, and tools of natural language processing (NLP). To ensure replicability and practical implementation, a generalised algorithmic structure accompanied by pseudocode is included. The proposed system is evaluated using datasets derived from simulations, authentic scenarios, and extremist digital platforms, annotated through manual and automated methods. Comprehensive experimental results indicate that hybrid ML models can reliably flag suspicious communication patterns using only metadata, packet dimensions, and traffic flow characteristics, thus obviating the necessity for content access. The findings highlight AI's capacity to furnish intelligence agencies with novel capabilities for the proactive identification of encrypted terrorist activities, thereby reinforcing counterterrorism operations. The study further stresses the importance of continuous model refinement and ethical governance for responsible and effective deployment.

Keywords: Machine Learning, AI, Cybersecurity, Encrypted Communications, Jihadist Networks, Counterterrorism, Deep Learning, Natural Language Processing, Anomaly Detection, Encryption Analysis

1. Introduction

The demand for a strong cybersecurity has become a pillar of modern-day digital society with diverse aspects of daily life, government, and business increasingly gravitating toward an internet-based system. Since the Internet is the central infrastructure for communications systems, business transactions, and administrative activity, protecting the digital world needs to be taken seriously. Not just for ensuring the privacy of individual data but for keeping sensitive information as well as the operational integrity of the economies of nations safe, cyberspace protection is essential. The growing significance of cyberspace demands the convergence of comprehensive and articulated cybersecurity policies that will neutralize advanced threats [1].

Parallel to that is an increased need for awareness in cyberspace, with individuals and organizations needing more information on digital protection methods. In a bid to neutralize potential threats, the users need to get accustomed to threats like data breach and focused cyber-attacks and put in place equally efficient protective methods as well. Moreover, adopting such protective methods requires good knowledge of the offensive methods employed by the attackers that misuse digital platforms for terrorist attacks [2]. Online extremism has emerged as a significant concern, with terrorist entities leveraging digital tools to disseminate violent ideologies and recruit operatives for coordinated assaults. Jihadist organisations exploit the anonymity and global reach of the Internet to spread extremist propaganda and orchestrate operational activities. This misuse of digital technologies by extremist factions necessitates the prompt formulation of targeted countermeasures, as the urgency for effective responses has never been more acute [3]. The biggest challenge for the purposes of the interception and analysis of terrorist communications is the employment by terrorist organizations of advanced encryption technologies. These players are effectively evading counter-surveillance by using encrypted communications platforms, encrypted video communications services, encrypted email services, and virtual private networks (VPNs). While personal privacy is being safeguarded by encryption, it is simultaneously creating serious problems for the law-enforcement and intelligence agencies that are expected to intercept and analyse terrorist communications.

Such agencies face considerable difficulties in monitoring covert activities, particularly as criminal elements exploit encryption to obscure their conduct on digital platforms, including those operating within the dark web [4; 5]. Technologically adept jihadist cells have demonstrated proficiency in employing encryption and anonymisation tools to shield their digital footprints. The enormous volume of encrypted data, compounded by the multiplicity of cryptographic methods and the rapid evolution of encryption technologies, creates complex conditions for surveillance and counterterrorism initiatives. The very encryption mechanisms intended to secure intelligence communications now pose considerable impediments to tracking terrorist behaviours, deciphering encrypted content, and disrupting plots before execution [6]. ML has emerged as a promising avenue to address these growing technological and security-related challenges. ML algorithms can detect anomalous patterns in encrypted datasets and classifying potentially terrorism-related communications. Their ability to uncover hidden irregularities and systematically categorise encrypted exchanges endows intelligence entities with powerful analytical tools. This study explores ML-based techniques for the detection of jihadist encrypted communications, evaluating their effectiveness and assessing their potential utility in supporting law enforcement and intelligence counterterrorism operations [7; 8].

Section 2 presents a literature review of previous research addressing encrypted communication detection strategies. Section 3 identifies key challenges impeding the detection of terrorist exchanges facilitated through encryption. Section 4 outlines the study's methodology, detailing the algorithmic system and rationale behind the selected techniques. The model's performance evaluation is reported in Section 5, followed by a discussion of recommendations and concluding remarks in Section 6.

2. Literature Review

The detection of encrypted terrorist communications has been widely studied, with studies including traditional and machine learning-based methods. The section gives an overview of recent developments within this field. Traditional methods for encrypted traffic detection often rely upon the identification of known traffic patterns or learned behaviour. However, such techniques are becoming less effective when faced with terrorist organizations adopting sophisticated encryption techniques. Therefore, there has been a strong trend towards using ML methods for overcoming the inherent weaknesses of previous approaches.

In [9], the authors employed a deep learning approach using CNNs, drawing upon the foundational work in [10], to detect encrypted network traffic. Their findings demonstrated that CNNs, when trained on features such as packet sizes and communication durations, were capable of distinguishing between benign and potentially malicious encrypted data flows. The study in [10] further investigated using NLP techniques for analyzing encrypted jihadist communications. The authors suggested that NLP would be able uncover latent semantic patterns within encrypted messages, and their linguistics-based methodology indicated potential for detecting covert communications about extremist activity. In turn, [11] used an unsupervised anomaly detection approach to analyze encrypted communication patterns above the physical layer. The approach was applied to learning models that do not require labelled data so that any type of behaviour anomaly matching terrorist communication patterns could be detected.

Cyber intelligence is a key element of counterterrorist tactics, specifically in the context of identifying and responding to encrypted jihadist threats. Both AI and ML have become groundbreaking tools in this context, providing the ability to handle large datasets and find patterns that signal potential danger [12; 13]. Different techniques using AI have been proposed for the identification of encrypted extremist threats. For instance, the INSPECT framework [14] combines ML techniques with behaviour analysis and graph pattern matching in identifying risk profiles and suspect networks. It has been proven with respect to dynamic datasets of domestic jihadist behaviour. Another method is the use of dynamic systems for the identification of unsupervised anomalies, that is valuable for the recognition of emerging threats with respect to unlabelled as well as temporally correlated datasets [15]. This methodology has shown efficacy in detecting indicators of compromise within Dark Web forum activity. AI and ML are also employed in the analysis of Dark Web environments to generate cyber threat intelligence (CTI), extracting meaningful insights by identifying anomalous or recurring patterns [15]. Through the use of predictive analytics, these technologies can support a proactive stance in cybersecurity, offering early detection of threat indicators [16].

Advanced technologies, including ML, deep learning (DL), and NLP, enable AI to process structured and unstructured data sources, discovering potential threats by detecting anomalous behaviours and patterns. However, implementing AI within cybersecurity ecosystems is challenging. These challenges include making available and ensuring the quality of data used for AI training, facilitating transparency within AI decision-making processes, and integrating AI platforms with existing security platforms. AI models need to be constantly updated because of the dynamic evolution of cyber threats [17]. Moreover, AI-enabled platforms can incorporate predictive analytics to evaluate threat trajectories and forecast future incidents [18]. Such systems are capable of autonomously detecting and responding to threats by processing vast datasets and identifying threat-relevant patterns [19]. AI and ML are also important for minimizing false positives, refining discrimination among legitimate threats versus suspicious anomalies. AI and ML tools automate threat prioritization and alleviate the workload for security staff by accurately filtering out normal versus malicious behaviour [20; 21]. In conclusion, AI and ML hold great potential for redefining detection and mitigation of cyber-attacks, making counterterrorism operations more effective. However, realising this potential requires addressing the technical and ethical challenges associated with their implementation in cybersecurity frameworks [22; 23].

2.1 Comparative Analysis of Prior Studies

Recent studies have explored a variety of ML and DL methods for detecting encrypted communications associated with terrorist activities. In spite of a majority of such models being satisfactory from an accuracy point, their performance and usability are quite different based on context. For example, studies in [24; 25] employed deep learning methods for traffic classification and the detection of previously unseen data. While their systems achieved high recall rates, they were not designed to function efficiently in real-time environments. In contrast, the work in [26] concentrated on identifying malicious encrypted traffic through feature mining, achieving high precision; however, the model's dependence on large volumes of labelled data restricts its scalability and broader applicability. Research presented in [27; 28; 29; 30] focused on NLP-driven approaches using social media content, identifying linguistic markers associated with radicalisation. These methods proved effective for early behavioural detection but showed limited capability in analysing encrypted data at the binary level, where no semantic information is accessible. Concerns regarding the ethical implications of false positives in automated surveillance systems, particularly within high-volume monitoring contexts, were raised in [8]. This observation aligns with the present study's methodology, which integrates anomaly detection with classification to minimise bias while maintaining high recall.

In contrast to earlier studies, this research presents the following key distinctions:

- It integrates both supervised and unsupervised learning within a unified system.
- It focuses specifically on encrypted binary traffic rather than textual data from social media platforms.
- It proposes a real-time operational framework, as opposed to a static analytical approach.

The combination of pattern-based traffic analysis with behaviour-driven feature extraction represents a novel contribution, linking technical performance with practical counterterrorism applications.

2.2 Problem Definition

Two primary challenges arise in the detection of encrypted jihadist communications across communication networks. These difficulties are highlighted in [31].

Problem 1: The detection of encrypted traffic is challenging due to the difficulty in accurately distinguishing between legitimate and malicious traffic within encrypted data streams.

Problem 2: Classifying Suspicious Communications [32].

The process of identifying encrypted traffic involves two successive steps towards assessing the terrorist affiliation of the communications. Models must analyse encrypted data by detecting abnormal patterns, alongside linguistic and behavioural features that indicate harmful intentions. The study seeks to develop robust machine learning systems capable of instant analysis of encrypted information, as these challenges necessitate such advancements [33]. In their current research, the authors explored how AI-powered technologies function in the context of counterterrorism efforts, addressing both operational and ethical concerns. Their research highlighted the significance of combining detection mechanisms with legal and ethical oversight, especially when AI technology is used for surveillance and threat identification. The paper presents an elaborate approach for ensuring responsible AI application for counterterror measures that are democratic and in line with human rights. Research in [34] explored AI-counterterror dynamics with a focus on technical and ethics-oriented analysis. The paper highlighted major flaws with AI deployment for surveillance and threat identification, particularly within sophisticated cybersecurity. In addition, it highlighted the importance of ethical responsibility when using AI tools for surveillance of extremism-related encrypted communications.

3. Methodology

This paper outlines a strong combination of supervised and unsupervised machine learning models for effective identification of terrorist-related encrypted communications. The detection mechanism applies this methodology to multiple sequential steps, which enhance both its working performance and flexibility. A comprehensive explanation of such methodological steps is given in the next section [34].

3.1 Data Collection and Pre-Processing

The methodology starts with the extensive gathering of varied encrypted communication data being its first step. The detection model draws information from published encrypted traffic datasets, supplemented by simulated encrypted traffic originating from terrorist networks and real-life network data. Original and simulated databases collectively are used, subjecting the model to a vast universe of terrorist group communication patterns and adopted encryption methods [35]. After collection of the data is complete, the pre-processing phase is initiated. Several preparatory tasks of utmost importance are performed on the data before progressing towards feature extraction and modelling. Pre-processing starts with noise reduction with a view to removing irrelevant data points and preserving encrypted traffic of relevance for analysis. Feature engineering is subsequently conducted in an effort to derive meaningful features from raw data such as packet size, transmission times, flow behaviours of destination-source pairs, and

frequency of communication [36]. These features are selected with utmost care on the basis of how they capture behaviour patterns that have the potential to differentiate the analysis of benign traffic from terrorist communications.

3.1.1 Feature Extraction

Feature: A crucial element for the success of the Model is feature extraction, which involves deriving distinctive characteristics from raw encrypted data. These extracted features serve as the foundation for both supervised classification methods and unsupervised anomaly detection.

Packet Size Variations: The size of encrypted packets provides key indicators that reflect the operational methods and communication strategies employed by entities [26].

Inter-packet Time Intervals: The traffic patterns exhibit deviations from normal behaviour, revealing critical communication details through the time intervals between packets.

Flow Patterns: Communication sessions follow specific flow patterns, alternating between periods of intense activity and inactivity. Firewalls can identify these patterns, helping to uncover the nature of the traffic [37].

Statistical Analysis of Encrypted Message Lengths: An analysis of statistical features of encrypted message lengths allows for detection of terrorist coordination based on scheduled communication patterns. Statistical patterns within message length, when taken over sessions and communication paths, tend to indicate synchronized operations, which are indicative of planned operations. For example, brief, repeated messages tend to represent the delivery of commands, while longer, erratic packets tend to indicate elaborate logistical planning. These patterns are particularly important, especially when explicit decryption of content is precluded by strong encryption techniques.

The features derived provide a wealth of data, allowing the model to discern normal communication profiles from those affected by malicious behaviour. Average packet size, packet size variability and periodicity of transmission are key features that facilitate the creation of distinct behaviour profiles. This assists the system in differentiating normal encrypted communications from communications with terrorist-related activity indications. Furthermore, the flow-based features (time intervals between messages, sender-receiving pairs, and bidirectionality) offer a further inferential layer for the model. These features depend on suspicious parties employing encrypted networks for long-term planning. With geospatial metadata and frequency analysis added into the equation, features facilitate anomaly detection even with decrypted text missing. Domain expertise and the employment of advanced statistical techniques offer a vital input for the extraction of features such that the data remains pertinent and complete. This enables the learning model to generalise properly across disparate crypto schemes and trending traffic patterns and maintain high-quality accuracy for both classification and anomaly recognition. Strong feature engineering is a prime ingredient for self-sufficient encrypted menace detection platforms [38].

3.1.2 Model Training

The methodology focuses on training machine learning models for classification and anomaly detection, which is divided into two key components: supervised learning and unsupervised learning [39].

Supervised Classification: This step trains different classification models on a labeled dataset with examples of encrypted communications (benign or terrorist-related). The aim is to instruct the model to learn patterns that distinguish non-hostile communications from communications related to terrorism. A few such classification methods that are used for comparisons are:

Random Forest: A machine learning method that can deal with high-dimensional datasets and complex features with superior generalization performance.

Support Vector Machines (SVM): A robust classification technique that is good for high dimensional spaces and ideal when there are a large number of features.

Deep Neural Networks (DNN): Deep learning models are investigated to find non-linear relationships in the data that might enhance classification performance.

Unsupervised learning methods detect unusual traffic patterns that reflect terrorist activities. Unsupervised learning methods don't rely on labelled information and aim to find instances of deviation from normal traffic behaviour. Some of the key methods involved

K-Means Clustering: It clusters similar instances and marks any instances that cannot be categorized into pre-defined clusters as points of anomaly.

Autoencoders: These neural networks act as compression and reconstruction systems. Anomalies are identified by observing the reconstruction error, which when greater than some threshold indicates abnormal traffic patterns.

Both supervised and unsupervised models undergo iterative training, refining their parameters to improve both classification accuracy and anomaly detection capabilities.

3.1.3 Model Evaluation

A comprehensive evaluation framework ensures the quality and reliability of the trained models. Multiple assessment techniques are employed to evaluate the performance of the model, focusing on accuracy and reliability. The primary metrics include:

Accuracy: The percentage of correctly classified instances, encompassing both benign and terrorist-related communications.

Precision: Measures the model's ability to correctly identify true terrorist-related instances, thereby reducing false positives.

Recall: Represents the proportion of actual terrorist-related communications correctly identified by the model, aiming to reduce false negatives.

F1-Score: A metric that uses the harmonic mean to balance precision and recall, providing a unified performance measure.

Cross-validation techniques are applied to ensure the model's robustness and to confirm its ability to generalise effectively. This evaluation process helps assess the model's capability to predict unseen data while mitigating the risk of overfitting, thus ensuring its successful deployment for encrypted traffic detection.

3.2 System Architecture and Technical Implementation

The implementation of the detection framework utilizes a multi-component architecture designed for scalable and real-time traffic processing.

3.3 Programming Environment and Tools

The system was developed using Python 3.11. Core libraries include:

- Scikit-learn for traditional ML models (Random Forest, SVM)
- TensorFlow for DNNs and autoencoders
- Pandas and NumPy for pre-processing
- Wireshark and PyShark for traffic capture and analysis

3.4 Data Pipeline

Encrypted traffic is collected through virtual network environments. Following feature extraction, a central processing module directs the extracted features to the appropriate model pipeline for further analysis.

3.5 Deployment Setup

The architecture is containerized using Docker and deployed on a Linux-based cloud server (Ubuntu 22.04), with GPU acceleration to support DNN models. Real-time detection is facilitated through:

- Queue-based data streaming (using Apache Kafka)
- API-triggered classification modules
- Immediate alerting to SIEM systems used by intelligence agencies

3.6 System Performance

The framework ensures sub-second latency for classification and keeps anomaly detection time under three seconds, enabling near-instantaneous identification of encrypted terrorist traffic.

3.6.1 Real-Time Detection Framework

The functioning of real-time detection systems demands the deployment of fully trained, optimized, and tested machine learning models. The system runs continuously scanning the arriving encrypted traffic without disrupting and classifying and detecting the anomalies in real-time. Running with high-traffic capacities, the system offers prompt responses and true detections of suspicious communications. With the recognition of terrorist-related communications, the system instantly raises an alarm enabling action from the law enforcement agencies in real-time. With the passage of time, the predictive capacity of the system just keeps growing and extracting the intelligence from the sheer amount of encrypted traffic without affecting the pace of the operation. The real-time detection system runs continuously scanning terrorist-related communications and keeping an eye on any other forms of malicious activities.

The system notifies and alerts responsible organizations with timely actionable information. The alarms get seamlessly integrated into process operations with pre-defined policies enabling immediate assessment, scoring, and mitigation of cyber threats. The alarms provide metadata aggregates like correlation-count fields, timestamps, packet-length profiles, and scoring for anomaly scores, providing context for the likelihood and severity of a possible threat. The system has a built-in feedback loop for false-positive identification as well as validation of true threats. Feedback is cycled back into the learning process in order to enhance the precision of the detections as well as class definition. The system thus becomes more proficient with time in detecting threats, particularly in rapidly changing environments.

One of the key advantages of the system is that it can maintain processing speed even when it is increased for increased traffic. The pipeline of the system applies parallel processing, model pruning, and replastering techniques for low-latency computation even during high-throughput traffic. This enables the system to continue being responsive towards arriving data, whether a high-traffic load or a low-traffic load. In addition, there is auditing and logging support, facilitating retrospective analysis and maintaining regulatory compliance. By retaining traffic fingerprints and alert histories, the system allows historical communication patterns to be cross-referenced across incidents, enhancing overall threat analysis. Ultimately, this real-time detection system serves as a vital monitoring tool and a dynamic intelligence asset, continually adapting to safeguard security operations against encrypted, covert terrorist communications in a scalable, repeatable, and auditable manner.

3.6.2 Proposed Algorithm and Pseudocode

The following pseudocode outlines the steps involved in detecting encrypted terrorist communications.

Algorithm: Detecting Encrypted Terrorist Communications

Input:

- Encrypted communication dataset (D)
- Pre-trained machine learning model (Model)

Output:

- Label (Terrorist/Non-Terrorist)

Steps:

1. Pre-Process Data D:
   - Extract relevant features (packet size, flow patterns, etc.)
   - Normalize features
2. Apply Feature Selection:
   - Select key features indicative of terrorist activity (e.g., frequency of communication)
3. Split Dataset D into Training and Test Sets (80%-20%):
   - Training set: Used to train the model
   - Test set: Used to evaluate the Model's performance
4. Train the Machine Learning Model (Model) using the Training Set:
   - Apply algorithms like Random Forest, SVM, or Deep Learning
5. Evaluate the Trained Model on the Test Set:
   - Use accuracy, precision, recall, and F1-score to assess performance
6. Use the Trained Model to Classify New Encrypted Communication:
   - If the Model predicts "Terrorist", flag it as suspicious
   - Otherwise, classify as non-terrorist

4. Results and Discussion

The encrypted communications utilised by jihadist networks present substantial challenges for counterterrorism efforts. The very nature of encryption, when employed by adversaries, complicates the ability of traditional intelligence agencies to effectively intercept or monitor these communications. In this context, machine learning offers a potential solution, providing a means to identify patterns and detect anomalies in encrypted data that may indicate terrorist activity. The proposed approach employs supervised learning for classification tasks alongside unsupervised learning for detecting unusual patterns that may signal terrorist communications. The supervised learning component uses labelled samples to train the model to differentiate between terrorist and non-terrorist content. However, the designed approach also incorporates unsupervised learning to identify threats that may not be present in the training data. The integrated strategy increases the system's strength and versatility. Through the analysis of indicators of packet size, flow patterns, and communication frequency within a filter system, the model adds intelligence that increases the effectiveness of counterterrorist operations. The technology shows potential for applicability across all manner of covert communications and is thus a highly valuable tool for counterterrorist intelligence and security. Each model was assessed using the same dataset and validated through five-fold cross-validation. Key results are summarised in Table 1.

Table 1: Performance Comparison of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	91.3%	89.5%	90.8%	90.1%
SVM	88.9%	86.4%	88.1%	87.2%
Deep Neural Net	93.5%	91.7%	92.4%	92.0%
Autoencoder	–	–	89.2%	–

5. Discussion

- DNNs demonstrated superior accuracy, but they required greater computational resources and longer training times.
- Random Forest provided rapid execution and interpretability, making it well-suited for field deployment.
- Autoencoders effectively detected previously unseen patterns, although they exhibited higher false-positive rates when training data was limited.

6. Limitations

- Unsupervised methods are highly sensitive to threshold tuning.
- Model performance decreased by 7–10% when applied to entirely new datasets featuring novel encryption schemes.
- The absence of labelled terrorist traffic data continues to pose a significant challenge for generalisation.

7. Future Work

The proposed hybrid machine learning system shows promising results, but there are several opportunities for further research and improvement. The approach should be expanded to incorporate additional machine learning techniques, such as deep learning and reinforcement learning, to enhance its ability to detect advanced communication patterns. The research would benefit from broadening the dataset by including encrypted communication examples using various encryption algorithms, thereby improving the model's effectiveness across different scenarios. This would enable the system to better adapt to evolving and emerging terrorist communication methods. A key area for future work is integrating the model with workplace surveillance systems for on-site deployment. By integrating the model directly with intelligence agencies, it would trigger immediate notifications for prompt security action. AI functionalities should augment, not replace, human specialists. They should augment professional expertise by infusing real-time monitoring with manual approbation steps for ensuring authentic detection and low false alarms during surveillance. Along with addressing bandwidth, load, and other resource constraints (such as the surveillance vs. freedom dilemmas and the security vs. privacy trade-offs), more work needs to address ethical issues of delegating surveillance tasks to machine learning machines. For such technologies to deploy responsibly, an open and well-defined ethical framework needs to be practiced and maintained throughout deployment.

7.1 Security, Privacy, and Societal Implications

The integration of machine learning in counterterrorism brings both benefits and risks. Although AI enhances detection accuracy, it also raises substantial ethical, legal, and societal concerns.

7.2 Surveillance and Privacy

Though machine learning brings significant advantages for counterterrorist purposes, it is entangled with positive as well as negative risks. Despite significant performance and capacity improvement across a range of uses such as the detection of encryption threats like anonymization, AI is coupled with extra ethical, legal, and social problems requiring resolution. Finding a suitable balance among national security demand, individual privacy, and generally encrypted communications is a primary challenge. In many cases, machines would handle non-terrorism messages and doing that would intrude on privacy rights entitled under the nation's constitution or conventions. Though the content is kept encrypted, the attached metadata such as the frequency of communications, an individual's digital movement positions, contacts, and session times can still be obtained, with an invasive behaviour pattern eroding trust in ICT regulation. There is an added risk of algorithmic bias as well. Machine learning models trained with skewed or inadequately representative databases will disproportionately focus on a particular group or geographic area. It is not merely unequal but potentially induces social isolation, aggravates tensions, and becomes a new source of long-term threats. Finally, AI's lack of transparency adds

issues for legal accountability. Even though detection rates with machine learning can be slid using a dial analogous to a thermostat, the so-called "black box" character of such systems, especially deep learning [7], makes it hard for judges, authorities, or institutions to know what leads a detection alert. Such a lack of transparency hinders judicial scrutiny, due process, and oversight by institutions, making them difficult, if not impossible, to execute.

7.3 Bias and False Positives

Unbalanced training datasets can cause AI models to over-predict certain language groups, geographical areas, or communication patterns, resulting in algorithmic profiling. When training data overrepresents certain examples of threats—usually taken from past investigations against limited areas or groups—the model learns biased presuppositions about what suspicious behavioral patterns look like. In turn, legitimate encrypted communications within groups are then falsely identified as threats. False positives misdirect investigation efforts and cause severe harm to innocent people. False positives can occur in the form of unwarranted surveillance, digital flagging in security databases, or even travel bans, prison sentences, or reputational damage. It destroys public trust and generates apprehensions regarding possible misuse of the national security programs. Repeated false positives legally curtail principles of necessity and proportionality if there is no accessible avenue for recourse or appeal. Repeated occurrence of unfair classification can moreover drive vulnerable communities away from the hands of the security agencies and interfere with intelligence collection activities and decrease long-term societal acceptance of extremism. To stem it, datasets should carefully be reaped such that they become diversified, contextually meaningful, and representative. Bias correction techniques, adversarial learning, and explainability techniques should be used so that the models don't perpetuate bias. Unmonitored civil society oversight of the input and the training process will maintain fairness along with preserving the effectiveness of the detections. Bias reduction within counterterror activities is a technical imperative but a vital ethical as well as an operational imperative.

7.4 Operational Dependence

Over-reliance on automated systems may undermine human oversight. AI should complement, rather than replace, human analysts, who contribute essential contextual reasoning to the evaluation of threats.

7.5 Policy Recommendations

- AI systems must be transparent, auditable, and open to external review.
- Human-in-the-loop models should be integrated into all surveillance workflows.
- Legal frameworks must strike a balance between counterterrorism objectives and civil liberties.

This research emphasises that ethical deployment is a strategic imperative, rather than a secondary consideration.

8. Conclusion

Jihadist networks employ encryption methods that obstruct counterterror efforts because they allow members to hide communications from conventional intelligence detection techniques. In this paper, a double machine-learning model is proposed for detecting blockchain-based encryption applied within Jihadist network systems. The proposed method shows promise as it incorporates both supervised and unsupervised learning techniques. These machine learning modules allow for the timely detection of suspicious communications, helping intelligence agencies thwart terrorist operations. Intelligence agencies now have the capability to identify threats before attacks occur through machine learning and encryption detection technology, marking a critical advancement in counterterrorism efforts. Future developments of this approach hold the potential to significantly enhance security operations by providing a flexible system capable of adapting to changes in terrorist organizational strategies. However, this strategy must also address the ethical and organizational implications of implementing AI systems in surveillance environments.

References

[1] Malik W, & Gul S. Bridging the Gap: Exploring the Intersection of Cybersecurity and Human Security in the Digital Age. Competitive Research Journal Archive. 2024;2(04):195-202 https://thecrja.com/index.php/Journal/article/view/49

[2] Sabillon R, & Bermejo Higuera JR. The Importance of Cybersecurity Awareness Training in the Aviation Industry for Early Detection of Cyberthreats and Vulnerabilities. Lecture Notes in Computer Science: Springer Nature Switzerland; 2023. p. 461-479.10.1007/978-3-031-48057-7_29. http://dx.doi.org/10.1007/978-3-031-48057-7_29

[3] Antonova EY. Terrorist Crimes in the Era of Digitalization: Forms of Activity and Measures for Counteraction. Journal of Digital Technologies and Law. 2023;1(1):251-269. Doi: https://doi.org/10.21202/jdtl.2023.10

[4] Bridgelall R. An Application of Natural Language Processing to Classify What Terrorists Say They Want. Social Sciences. 2022;11(1):23. Doi: https://doi.org/10.3390/socsci11010023

[5] Cohen K, Johansson F, Kaati L, & Mork JC. Detecting Linguistic Markers for Radical Violence in Social Media. Terrorism and Political Violence. 2013;26(1):246-256. Doi: https://doi.org/10.1080/09546553.2014.849948

[6] Dave D, Sawhney G, Aggarwal P, Silswal N, & Khut D. The New Frontier of Cybersecurity: Emerging Threats and Innovations. 2023 29th International Conference on Telecommunications (ICT); 2023/11/08: IEEE; 2023. p. 1-6.10.1109/ict60153.2023.10374044. http://dx.doi.org/10.1109/ict60153.2023.10374044

[7] Esmailzadeh Y, & Motaghi E. International Terrorism and Social Threats of Artificial Intelligence. Journal of Globalization Studies. 2024;15(1):168-179. Doi: https://doi.org/10.30884/jogs/2024.01.09

[8] Fernandez M, & Alani H. Artificial intelligence and online extremism. Predictive Policing and Artificial Intelligence: Routledge; 2021. p. 132-162.http://dx.doi.org/10.4324/9780429265365-7

[9] Guna WJA. Kufal Symbiosis: Collaboration of Artificial Intelligence and Terrorist Organizations. Security Intelligence Terrorism Journal (SITJ). 2025;2(1):20-26. Doi: https://doi.org/10.70710/sitj.v2i1.28

[10] Irfan M, Almeshal ZA, & Anwar M. Unleashing transformative potential of artificial intelligence (AI) in countering terrorism online radicalisation extremism and possible recruitment. 2024. Doi: https://doi.org/10.34961/researchrepository-ul.25451590.v1

[11] Kaur H. The Evolution Of Terrorism In Digital Age: Cyber Jihad And Emerging Threats. International Journal of Multidiciplinary Education Research. 2025;14(1(3)) https://ijmer.s3.amazonaws.com/pdf/volume14/volume14-issue1(3)/3.pdf

[12] Montasari R. Exploring the Current Landscape of Cyberterrorism: Insights, Strategies, and the Impact of COVID-19. Advanced Sciences and Technologies for Security Applications: Springer International Publishing; 2024. p. 65-90. http://dx.doi.org/10.1007/978-3-031-50454-9_5

[13] Trifunović D. Cybersecurity–virtual space as an area for covert terrorist activities of radical islamists. Teme-Časopis za Društvene Nauke. 2021;45(1):95-109 https://www.ceeol.com/search/article-detail?id=949985

[14] Muramudalige SR, Hung BW, Libretti R, Klausen J, & Jayasumana AP. Investigative Pattern Detection Framework for Counterterrorism. arXiv preprint arXiv:2310.19211. 2023. Doi: https://doi.org/10.48550/arXiv.2310.19211

[15] Sangher KS, Singh A, Pandey HM, & Kumar V. Towards Safe Cyber Practices: Developing a Proactive Cyber-Threat Intelligence System for Dark Web Forum Content by Identifying Cybercrimes. Information. 2023;14(6):349. Doi: https://doi.org/10.3390/info14060349

[16] Jimmy FNU. The Role of Artificial Intelligence in Predicting Cyber Threats. International Journal of Scientific Research and Management (IJSRM). 2023;11(08):935-953. Doi: https://doi.org/10.18535/ijsrm/v11i08.ec04

[17] Danish M. Enhancing Cyber Security through Predictive Analytics: Real-Time Threat Detection and Response. arXiv preprint arXiv:2407.10864. 2024. Doi: https://doi.org/10.48550/arXiv.2407.10864

[18] Ofili BT, Obasuyi OT, & Osaruwenese E. Threat intelligence and predictive analytics in USA cloud security: mitigating AI-driven cyber threats. International Journal of Engineering Technology Research & Management. 2024;8(11):631. Doi: https://doi.org/10.5281/zenodo.14991864

[19] Ahmetoglu H, & Das R. A comprehensive review on detection of cyber-attacks: Data sets, methods, challenges, and future research directions. Internet of Things. 2022;20:100615. Doi: https://doi.org/10.1016/j.iot.2022.100615

[20] Fraiwan M. Identification of markers and artificial intelligence-based classification of radical Twitter data. Applied Computing and Informatics. 2022. Doi: https://doi.org/10.1108/aci-12-2021-0326

[21] Khan FA, Li G, Khan AN, Khan QW, Hadjouni M, & Elmannai H. AI-Driven Counter-Terrorism: Enhancing Global Security Through Advanced Predictive Analytics. IEEE Access. 2023;11:135864-135879. Doi: https://doi.org/10.1109/access.2023.3336811
[22] Manoharan A, & Sarker M. Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection. International Research Journal of Modernization in Engineering Technology and Science. 2024. Doi: https://doi.org/10.56726/irjmets32644
[23] Olakunle Abayomi A, Chinwe Chinazo O, Onyeka Chrisanctus O, Chuka Anthony A, & Obinna Donald D. Review of AI and machine learning applications to predict and Thwart cyber-attacks in real-time. Magna Scientia Advanced Research and Reviews. 2024;10(1):312-320. Doi: https://doi.org/10.30574/msarr.2024.10.1.0037
[24] Mohiuddin Qadri SSS, Almusawi A, Albdairi M, & Esirgün E. Optimizing Traffic Signal Timing at Urban Intersections: A Simheuristic Approach Using GA and SUMO. 2024 Innovations in Intelligent Systems and Applications Conference (ASYU); 2024/10/16: IEEE; 2024. p. 1-6.10.1109/asyu62119.2024.10757086. http://dx.doi.org/10.1109/asyu62119.2024.10757086
[25] Pathmaperuma MH, Rahulamathavan Y, Dogan S, & Kondoz AM. Deep Learning for Encrypted Traffic Classification and Unknown Data Detection. Sensors (Basel). 2022;22(19):7643. Doi: https://doi.org/10.3390/s22197643
[26] Vashishtha N. (2023). *Artificial Intelligence-assisted Terrorism: A New Era of Conflict*. Vivekanda International Foundation. https://www.vifindia.org/article/2023/august/29/Artificial-Intelligence-assisted-Terrorism-A-New-Era-of-Conflict?slide=$slideshow$
[27] Khattri V, Bhushan N, Singh DK, & Shiblee M. Landscape of Cyber Terrorism on Internet and AI-Powered Countermeasures. Cyberology: Chapman and Hall/CRC; 2025. p. 282-296. http://dx.doi.org/10.1201/9781003409571-17
[28] Kummerow A, Abrha E, Eisenbach M, & Rösch D. Unsupervised Anomaly Detection and Explanation in Network Traffic with Transformers. Electronics. 2024;13(22):4570. Doi: https://doi.org/10.3390/electronics13224570
[29] Lewinsky D, Te'eni D, Yahav-Shenberger I, G. Schwartz D, Silverman G, & Mann Y. Detecting terrorist influencers using reciprocal human-machine learning: The case of militant Jihadist Da'wa on the Darknet. Humanities and Social Sciences Communications. 2024;11(1). Doi: https://doi.org/10.1057/s41599-024-03920-7
[30] Macdonald S, Mattheis A, & Wells D. Using Artificial Intelligence and Machine Learning to Identify Terrorist Content Online. Tech Against Terrorism Europe-15 January; 2024.Retrieved from: https://read-me.org/s/TATE-AIREPORTFINAL1.pdf
[31] Qawasmeh SA-D, AlQahtani AAS, & Khan MK. Navigating cybersecurity training: A comprehensive review. Computers and Electrical Engineering. 2025;123:110097. Doi: https://doi.org/10.1016/j.compeleceng.2025.110097
[32] Sear R, & Johnson NF. Unprecedented reach and rich online journeys drive hate and extremism globally. arXiv preprint. 2023. Doi: https://doi.org/10.48550/arXiv.2311.08258
[33] Singer T. Visual Generative AI in Warfare and Terrorism: Risk Mitigation through Technical Requirements and Regulatory Insights: Technische Universität Wien; 2024.Retrieved from: https://doi.org/10.34726/hss.2024.126126
[34] Syllaidopoulos I, Ntalianis K, & Salmon I, editors. AI-Powered Solutions in Counter-Terrorism and Cybersecurity: Ethical and Operational Challenges. 5th International Ethics Congress; 2024 December 16-18; Adana, Türkiye.https://www.izdas.org/_files/ugd/614b1f_fdbf9265df4e471da9bee5170b6e48e5.pdf
[35] Macdonald S, Mattheis A, & Wells D. Using Artificial Intelligence and Machine Learning to Identify Terrorist Content Online. Tech Against Terrorism Europe-15 January; 2024.Retrieved from: https://read-me.org/s/TATE-AIREPORTFINAL1.pdf
[36] Torregrosa J, Bello-Orgaz G, Martínez-Cámara E, Ser JD, & Camacho D. A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges. J Ambient Intell Humaniz Comput. 2022:1-37. Doi: https://doi.org/10.1007/s12652-021-03658-z
[37] Wang Z, & Thing VLL. Feature mining for encrypted malicious traffic detection with deep learning and other machine learning algorithms. Computers & Security. 2023;128:103143. Doi: https://doi.org/10.1016/j.cose.2023.103143
[38] Yeneakal TY. AI and extremism in social networks [Master theses]: The University of Bergen; 2019.Retrieved from: https://hdl.handle.net/1956/21061
[39] Zamanzadeh Darban Z, Webb GI, Pan S, Aggarwal C, & Salehi M. Deep Learning for Time Series Anomaly Detection: A Survey. ACM Computing Surveys. 2024;57(1):1-42. Doi: https://doi.org/10.1145/3691338