NFSU JOURNAL OF FORENSIC SCIENCE
ISSN: 3049-2408 (Online)
Journal Homepage: https://jfs.nfsu.ac.in/

Revolutionizing Digital Forensics: The Role of AI and ML in Evidence Analysis

Niza Italiya[1], Jay Makwana[1], Bhumi Thakor[2], Harsh Panchal[1,*]
[1]Ansh Tech Labs Pvt. Ltd., Gandhinagar, Gujarat, India
[2]Qualcomm Inc, Hyderabad, India
[harshp2035@gmail.com] (Corresponding Author)*

Abstract

Evidence Analysis plays a valuable role in any investigation carried out for individual or organizations post criminal activities, leveraging AI/ML extensively which helps in automating repetitive and laborious tasks aiding investigators to focus on the significant course of action. AI/ML came up for significant transformations to the field by various techniques for simplification. Advancements allowed experts to dedicate time and other resources to critical aspects of investigation instead. Modern methodologies offer solutions to variegated domains of digital forensics, such as network analysis, device forensics, cybercrime investigations and many more. This paper gives you an overview of how AI and ML techniques can be implemented in digital forensics, by use of different approaches. For example, NLP can be used to analyse large volume of text data, extracting information or identifying patterns. AI driven image and video surveillance tools can detect anomalies, recognize faces or analyse patterns in recordings of live feed. Pattern recognition helps in identifying recurring events or correlations in evidence, like tracing data or detecting frauds in or during cyberattacks.

Keywords: *Digital Forensics, Evidence Analysis, Artificial Intelligence, Machine Learning*

1. Introduction

Digital Forensics as offshoot of forensic science, mainly focuses on recovering and investigations of collected electronic data to gather evidence. Network Forensics, Mobile Forensics and image/video forensics all come under the same umbrella. Special hardware and software tools are designed for recovery and safeguarding of digital evidence. It aids law enforcement investigations using digital evidence which is the core unit of almost all criminal activities [1]. In modern business operations, data analytics plays a significant role in retrieving information, avoiding unwanted data, and addressing issues such as missing values. It also assists in making operations simple by effective visualization and interpretation, decision making, or any other social needs. Cyber-attacks have gradually increased as the organizations are increasingly exposing themselves to the internet, putting a spotlight on the need for more effective detection and evidence collection. Through traditional methods of forensics, it is challenging to detect specific spots out of all the data, which is increasing every microsecond [2]. Cyber

Attacks are complex, resulting in development of cyber forensic investigations assisted by Artificial Intelligence and Machine Learning. Machine Learning techniques are implemented to detect and analyse data including behaviour of user, resulting in detection of anomalies. Intrusion detection algorithms can easily differentiate the changes in patterns or behaviour [3].

Studies reveal that businesses adopting assistance of these technologies experiences remarkable improvements by achieving accuracy and precision in investigations. Implemented AI/ML techniques aided in accelerating threat detection and analytical procedures to guard digital assets of organization. [4, 6] Result proves that traditional techniques were time consuming, prone to errors, complex, resource intensive and laborious which used to require specialized professional knowledge. Artificial intelligence and Machine learning assists by using Natural Language Processing (NLP) techniques for text-based logs to examine extent of data breach using patterns and spot anomalies by continues network monitoring and analysing the data with accuracy in a minimal timeframe. Apart from any other devices or data sources, mobile devices are widely integrated into both personal and professional aspects, bringing a wave of abundant and easily accessible data. With increasing number of users and high frequency of usage that leads to their exceptional value in investigative analysis. Mobile devices have always been in spotlight during investigations, and implementing these technological approaches has enhanced analysing multiple devices smoothly and simultaneously [5]. AI/ML has powerful influence in Digital Forensics and therefore in Mobile Forensics. Instead of using traditional methodologies, advanced AI/ML tools have empowered legal teams and investigators to conduct analysis accurately and within the expected time frame.

Key application of AI/ML in Digital Forensics: Malware Detection and Network analysis: The network forensics process beginning with data acquisition and continues through the analysis and reporting phases as illustrated in Fig. 1. To create valuable insights by analysing digital evidence collected from sources while investigations, Digital Forensics has become an inescapable subject. Popular tools including Encase and Helix with appropriate cybercrime methodologies provide practical forensic analysis [10]. Regardless of capabilities to examine digital evidence at all levels with forensic devices having high compatibility, sharing valuable digital assets and other cybercrimes are increasing simultaneously. To the rescue, Deep Learning models promise to perform malware classification, phishing detection, and intrusion detection by automating the procedures to extricate precise values from network traffic. Facilities of evidence preservation, analysis and interpretation on evidential basis are carried out smoothly using Deep Learning based cyber forensic investigations [12].

As the illustration in Fig. 1 shows, flow of Network forensics in details steps. Identification and investigation of internal and external network attacks, examination of networked devices, reverse engineering of protocols and then analysing it, an upright outcome and overall dealing with cybercrime fits into Network Forensics [11]. Network attacks mainly focus on exploiting communication protocols and other hardware and software vulnerabilities with a range from single device network traffic to wide Internet traffic. It is to deal with information retrieved for both Network level attack detection solutions which focuses on headers of network packets and Application-level attack detection solutions which focuses on investigating data fragments carried out in payloads. Network analysis tools (NATs) provide quality assistance in capturing, identification and analysis of network traffic. Identifying Detection solution with gradually increasing of attacks and sheer volume of data to be analysed is a challenge, and investigations of network traffic is quite strenuous despite of built in NAT automation services as consequential time remains wasted in investigating false positive alerts. To mitigate impacts, Artificial Intelligence centred approaches came with a career to detect and analyse network traffic with accuracy. Results show that proposing Shannon Entropy to identify executable file content to detect anomalies if present, was a considerable and noteworthy implementation [10-13].

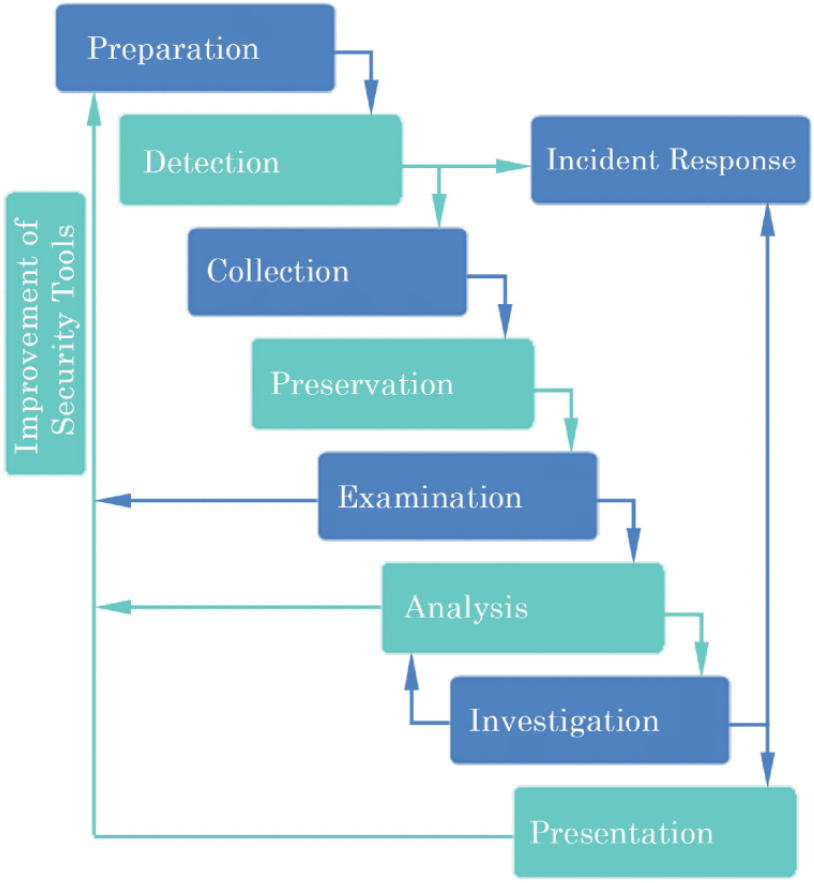

Figure 1. Overview of Network Forensics process that illustrates the flow from acquisition of data to analysis and reporting [13].

2. File Classification and Content Analysis

Influence of Artificial Intelligence and Machine Learning is boosting diurnally in digital forensics investigations. Complexity in investigations caused by immense volume of data and documents to be examined, lacking efficiency by virtue of existing network forensics method which has limitations while processing with large variety of data. It made network forensics an exhausting task. Machine learning techniques intervened to confront the issue of storage and processing of big data using cloud computing.

Performance of image recognition and categorization software using variety of evaluation metrics is exceptionally rising in digital forensics including F1 score, precision, recall, confusion matrix and accuracy. Especially image categorization utility has notable impacts on outcome of legal cases [14]. Furthermore, pattern recognition and Natural language processing utilities in tools enhanced the efficiency. For this reason, software for forensics purposes must be designed with robustness, transparency and interpretability in mind [15].

3. Anomaly Detection in Digital Evidence

In-depth investigations are hindered by data volumes caused by digitalization which directed towards difficulties in analysing data. Anomalies vary and there is no specific general method applicable to detect them all [2]. Machine learning is the fit which facilitates accurate results after analysing. Algorithms and free databases are available to study and test different methods for network anomalies. Tested data with high accuracy can be used as validated and reliable data for forensic purposes. In subjects like network forensics and mobile forensics, Machine Learning methodologies are implemented for ease [15].

4. Challenges and Limitations in AI/ML in Digital Forensics

4.1 Data Privacy and Ethical Concerns: The focal point of Data Privacy is to ensure database security; however, we are facing multitudinous instances of unauthorized access, data breaches and illegal use of individuals or organizations data. Sensitive data should be protected, and the rights of individuals must be respected throughout the procedure of collecting, managing and using information. The role of database administrator is to help secure critical data stored in database and to maintain its security, integrity and confidentiality. For the same reason, regulations such as General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPPA) assists organizations. Robust encryption methods, strict access controls and intrusion detection systems are rigorously implemented by these regulations for safeguarding data.

The crucial part of investigations are ethical concerns and issues raised by dint of storage, use and sharing of genomic data. Bioinformatic data banks are all about unique and priceless information about individual's genetic composition. Privacy and autonomy concerns caused by sensitivity of genetic data lead to demand of addressing challenges in this field. Co-operation and data sharing are crucial while working in advancements of generic research, which directly is concerned with permissions, security and privacy of data. Consent of usage and sharing of sensitive data of contributors is challenging in cases of working with confidential data like genetic information.

There have been multiple concerns on the ethics of Deep Learning, never known before success ranging from data privacy to biased predictions. Researchers have gradually come up with solutions to tackle these issues by introducing fairness metrics, federated learning or differential privacy [16]. Initially, the process presents an ethical federated learning model that incorporates all three measures at the same time. The experiments conducted on the Adult, Bank and Dutch datasets highlight the resulting "empirical interplay" between accuracy, fairness and privacy [18].

4.2 Algorithm bias and Legal Admissibility: Deploying Deep Learning technologies raised ethical issues in HR practices. Algorithmic bias was amongst the issues with data privacy concerns and level of transparency in decision making. Unfairness in algorithms decision making which emanates from biased data training, triggers legal admissibility and unfair outcomes [19]. The need for reliability and fairness results in excluding evidence generated by biased algorithms in courts. Explanation of AI outputs is major factor affecting legal admissibility. Difficulties are to be faced in complying with regulations like GDPR, if using AI based forensics tools caused by poor data processing and storage [20].

5. Emerging trends and innovations in AI/ML in Digital Forensics

5.1 AI driven automation and predictive analysis: Rapid expansion of digital forensics with technology is thriving for analysing digital evidence during investigations. Call for effective investigation strategies with higher level of efficiency leads to inculcate technologies like Artificial Intelligence and Machine Learning. Automated analysis of data facilitates more accuracy as algorithms performs swift scans, making the procedure more streamlined [1].

In recent years, significant application of Artificial Intelligence, Machine Learning and Deep Learning amplified Digital Forensics. Identification and classification functionalities are way more unchallenging, using trained and tested AI models with great resolution and precision [4]. YOLOv8, advanced and pre-trained object recognition models are professionals, dealing with real-time identification and classification of suspects. Digital materials and datasets were considered for training these nano, small, medium, large and extra-large sized models. And for ease of use, desktop applications were developed to aid digital forensics. Machine Learning technology is used to analyse diverse historical datasets to predict or reveal potential criminal behaviour and future criminal activities [5, 7, 21].

6. Blockchain Integration for Evidence Integrity

Ensuring integrity and security of digital evidence with Blockchain Technology is growing immensely in Digital Forensics. Use of Cloud storage services is gradually increasing and causing integrity and security concerns. With transparent and immutable transactions, blockchain allows verification, perseverance of chain of custody, evidence credibility, reliability and tamper resistance as result of decentralization and real time audits of transactions. Integration of technology contributes to scalability, security, performance and enhances trust in cloud storage environments.

Video footage as evidence is sourced from public or private surveillance systems for criminal investigations, which might be stated as prime facet subsequently. Untrusted video sources as information or evidence while investigation raises integrity issues. Blockchain technology implements storage of metadata as blockchain transactions, allowing experts and authorized entities to validate unauthorized alterations making it tamper-proof. Metadata comprises of hash value and message authentication code where SHA256 is used to store into blockchain with public key after encryption by asymmetric cryptography.

In Information Management Systems, Data Integrity being the principal issue, recurrent database auditing is the exclusive approach. Smart contracts automate transactions across many industries when certain conditions are met based on terms of agreement. In the traditional Land Registry System, brokers used to play a vital role in ensuring registration of land/property where everything is marked down for archives regarding transactions between both parties by authorized government officials [22]. Chances of alterations in documents with such powerful controls amplifies raising concerns with tangible proof of land. Smart contracts came up as a remedy that deals with transactions regarding assets and making procedure more secure, cost-efficient and rapid. Smart contracts are easily traceable, irreversible and transparent as they are stored on distributed database that ensures terms of contract immutability [23-27].

7. Conclusion

Even in its budding stage, AI/ML methods entangled cybercrime timelines, pattern recognition, orchestrating responsive strategies to incidents ensuring sharp and rapid responses to cyber threats [7]. Artificial intelligence can be considered as game changer in a drive of these challenges, reshaping cybersecurity by radical improvements in threat detection, mitigation and incident response [8,9]. Transformative advantages offered by AI/ML includes anomaly detection, predictive analysis, and efficient evidence in Digital Forensics. Modern technology works well to evaluate security breaches. Nevertheless, data privacy concerns, legal admissibility, and algorithmic bias are still a challenge, but overall provides a remarkable benefit for individuals and various industries. However, hybrid approaches of combined, traditional and modern techniques enhance transparency and developing standards for ethical AI deployment in forensic investigations. This paper signifies efficiency of AI blended with human efficiency resulting into accuracy and speed despite of challenges, ethical, technical and legal limitations.

References

[1] D. Dunsin, M. C. Ghanem, K. Ouazzane, and V. Vassilev, "A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response," Forensic Science International: Digital Investigation, 2024.

[2] S. Qadir and B. Noor, "Applications of Machine Learning in Digital Forensics," 2021 International Conference on Digital Futures and Transformative Technologies, ICoDT2 2021, 2021.

[3] S. Abraham, K. Alakananda, and N. A. Amdalli, "A comprehensive review on digital forensics intelligence," Advancements in Cybercrime Investigation and Digital Forensics, 2023.

[4] A. M. Qadir and A. Varol, "The Role of Machine Learning in Digital Forensics," 8th International Symposium on Digital Forensics and Security, ISDFS 2020, 2020.

[5] S. Misra, C. Arumugam, S. Jaganathan, and S. Saraswathi, "Confluence of Al, Machine, and Deep Learning in Cyber Forensics," in Confluence of Al, Machine, and Deep Learning in Cyber Forensics, 2020.

[6] B. Fakiha, "Enhancing Cyber Forensics with Al and Machine Learning: A Study on Automated Threat Analysis and Classification," International Journal of Safety and Security Engineering, 2023.

[7] A. Vasilaras, N. Papadoudis, and P. Rizomiliotis, "Artificial intelligence in mobile forensics: A survey of current status, a use case analysis and Al alignment objectives," Forensic Science International: Digital Investigation, 2024.

[8] O. Mykhaylova, T. Fedynyshyn, V. Sokolov, and R. Kyrychok, "Person-of-Interest Detection on Mobile Forensics Data-Al-Driven Roadmap," CEUR Workshop Proceedings, 2024.

[9] H. Henseler and J. Hyde, "Technology-assisted analysis of timeline and connections in digital forensic investigations," CEUR Workshop Proceedings, 2019.

[10] S. Ramesh, K. Prathibanandhi, P. Hemalatha, and A. R. Basha, "The Convergence of Novel Deep Learning Approaches in Cybersecurity and Digital Forensics," in Simulation and Analysis of Mathematical Methods in Real-Time Engineering Applications, 2021.

[11] K. Nguyen, D. Tran, W. Ma, and D. Sharma, "An approach to detect network attacks applied for network forensics," in Proceedings of the 11th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2014, 2014.

[12] K. Nguyen, D. Tran, W. Ma, and D. Sharma, "A new approach to executable file fragment detection in network forensics," in Lecture Notes in Computer Science, 2014.

[13] S. Rizvi, M. Scanlon, J. McGibney, and J. Sheppard, "Application of Artificial Intelligence to Network Forensics: Survey, Challenges, and Future Directions," IEEE Access, 2022.

[14] G. S. Chhabra, V. Singh, and M. Singh, "Hadoop-based analytic framework for cyber forensics," International Journal of Communication Systems, 2018.

[15] N. Kumar, P. K. Keserwani, and S. G. Samaddar, "A Comparative Study of Machine Learning Methods for Generation of Digital Forensic Validated Data", in Proceedings of the 9th International Conference on Advanced Computing, 2017.

[16] E. Pina, J. Ramos, H. Jorge, and P. Martins, "Data Privacy and Ethical Considerations in Database Management," Journal of Cybersecurity and Privacy, 2024.

[17] Y. Balagurunathan and R. R. Sethuraman, "An Analysis of Ethics-Based Foundation and Regulatory Issues for Genomic Data Privacy," Journal of The Institution of Engineers (India): 2024.

[18] M. Padala, S. Damle, and S. Gujar, "Federated Learning Meets Fairness and Differential Privacy," in Communications in Computer and Information Science, 2021.

[19] B. Singh, M. Neti, and S. Choudhury, "Ethical Considerations in the Use of Deep Learning for HR Decision-Making," in Proceedings of the International Conference on Computing, Power, and Communication Technologies (IC2PCT 2024), 2024.

[20] Z. Geradts, "Forensic Challenges on Multimedia Analytics, Big Data and the Internet of Things," in Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018), 2018.

[21] S. Karakuş, M. Kaya, and S. A. Tuncer, "Real-Time Detection and Identification of Suspects in Forensic Imagery Using Advanced YOLOv8 Object Recognition Models," Traitement du Signal, 2023.

[22] Z. Li, M. N. Ahmad, Y. Jin, and Z. Lantu, "Unleashing Trustworthy Cloud Storage: Harnessing Blockchain for Cloud Data Integrity Verification," in Lecture Notes in Computer Science, 2024.

[23] R. A. Michelin, N. Ahmed, S. S. Kanhere, and S. Jha, "Leveraging Lightweight Blockchain to Establish Data Integrity for Surveillance Cameras," in Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency, ICBC 2020, 2020.

[24] S. Salagrama, V. Bibhu, and A. Rana, "Blockchain-Based Data Integrity Security Management," Procedia Computer Science, 2022.

[25] K. P. Chairinnisa, H. N. Hamdani, I. S. Edbert, and A. Aulia, "Kereta Api Indonesia (KAI) Boarding Gate Face Recognition Data Security Using Blockchain Technology," in Proceedings of the 2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics, ICoABCD 2024,

2024.

[26] S. Verma, S. K. Pathak, P. Chaudhary, and S. Yousuf, "Land Registry System Using Blockchain Technology," in Proceedings of the 2024 International Conference on Emerging Innovations and Advanced Computing, INNOCOMP 2024, 2024.

[27] A. M. Palanisamy and R. V. Nataraj, "A Novel Methodology to Ensure Data Integrity in Enterprise Information Systems Using Blockchain Technology," in Proceedings of the 1st International Conference on Electrical, Electronics, Information, and Communication Technologies, ICEEICT 2022, 2022.

About the Authors

Niza Italiya:

Author holds BSc. in CA and IT and is currently working as Software Developer at Ansh Tech Labs. Research interest includes Digital Forensic, OSINT, Incident Response and Web Filtering.

Jay Makwana:

The author holds an M. Tech Cybersecurity, He holds 7+ years of experience in Product development and designing. His research interests include Digital Forensic, Incident Response, OSINT, and IAM.

Bhumi Thakor:

The author holds an M. Tech in Cybersecurity and Incident Response and is currently working as a Senior Cybersecurity Engineer at Qualcomm. Their research interests include IoT security, Cloud security, and the dark web.

Harsh Panchal:

The author holds an M. Tech in Cybersecurity and Digital Forensics. He holds 5+ years of experience in product development. Research interests include OSINT, Drone Forensics, Dark Web AI and Cyber Security.