10.48047/jocaaa.2025.34.05.27

Machine Learning-Driven Forensics and Evidence Analysis for Cybercrime Investigations

[1] Amtabh Srivastava
Research Scholar, Dept of Computer science

[2] Dr. Jitender Rai
Dept of Computer science

[1.2] Sunrise University, Alwar

Corresponding Author Email id: Amitabh7500@Yahoo.Com

Abstract: The increasing complexity and volume of cybercrimes necessitate innovative forensic techniques capable of handling multifaceted digital evidence. This paper explores how machine learning (ML), a subdomain of artificial intelligence (AI), revolutionizes digital forensic science through intelligent data analysis, pattern recognition, and automated decision-making. Drawing on extensive literature from 2020 to 2025, the study highlights how ML and ML technologies, including convolutional neural networks, natural language processing, and 3D image analysis, enhance various forensic applications from skeletal identification to anomaly detection and timeline reconstruction. The paper also discusses the integration of ML in legal procedures, challenges like adversarial attacks and data bias, and ethical concerns surrounding algorithm transparency and evidence admissibility. Through a structured ten-step forensic workflow, the study proposes a framework that combines technical precision with investigative intelligence, improving efficiency in cybercrime resolution. The findings underscore the need for interdisciplinary collaboration and regulatory refinement to responsibly harness ML in forensic investigations.

Keywords: Cybercrime Forensics, Machine Learning, Digital Evidence Analysis, ML Investigative Tools

I. Introduction

In recent years, the digital revolution has profoundly altered the landscape of criminal investigation, especially in the realm of cybercrime. With cyber threats escalating in complexity and scale, traditional forensic techniques are increasingly found wanting in their ability to handle the vast, multidimensional, and evolving nature of digital evidence. This gap has necessitated a paradigm shift in investigative methods, one in which Artificial Intelligence (AI) and Machine Learning (ML) are playing transformative roles. Machine learning a subfield of ML concerned with developing algorithms that improve from data experience has emerged as a vital tool for forensic analysts, offering the ability to detect patterns, predict events, and classify data at scales that would be humanly unmanageable. Machine learning-driven forensics and evidence analysis present a compelling convergence of technology and criminology, integrating the computational power of ML with the methodical scrutiny of forensic science. As cybercrime becomes more prevalent ranging from identity theft and financial fraud to ransomware and deepfake-based misinformation the need for intelligent, adaptable, and scalable forensic tools is more pressing than ever. This introduction delves into the theoretical foundations, technological advancements, and practical implementations of machine learning within digital forensics, drawing upon a robust body of scholarly literature from the past five years.

Foundations and Scope of ML in Forensics: Artificial Intelligence, as a broad computational discipline, has long aimed to simulate cognitive processes such as learning, reasoning, and decision-making (Iqbal et al., 2020).

10.48047/jocaaa.2025.34.05.27

Iqbal et al. (2020) emphasize that AI's ability to extract insights from vast data volumes is especially relevant to forensic analysis, which often involves mining through terabytes of heterogeneous data. They describe how ML subfields like machine learning and deep learning have advanced digital forensics by automating processes that were once manual, time-consuming, and error-prone. Similarly, Jeong (2020) provides a comprehensive taxonomy of ML crimes, classifying them into categories where ML acts as either a tool or a target. These classifications help contextualize how machine learning can both enable and combat cybercriminal activities. Jeong underscores the dual nature of ML in cybersecurity for every advancement in AI-based detection methods, adversaries exploit vulnerabilities in ML systems to develop more sophisticated attacks. This interplay between defence and offense has catalysed the evolution of forensic strategies.

Machine Learning in Human Identification and Biometric Forensics: The intersection of ML and biomedical imaging has also impacted forensic anthropology. Mesejo et al. (2020) present one of the first comprehensive surveys on applying ML to skeletal-based human identification, highlighting the potential of ML and computer vision in estimating biological profiles and reconstructing facial features. Their work underscores the role of ML in medico-legal investigations, especially in scenarios involving mass disasters or unidentified human remains. In a related vein, Khanagar et al. (2021) reviewed the application of ML models primarily convolutional neural networks (CNNs) and artificial neural networks (ANNs) in forensic odontology. Their analysis reveals that ML systems match, and sometimes exceed, the performance of trained human examiners in identifying bite marks, estimating age and gender, and predicting mandibular structures.

Advancing Forensic Automation and Digital Intelligence: The push for automation in forensics is echoed by Jarrett and Choo (2021), who note that AI-powered tools are increasingly being adopted by law enforcement agencies to reduce costs and improve accuracy in digital evidence collection. The authors also discuss how automation complements forensic workflows by enabling continuous monitoring, real-time alerts, and predictive modelling. Gupta et al. (2020) emphasize the ubiquity of ML across all stages of the forensic process from initial scene investigation to courtroom judgment. They identify various ML applications such as blood spatter analysis, image enhancement, and satellite-based crime mapping, positioning ML as a holistic forensic assistant.

3D Analysis and Cross-Disciplinary Applications: The transformative potential of ML is perhaps most evident in deep learning's application to three-dimensional imaging. Thurzo et al. (2021) explore how 3D convolutional neural networks (3D CNNs) can automatically extract features from CBCT scans, a breakthrough for forensic anthropology. Their research promotes interdisciplinary collaboration, encouraging non-technical medical professionals to engage with AI-powered tools for tasks such as sex determination, age estimation, and tissue reconstruction.

Natural Language Processing and Social Media Forensics: As digital communication becomes a principal medium of interaction, the role of Natural Language Processing (NLP) in forensic investigations grows. Ukwen and Karabatak (2021) conducted a literature review on NLP in digital forensics, noting its application in analyzing text-based data for sentiment, authorship, and deception. Bokolo and Liu (2024) extended this concept into social media forensics, where ML is used to detect extremist propaganda, hate speech, and cyberbullying. Their work highlights the use of Graph Neural Networks (GNNs) for social network modelling a powerful tool in mapping criminal networks.

Security, Ethics, and ML Attacks: Despite these advantages, the integration of ML into forensic systems is not without challenges. Manasa and Kumar (2022) explore adversarial attacks targeting ML models, such as image

10.48047/jocaaa.2025.34.05.27

tampering and deepfake generation. They propose robust forensic tools like EXIF-SC and Noiseprint, which use deep learning to authenticate image sources and detect anomalies. Their work calls for a shift toward secure-bydesign ML systems that anticipate and resist manipulation. Similarly, Alnafrani and Wijesekera (2022) tackle the limitations of memory-constrained IoT devices in forensic investigations. Their AI-driven system emulates cyberattacks to predict potential vulnerabilities, providing forensic examiners with synthetic yet realistic data to identify root causes. Ethical considerations also arise in the forensic use of AI. Solanke and Biasiotti (2022) caution that the opacity of ML models may hinder legal admissibility and standardization. They propose three core tools for validating AI-generated evidence and stress the importance of interpretability, fairness, and accountability.

Machine Learning for Evidence Mining and Case Management: In practical applications, ML models streamline complex investigative tasks. Dunsin et al. (2022) introduce a multi-agent digital investigation framework using case-based reasoning (CBR) to accelerate forensic tasks. Through deploying intelligent software agents that handle sub-tasks autonomously, their framework demonstrated faster evidence analysis compared to conventional tools. Faqir (2023) elaborates on AI's influence on legal systems, especially in optimizing decisionmaking processes such as sentencing and parole. He advocates for the integration of biometric systems and intelligent surveillance to proactively deter crime and improve public safety outcomes.

Real-World Challenges and Country-Specific Issues: Not all regions benefit equally from these advancements. Obidimma and Ishiguzo (2023) explore the challenges Nigeria faces in adopting ML for cybercrime investigation. Their findings cite a critical skills gap among legal professionals and law enforcement, as well as outdated legal frameworks that fail to accommodate AI-generated evidence. These systemic barriers limit the utility of ML and raise concerns about fairness, privacy, and due process. Gogia and Rughani (2023) highlight a similar need for ML in incident response teams, particularly in navigating technologies like smart wearables, blockchain, and home automation. Their exploratory survey finds overwhelming support for AI-enhanced automation in cybercrime investigations, reinforcing the need for continual tool development and expert training.

Evaluation Metrics and Digital Forensic Pipelines: Vasilaras et al. (2024) assess ML performance in mobile forensics using standard classification metrics such as accuracy, precision, recall, and F1-score. Their ML Alignment Framework proposes design principles for transparency and robustness in forensic ML systems, aligning them with legal expectations and investigative needs. Zangana and Omar (2025) offer a contemporary lens on the convergence of ML and traditional digital forensics. They explore how ML technologies, especially deep learning and NLP, empower investigators to analyse massive datasets efficiently. The authors also examine privacy, ethics, and accountability in the deployment of forensic AI. Bansal et al. (2025) provide a broader view of AI's role in law enforcement. They detail how surveillance systems, digital forensics, and genetic profiling have transformed modern criminal investigation. Nevertheless, they caution against overreliance on opaque technologies, calling for oversight mechanisms to mitigate unintended consequences.

10.48047/jocaaa.2025.34.05.27

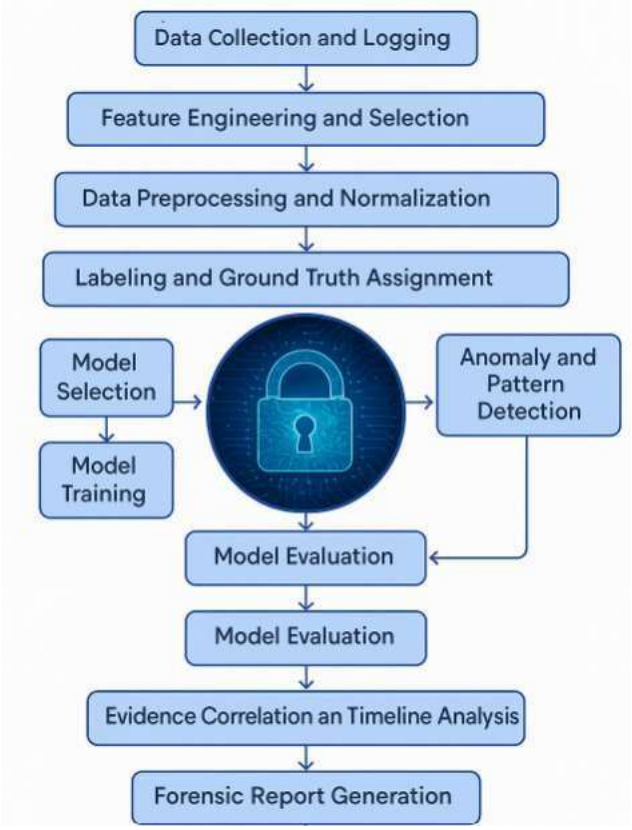

Fig. 1: ML-Driven Forensics and Evidence Analysis Framework for Cybercrime Investigations

II. Related Reviews

Iqbal et al. (2020). discussed artificial intelligence (AI) as an established domain within computer science, concerned with enabling machines to handle computationally intensive or complex tasks typically requiring human intelligence. They noted that ML involved a blend of technologies capable of extracting insights and patterns from vast amounts of data, a key aspect in forensic analysis. The chapter was reported to have focused broadly on ML and its subfields, including machine learning and deep learning, while also elaborating on ML and data mining techniques relevant to digital forensics. The authors were said to have identified limitations in existing methods and proposed a novel approach aimed at enhancing the identification of potential data, detecting variables of interest, and offering clearer insights. The concluding section was described as assessing the future trajectory of digital forensics.

Jeong (2020) reviewed how advances in Artificial Intelligence (AI) had impacted numerous fields, including computer science, robotics, social engineering, psychology, and criminology. The study highlighted that while ML had addressed various challenges, researchers had increasingly emphasized potential security threats associated with ML algorithms and training data. It was noted that since ML systems inherited the security vulnerabilities of traditional computer systems, concerns about novel cyberattacks enhanced by ML were also growing. Moreover, due to AI's integration into physical environments—such as autonomous vehicles and

10.48047/jocaaa.2025.34.05.27

intelligent virtual assistants—the risks extended beyond cyberspace, posing potential physical harm. In this context, the article presented a literature review on AI-related security threats and crimes, defining the term "ML crime" and categorizing it into two types: ML as tool crime and ML as target crime, drawing inspiration from existing cybercrime taxonomies. The review systematically explored foreseeable ML crimes, examined relevant forensic techniques, and analyzed the unique characteristics of these crimes. It also identified challenges that traditional forensic methods struggled to address and concluded by outlining open issues, stressing the need for novel strategies in ML forensics.

Mesejo et al. (2020) conducted what was regarded as the first comprehensive survey on the application of artificial intelligence techniques in the analysis of biomedical images for forensic human identification. The study emphasized the critical role of human identification in modern society, especially within medico-legal contexts, and noted that technological advancements in this area could significantly aid the growing demand for precise and reliable identification tools. The authors had outlined the relevance and applicability of forensic anthropology in diverse identification scenarios and subsequently reviewed key trends involving computer vision, machine learning, and soft computing methods. These techniques were examined in relation to estimating biological profiles, conducting comparative radiography, performing craniofacial superimposition, and analyzing trauma, pathology, and facial reconstruction. The paper also addressed both the capabilities and limitations of the methods employed, and concluded with a reflection on methodological concerns and directions for future research.

Gupta et al. (2020) aimed to explore improved and expansive methods to enhance, broaden, and integrate forensic science techniques across its various branches by leveraging current scientific advancements and incorporating emerging technologies, such as Artificial Intelligence (AI). The study discussed the contemporary and potential future applications of ML within forensic science. It was highlighted that ML could be utilized in areas like blood pattern recognition and analysis, crime scene reconstruction, digital forensics, image processing, and satellite monitoring. The authors indicated that ML held a wide range of applications throughout the forensic process—from the initial investigation of a crime scene to the final judgment in a court of law—demonstrating significant potential for improving the precision of forensic outcomes.

Jarrett and Choo (2021) reviewed that artificial intelligence (AI)—broadly defined to encompass machine learning and deep learning—and automation had been regarded as complementary computing disciplines. They noted that AI-powered software, programs, operating systems, and devices were being developed extensively to automate a wide range of processes and operations. The main objectives of integrating ML and automation were said to include enhancing efficiency, improving accuracy, and reducing costs. Although some ongoing costs remained associated with automation, these were generally reported to be significantly lower than the expenses involved in performing tasks manually, thereby increasing the potential for a high return on investment. An emerging area of application identified in their review was digital forensics, where U.S. federal and state law enforcement agencies had begun examining the benefits of AI-powered technologies. This development was believed to enhance the accuracy and effectiveness of digital forensic investigations, potentially aiding in the resolution of more cases.

Thurzo et al. (2021) reviewed the application of three-dimensional convolutional neural networks (3D CNNs) in forensic medicine, emphasizing their potential in image processing and recognition through deep learning for both generative and descriptive tasks. They noted that, unlike earlier models, CNNs had the advantage of automatically detecting key features without human supervision. Their research highlighted how 3D CNNs were employed to extract features from 3D volumes or sequences of 2D images, such as slices from cone-beam computed

10.48047/jocaaa.2025.34.05.27

tomography (CBCT) scans. The study aimed to foster interdisciplinary collaboration between forensic medical professionals and ML engineers, encouraging clinicians with limited ML knowledge to engage with advanced technologies in pursuit of forensic advancements. The authors introduced a novel 3D CNN workflow for full-head CBCT analysis and explored customized applications of 3D CNNs in forensic research from five angles: sex determination, biological age estimation, 3D cephalometric landmark annotation, growth vector prediction, and soft-tissue estimation from the skull and vice versa. They concluded that the integration of 3D CNNs could mark a transformative shift in forensic medicine by significantly enhancing analysis workflows through neural network-based approaches.

Khanagar et al. (2021) reviewed the application and performance of artificial intelligence (AI) technology in forensic odontology (FO), which primarily involved the identification of individuals through dental remains such as teeth and jawbones. The review highlighted that ML had emerged as a significant advancement in supporting reliable decision-making within forensic sciences. The authors had conducted a systematic search of articles published between January 2000 and June 2020 using established search engines. To assess the risk of bias in the selected studies, the QUADAS-2 tool had been utilized. The review indicated that ML had been extensively applied in FO for tasks such as bite-mark identification, mandibular morphology prediction, gender determination, and age estimation. Most ML models reported in these studies were based on artificial neural networks (ANNs) or convolutional neural networks (CNNs). The findings from the reviewed studies had been promising, with several indicating that the ML models demonstrated accuracy and precision comparable to that of trained human examiners. Such models were considered potentially valuable tools in the identification of victims in mass disasters and could serve as supportive aids in medico-legal investigations.

Ukwen and Karabatak (2021) reviewed the growing use of Artificial Intelligence (AI) by digital forensics and cybersecurity professionals in combating cybercrime. They noted that, over the years, there had been an increase in the adoption of ML technologies, particularly Natural Language Processing (NLP), for various applications such as data mining, knowledge representation, pattern recognition, and expert systems. Their research focused on surveying existing literature related to NLP-based systems within the context of digital forensics and cybersecurity, highlighting their roles, applications, challenges, and future directions. The article was intended to serve both as a guide for researchers and practitioners to understand the current landscape of cybersecurity and digital forensics, and as a roadmap for future developments in the field.

Manasa and Kumar (2022) had emphasized the need for adopting new research approaches to address security threats in Artificial Intelligence (AI)-based systems. Their study had aimed to investigate ML attacks described as "malicious by design" and had focused on conceptualizing the problem alongside developing strategies for countering such attacks using digital forensic tools. They had highlighted a specific class of adversarial problems involving image tampering in computational processes related to digital photography, computer vision, and pattern recognition, including facial capping algorithms. The review had discussed state-of-the-art forensic developments such as the application of end-to-end neural network training pipelines for image rendering and provenance analysis, deepfake image analysis using frequency methods, wavelet analysis, and tools like Amped Authenticate, the use of capsule networks to detect forged images, information transformation for feature extraction via tools such as EXIF-SC, Splice Radar, and Noiseprint, and the employment of GAN-based models as anti-image forensic tools. These advancements had been studied in depth to inform the design of a novel research approach aimed at enhancing the utility of digital forensics.

10.48047/jocaaa.2025.34.05.27

Solanke and Biasiotti (2022). discussed the significant impact of artificial intelligence (AI) on various sectors, highlighting its potential in addressing complex challenges in digital forensics. They emphasized how AI, particularly through machine learning models, can enhance forensic analysis by detecting patterns and recognizing hidden evidence in digital artifacts, which might otherwise be overlooked in manual investigations. Despite several proposals for integrating ML into digital forensics, skepticism about its opacity has hindered the proper formalization and standardization of the field. The authors presented three essential tools for developing robust machine-driven methodologies in digital forensics, focusing on methods for evaluating, standardizing, and optimizing ML techniques. They also examined the strengths and weaknesses of these tools in the context of digital forensics applications, noting their implications for the admissibility of forensic evidence in legal proceedings.

Dunsin et al. (2022) discussed the challenges faced by digital investigators in identifying evidence within digital information, emphasizing that the evolving criminal tactics outpace the current forensic procedures and technologies. They noted the increasing difficulty in determining the relevance of sources to specific investigations, with criminals exploiting weaknesses in the digital investigation process. The paper highlighted the crucial role of artificial intelligence (AI) in digital forensics, particularly its effectiveness in detecting and preventing criminal activity. Through ML algorithms, risks could be identified, and criminal activities could be forecasted, supporting the development of theories for court presentations. The authors proposed a multiagent framework for digital investigations using intelligent software agents (ISA) to address tasks collaboratively, with each agent's knowledge and rules tailored to the investigation type. They introduced the case-based reasoning (CBR) technique to classify criminal investigations efficiently and implemented the framework using tools such as the Java Agent Development Framework, Eclipse, Postgres repository, and a rule engine. Experiments with the framework on the Lone Wolf image files demonstrated a significant reduction in processing time, with the MADIK framework completing tasks like integrity checks far faster than traditional forensic toolkits. The framework's integration with Python also facilitates the addition of other forensic tools, enhancing its versatility.

Alnafrani and Wijesekera (2022). highlighted the growing challenges faced by forensic investigators due to the limited memory and storage capabilities of most Internet of Things (IoT) devices. These limitations make it difficult to pre-process data and gather relevant evidence for reconstructing attacks. To address this, they proposed the use of artificial intelligence (AI)-inspired techniques to automate forensic analysis by emulating attacks to identify and collect forensic evidence. They employed a differentiable inductive logic programming (∂ILP) system to extract attack emulation information from various sources, including device- and subsystem-level vulnerabilities within an enterprise network. By predicting potential attacks based on previous incidents involving similar configurations, their methodology demonstrated the ability to generate rules that could aid forensic examiners in identifying evidence without needing to execute the attacks.

Faqir (2023) examined the integration of Artificial Intelligence (AI) in digital criminal investigations, highlighting its methodologies, legal implications, and impact on the justice system. The study, which adopted a multifaceted approach using qualitative, descriptive, and analytical methods, drew data from legal documents and scholarly literature. It revealed AI's significant role in law enforcement, influencing arrest procedures, release decisions, sentencing, recidivism prediction, criminal activity identification, and suspect apprehension through advanced audio analysis. The findings emphasized the transformative power of machine learning in case data analysis and organization, offering recommendations to optimize ML usage in criminal investigations. These included prioritizing high-risk cases with diverse data sources for informed decision-making, employing ML for crime prediction, suspect identification, and reinforcing security measures. It also advocated for AI-powered

10.48047/jocaaa.2025.34.05.27

biometric identification systems for identity verification and intelligent surveillance solutions to prevent crime proactively. Finally, the study underscored machine learning's role in enhancing case management efficiency within the criminal justice system.

Gogia and Rughani (2023) examined the complexities involved in investigating cybercrimes, highlighting the challenges posed by rapidly evolving technologies such as smart wearables, home automation, cloud computing, and cryptocurrencies. The authors explored the difficulties faced by incident response teams in keeping pace with these technological advancements used by criminals. Through an exploratory study, they gathered responses from cybercrime experts, law enforcement professionals, researchers in incident response teams, and academics via a voluntary and consensual online Google form. The findings underscored the growing need for artificial intelligence and machine learning to enhance current processes and develop automated solutions to tackle emerging cybercrimes.

Obidimma and Ishiguzo (2023) reviewed the growing threat of cybercrime in Nigeria, from online fraud to advanced hacking, which jeopardized national security, the economy, and individual rights. The authors highlighted the potential of Artificial Intelligence (AI) in enhancing cybercrime detection, investigation, and prevention efforts, particularly through machine learning, predictive analytics, and automated threat detection. However, they noted that the integration of ML into Nigeria's cybercrime landscape faced significant challenges due to a considerable skills gap, both technical and legal. Law enforcement agencies and cybersecurity professionals lacked the expertise to effectively deploy ML tools, and Nigerian legal practitioners were unfamiliar with the implications of ML in legal processes, including the admissibility of AI-generated evidence. The authors also emphasized the inadequacy of existing laws, such as the Cybercrime (Prohibition, Prevention, etc.) Act 2015, in addressing AI's complexities, which resulted in legal uncertainties and regulatory gaps. Additionally, ethical concerns like data privacy, ML bias, and over-surveillance were raised, underlining the need for legal safeguards. The paper explored these dual challenges and proposed reforms to better incorporate ML in Nigeria's legal framework while ensuring fairness and privacy.

Das et al. (2023). explored the significant advancements in forensic and criminal investigations, particularly emphasizing the transformative role of Artificial Intelligence (AI). They highlighted how ML has revolutionized the field by improving the efficiency and accuracy of investigations and evidence analysis. ML systems were noted for their ability to perform tasks traditionally requiring human skills such as strategic thinking, speech recognition, and cognitive reasoning—more quickly and with fewer errors. Despite the potential for ML to enhance investigative outcomes, the authors also acknowledged the challenges faced by experts, including the vast amounts of data and complex, often incomplete, information that could lead to flawed investigations or injustices. The paper underscored AI's growing importance in addressing these challenges and its capacity to elevate the field of forensic and criminal investigation.

Vasilaras et al. (2024) conducted a comprehensive study to assess the role and effectiveness of Artificial Intelligence (AI) and Machine Learning (ML) systems in mobile forensics, acknowledging their growing influence in digital forensic investigations. The researchers performed a survey to evaluate ML functions in mobile forensic software from practitioners' perspectives, aiming to deepen the understanding of current practices in the field. The study focused on the performance of image categorization software, assessing it through various metrics, including accuracy, precision, recall, F1-score, and the confusion matrix. Furthermore, the authors developed an ML Alignment framework to conceptualize strategies and solutions for Mobile and Digital Forensics, addressing technical and administrative challenges. The study highlighted the importance of

10.48047/jocaaa.2025.34.05.27

interpretability, transparency, and robustness in ML systems, stressing that these qualities significantly impact the outcome of legal cases. It also emphasized the need for ML governance and standardized procedures. The results indicated that the accuracy and robustness of image categorization software are critical for forensic investigations, underlining the necessity for designs that prioritize transparency and interpretability.

Bokolo and Liu (2024) discussed how social media platforms had revolutionized human communication and social interactions, highlighting their undeniable positive impacts. However, they also pointed out the rise of harmful antisocial behaviors, such as cyberbullying, misinformation, hate speech, radicalization, and extremist propaganda, which had caused significant harm to society, particularly vulnerable populations. To address these issues, the field of social media forensics emerged, enabling investigators and law enforcement to better tackle cybercrimes. The paper reviewed recent research in the field, focusing on how artificial intelligence (AI) techniques were being applied in social media forensics investigations, particularly through natural language processing to identify extremist ideologies, detect online bullying, and analyze deceptive profiles. Additionally, the authors explored the use of Graph Neural Networks (GNNs) for social network modeling in forensic contexts and concluded by discussing the challenges faced in the field while suggesting future research directions.

Zangana and Omar (2025) explored the intersection of digital forensics and artificial intelligence (AI), highlighting the transformative impact ML was having on digital investigative techniques. It examined how digital forensics, which traditionally focused on the collection, preservation, and analysis of electronic evidence, was encountering both challenges and opportunities in the age of AI. With the rapid proliferation of digital devices and the increasing sophistication of cyber threats, the need for enhanced methods of accurate and timely evidence gathering was emphasized. ML technologies, including machine learning, natural language processing, and deep learning, were described as empowering forensic professionals to detect, analyze, and interpret vast volumes of data more efficiently than ever before. The chapter also discussed the fundamentals of digital forensics, the integration of ML tools in forensic processes, and the ethical and privacy concerns that arose. The insights offered aimed to help readers understand how ML could bolster digital forensics, enabling more proactive and precise responses to cybercrime.

Bansal et al. (2025) explored the impact of technological advancements on criminal investigations, noting how their integration has significantly transformed law enforcement's approaches and enhanced their capacity to solve crimes. The authors examined the development and application of digital technologies and forensic science, emphasizing how these innovations have revolutionized the gathering and evaluation of evidence. Key topics discussed included the use of surveillance technologies, digital forensics, and DNA analysis, all of which have reshaped investigative practices. However, the paper also highlighted the challenges and limitations associated with these advancements, such as issues with reliability, privacy concerns, and the potential for digital injustices.

III. Author Findings from Existing Reviews

Author(s)	Year	Objective	Methodology	Findings
Iqbal et al.	2020	To examine ML and its subfields in digital forensics.	Literature review on ML and data mining techniques.	Identified limitations and proposed novel forensic approaches.

10.48047/jocaaa.2025.34.05.27

Jeong	2020	To explore AIâ€™s role in crimes and security threats.	Taxonomy of ML crimes, literature synthesis.	Highlighted ML as both crime tool and target; need for new strategies.
Mesejo et al.	2020	To review ML in biomedical image-based forensic ID.	Survey on machine learning and computer vision applications.	Outlined uses in forensic anthropology and facial reconstruction.
Gupta et al.	2020	To explore integration of ML in forensic science.	Discussion on ML applications in forensic processes.	ML seen as enhancing precision from crime scene to courtroom.
Jarrett and Choo	2021	To assess ML and automation in digital forensics.	Literature-based analysis of ML adoption in law enforcement.	Identified cost-effective benefits and increased forensic accuracy.
Thurzo et al.	2021	To explore 3D CNNs in forensic medicine.	Review of deep learning techniques for 3D image analysis.	Introduced full-head CBCT workflow and outlined ML applications.
Khanagar et al.	2021	To review ML use in forensic odontology.	Systematic review using QUADAS-2.	ANNs and CNNs show precision in dental-based identifications.
Ukwen and Karabatak	2021	To evaluate NLP in digital forensics.	Survey of NLP systems in cybersecurity contexts.	Outlined roles, applications, and challenges of AI-NLP.
Manasa and Kumar	2022	To address adversarial ML threats in forensics.	Exploration of ML attacks and forensic countermeasures.	Presented state-of-the-art forensic tools for image analysis.
Solanke and Biasiotti	2022	To assess AI's role and standardization in digital forensics.	Review on evaluation and optimization tools for AI.	Stressed need for interpretability and legal admissibility.
Dunsin et al.	2022	To develop AI-based framework for digital investigations.	Implemented multiagent system with CBR.	Showed improved efficiency and integration with forensic tools.
Alnafrani and Wijesekera	2022	To propose ML for attack emulation in IoT forensics.	Used â^,ILP to automate forensic data generation.	Enabled rule-based evidence identification without attack execution.
Faqir	2023	To examine AIâ€™s role in legal and criminal systems.	Descriptive analysis from legal documents and literature.	Recommended ML for prediction, biometrics, and intelligent surveillance.
Gogia and Rughani	2023	To identify cybercrime challenges and need for AI.	Exploratory survey of experts and researchers.	Found strong need for automated ML solutions.

10.48047/jocaaa.2025.34.05.27

Obidimma and Ishiguzo	2023	To analyze ML use in Nigeriaâ€™s cybercrime investigations.	Review on skills, legal gaps, and ethical concerns.	Proposed reforms for legal integration and ML ethics.
Das et al.	2023	To discuss ML advancements in forensic investigation.	Conceptual study on ML methods and challenges.	ML enhances investigation but data complexity remains a challenge.
Vasilaras et al.	2024	To assess ML in mobile forensics.	Survey of ML tools and ML Alignment framework.	Highlighted need for transparency and robustness in tools.
Bokolo and Liu	2024	To explore ML in social media forensics.	Survey of NLP and GNN applications.	Outlined methods to detect cyberbullying and extremism.
Zangana and Omar	2025	To integrate ML in digital forensic techniques.	Review of ML tools and ethical concerns.	Promoted ML for scalable and ethical digital forensics.
Bansal et al.	2025	To study tech advancements in criminal investigations.	Analytical study of forensic technology impact.	Acknowledged AIâ€™s role and its ethical-legal concerns.

IV. Procedure Forensic analysis using machine learning methods

Step 1: Data Collection and Logging

Capture cyber forensic data from various sources such as logs, emails, network traffic, etc.

$$D = \{d_1, d_2, ..., d_n\}, \quad d_i \in \mathbb{R}^m$$

Where *D* is the dataset of *n* forensic events, each with *m* features (timestamps, IPs, ports, etc.)

Step 2: Feature Engineering and Selection

Select relevant attributes using mutual information or variance thresholding.

$$F' = \text{Select}(F) = \{f_1, f_2, ..., f_k\}, \quad k \leq m$$

Step 3: Data Preprocessing and Normalization

Normalize or scale features for model compatibility.

$$x' = \frac{x - \mu}{\sigma}, \quad \forall x \in F'$$

Step 4: Labeling and Ground Truth Assignment

10.48047/jocaaa.2025.34.05.27

Assign labels for supervised learning (e.g., malicious = 1, benign = 0).

$$Y = \{yi \mid Ii \varepsilon \{0,1\}, i\text{-}1,2,\ldots\ldots n\}$$

Step 5: Model Selection

Choose appropriate ML model MMM for classification or anomaly detection.

$$M \varepsilon \{ SVM, RF, NN, KNN, XGBoost\}$$

Step 6: Model Training

Train the model using feature set *F′* and labels *Y*

$$\boldsymbol{\Theta^* = \text{Arg min}\ \mathcal{L}}\ (M(F', Y; \boldsymbol{\Theta}))$$

Where $\mathcal{L}$ is the loss function (e.g., cross-entropy for classification)

Step 7: Model Evaluation

Evaluate the trained model using performance metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Step 8: Evidence Correlation and Timeline Analysis Cluster and link forensic events across logs.

$$G = (V, E), \quad E_{ij} = \text{Similarity}(d_i, d_j)$$

Where *G* is a graph modelling correlation between evidence nodes.

Step 9: Anomaly and Pattern Detection

Use unsupervised ML for unknown threats

$$\text{Anomaly Score} = \|x_i - \mu_{\text{normal}}\|^2 > \delta$$

Step 10: Forensic Report Generation and Decision Support

Generate reports from model predictions

Report = Generate $(M(x_i)) \rightarrow$ Verdict, Evidence Trail, Source Attribution

10.48047/jocaaa.2025.34.05.27

V. Flow Chart

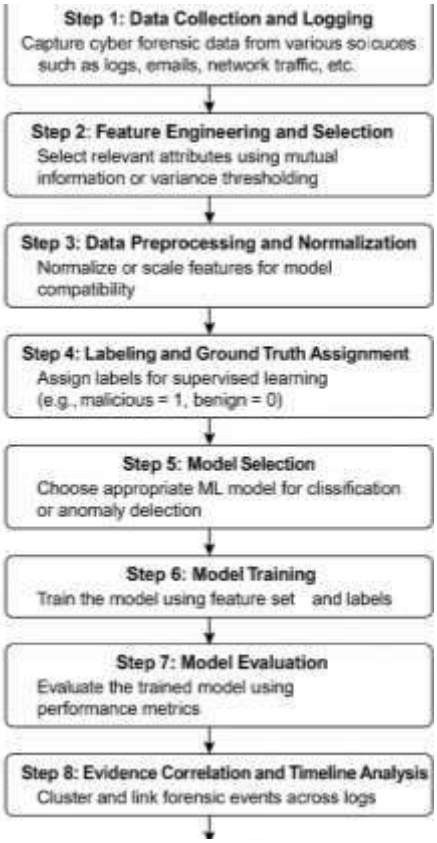

Fig. 2: Flow Chart of the forensic analysis process

The process of forensic analysis using machine learning methods involves a structured ten-step workflow that ensures the efficient identification, correlation, and interpretation of digital evidence. It begins with data collection and logging, wherein forensic data is extracted from diverse sources such as logs, emails, and network traffic. This raw data undergoes feature engineering and selection, which refines the dataset by choosing the most relevant attributes using statistical techniques. Following this, data preprocessing and normalization ensure uniformity in feature scaling, making it compatible with machine learning algorithms. Subsequently, labelling and ground truth assignment prepares the dataset for supervised learning by classifying data into categories (e.g., malicious, or benign). In model selection, suitable ML models like SVM, Random Forest, or XGBoost are chosen based on the problem type. The training phase then optimizes the model to fit the labelled data. Model evaluation uses performance metrics such as precision and recall to validate accuracy. Next, evidence correlation and timeline analysis link forensic events across time and space, while anomaly detection leverages unsupervised learning to uncover previously unknown threats. Finally, the system performs forensic report generation, summarizing outcomes with verdicts and evidence trails. This framework ensures robust, intelligent support for cybercrime investigations.

VI. Conclusion

The convergence of machine learning and digital forensics offers a paradigm shift in how cybercrime investigations are conducted and interpreted. This study has demonstrated that machine learning-driven methods

10.48047/jocaaa.2025.34.05.27

enhance the speed, scalability, and accuracy of forensic processes from feature engineering to final evidence reporting. As cyber threats evolve in sophistication, ML techniques such as natural language processing, neural networks, and 3D image analysis provide unmatched capability in detecting, classifying, and analysing complex datasets. Moreover, ML tools support forensic investigators in reconstructing events, predicting attack vectors, and identifying perpetrators with unprecedented precision. However, this integration is not without challenges. Issues related to algorithmic opacity, legal admissibility, and technological literacy among law enforcement professionals must be addressed. The research emphasizes the need for standardized protocols, ethical frameworks, and capacity-building to optimize the utility of ML tools in forensics. Ultimately, embracing machine learning in digital investigations marks a significant leap toward smarter, more resilient justice systems.

Reference

1. Iqbal, F., Debbabi, M., Fung, B. C., Iqbal, F., Debbabi, M., & Fung, B. C. (2020). Artificial intelligence and digital forensics. *Machine learning for authorship attribution and cyber forensics*, 139-150.
2. Jeong, D. (2020). Artificial intelligence security threat, crime, and forensics: taxonomy and open issues. *IEEE Access*, *8*, 184560-184574.
3. Mesejo, P., Martos, R., Ibáñez, Ó., Novo, J., & Ortega, M. (2020). A survey on artificial intelligence techniques for biomedical image analysis in skeleton-based forensic human identification. *Applied Sciences*, *10*(14), 4703.
4. Gupta, S., Sharma, M. V., & Johri, P. (2020). Artificial intelligence in forensic science. *International Research Journal of Engineering and Technology*, *7*(5), 7181-7184.
5. Jarrett, A., & Choo, K. K. R. (2021). The impact of automation and artificial intelligence on digital forensics. *Wiley Interdisciplinary Reviews: Forensic Science*, *3*(6), e1418.
6. Thurzo, A., Kosnáčová, H. S., Kurilová, V., Kosmeľ, S., Beňuš, R., Moravanský, N., ... & Varga, I. (2021, November). Use of advanced artificial intelligence in forensic medicine, forensic anthropology and clinical anatomy. In *Healthcare* (Vol. 9, No. 11, p. 1545). MDPI.
7. Khanagar, S. B., Vishwanathaiah, S., Naik, S., Al-Kheraif, A. A., Divakar, D. D., Sarode, S. C., ... & Patil, S. (2021). Application and performance of artificial intelligence technology in forensic odontology–A systematic review. *Legal Medicine*, *48*, 101826.
8. Ukwen, D. O., & Karabatak, M. (2021, June). Review of NLP-based systems in digital forensics and cybersecurity. In *2021 9th International symposium on digital forensics and security (ISDFS)* (pp. 1-9). IEEE.
9. Manasa, S., & Kumar, K. P. (2022). Digital forensics investigation for attacks on artificial intelligence. *ECS Transactions*, *107*(1), 19639.
10. Solanke, A. A., & Biasiotti, M. A. (2022). Digital forensics AI: evaluating, standardizing and optimizing digital evidence mining techniques. *KI-Künstliche Intelligenz*, *36*(2), 143-161.
11. Dunsin, D., Ghanem, M. C., & Ouazzane, K. (2022). The use of artificial intelligence in digital forensics and incident response (DFIR) in a constrained environment.
12. Alnafrani, R., & Wijesekera, D. (2022, June). AIFIS: Artificial intelligence (AI)-based forensic investigative system. In *2022 10th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.
13. Faqir, R. S. (2023). Digital criminal investigations in the era of artificial intelligence: A comprehensive overview. *International Journal of Cyber Criminology*, *17*(2), 77-94.
14. Gogia, G., & Rughani, P. (2023, October). Is Artificial Intelligence-Based Automation for Cybercrime Investigation the Need of the Hour? An Exploratory Study. In *The International Conference on Recent Innovations in Computing* (pp. 361375). Singapore: Springer Nature Singapore.
15. Obidimma, E. O., & Ishiguzo, R. O. (2023). Artificial Intelligence And Cybercrime Investigation in Nigeria: Addressing the Legal and Technical Skills Gaps. *African Journal Of Criminal Law And Jurisprudence*, *8*.
16. Das, S. S., Patnaik, S., Pattnayak, P., & Mohanty, A. (2023, December). An Advancement in Forensic and Criminal Investigation through Artificial Intelligence. In *2023 OITS International Conference on Information Technology (OCIT)* (pp. 920-926). IEEE.
17. Vasilaras, A., Papadoudis, N., & Rizomiliotis, P. (2024). Artificial intelligence in mobile forensics: A survey of current status, a use case analysis and ML alignment objectives. *Forensic Science International: Digital Investigation*,

10.48047/jocaaa.2025.34.05.27

49, 301737. 18. Bokolo, B. G., & Liu, Q. (2024). Artificial intelligence in social media forensics: A comprehensive survey and analysis. *Electronics*, *13*(9), 1671.

19. Zangana, H. M., & Omar, M. (2025). Introduction to Digital Forensics and Artificial Intelligence. In *Digital Forensics in the Age of ML* (pp. 1-30). IGI Global Scientific Publishing.

20. Bansal, S., Nayak, S. S., & Dave, I. (2025). The Role of Forensic Science and Digital Technology in enhancing Investigation Efficacy: An Analytical Study. *Cuestiones de Fisioterapia*, *54*(2), 2566-2592.