International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Use of Machine Learning in Digital Forensics

Khushi Sharma[1], Mayur Dattatray Mali[2], Sunita Shinde[3] , Saraswati Ghodke[4], Ruturaj Humane[5]

[1]Asst. Prof. , School of Science, G. H. Raisoni International Skill Tech University, Pune
[2] Asst. Prof. , School of Science, G. H. Raisoni International Skill Tech University, Pune
[3] Asst. Prof. , School of Science, G. H. Raisoni International Skill Tech University, Pune
[4] Asst. Prof. , School of Science, G. H. Raisoni International Skill Tech University, Pune
[5] Asst. Prof. , School of Science, G. H. Raisoni International Skill Tech University, Pune

---

Abstract - Digital forensics has emerged as a critical field in the investigation of cybercrimes, enabling the identification, preservation, and analysis of digital evidence from various sources such as computers, mobile devices, networks, and cloud systems. However, the exponential growth in data volume and complexity of cyber-attacks has made traditional forensic methods insufficient. Machine Learning (ML), a subset of Artificial Intelligence, introduces automated and intelligent techniques that significantly enhance the efficiency and accuracy of digital forensic investigations. This paper presents an in-depth study of ML applications in digital forensics, including malware detection, intrusion detection, log analysis, and multimedia forensics. It also explores various ML techniques such as supervised learning, unsupervised learning, and deep learning models. Furthermore, the paper discusses implementation methodology, experimental approaches, advantages, challenges, and future directions. The integration of ML into digital forensics enables faster decision-making, improved scalability, and the ability to detect unknown threats, making it a transformative technology in cybersecurity.

***Key Words*:** Digital Forensics, Machine Learning, Cybercrime, Malware Detection, Intrusion Detection, Deep Learning.

1.INTRODUCTION

Digital Forensics is the scientific process of collecting, preserving, analyzing, and presenting digital evidence in a legally acceptable manner. With the rapid increase in internet usage, cloud computing, and mobile technologies, cybercrime has become more sophisticated and widespread. Cyber-attacks such as ransomware, phishing, identity theft, and data breaches demand advanced forensic techniques for investigation.

Traditional forensic methods rely heavily on manual analysis, which is time-consuming and prone to human error. Moreover, the massive volume of data generated daily makes it difficult to process evidence efficiently.

Machine Learning (ML) provides a solution by automating data analysis and identifying hidden patterns. ML algorithms can process large datasets, detect anomalies, and improve the speed and accuracy of investigations. The integration of ML into digital forensics represents a significant advancement in combating cybercrime.

2. BACKGROUND AND FUNDAMENTALS

2.1 Digital Forensics Process

The standard forensic investigation process includes:

- Identification of evidence
- Preservation of data integrity
- Collection of digital data
- Examination and analysis
- Presentation in court

Maintaining the chain of custody is critical to ensure the integrity and admissibility of evidence.

2.2 Machine Learning Overview

Machine Learning is a technique that enables systems to learn from data without explicit programming. It involves:

- Training phase
- Testing phase
- Prediction phase

ML models improve performance over time by learning patterns from historical data.

3. LITERATURE REVIEW

Several researchers have explored the application of Machine Learning in digital forensics:

- Sommer and Paxson (2010) highlighted the role of ML in intrusion detection systems.
- Garfinkel (2010) discussed challenges in digital forensic investigations due to large-scale data.
- Behl & Behl (2017) emphasized the importance of AI in cybersecurity.

Existing research shows that ML significantly improves detection accuracy but faces challenges such as lack of datasets and interpretability issues.

4. MACHINE LEARNING TECHNIQUES

4.1 Supervised Learning

Supervised learning uses labeled datasets to train models. It is widely used in classification tasks.

Algorithms:

- Decision Tree
- Support Vector Machine (SVM)
- Naïve Bayes
- Random Forest

Use Cases:

- Spam detection
- Malware classification
- Fraud detection

4.2 Unsupervised Learning

Unsupervised learning identifies patterns in unlabeled data.

Algorithms:

- K-Means Clustering
- DBSCAN
- Hierarchical Clustering

Use Cases:

- Anomaly detection
- Network behavior analysis

4.3 Deep Learning

Deep learning models are capable of handling complex and high-dimensional data.

Models:

- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)

Use Cases:

- Image forensics
- Video analysis
- Deepfake detection

4. APPLICATIONS OF MACHINE LEARNING IN DIGITAL FORENSICS

4.1 Malware Detection and Analysis

ML algorithms analyze file behavior and detect malicious software. They classify malware into different categories, improving threat response time.

4.2 Network Intrusion Detection

ML models monitor network traffic and detect suspicious activities such as unauthorized access and data breaches.

4.3 Log Analysis and Event Correlation

System logs contain valuable forensic data. ML helps in identifying patterns and correlating events to detect cyber incidents.

4.4 Mobile Forensics

ML techniques are used to analyze data from smartphones, including messages, call logs, and app activity.

4.5 Multimedia and Deepfake Forensics

Deep learning models detect manipulated images and videos, which is crucial in preventing misinformation and fraud.

5. METHODOLOGY

The general workflow for applying ML in digital forensics is:

1. Data Collection: Gathering data from devices, networks, and logs
2. Data Preprocessing: Cleaning and normalizing data
3. Feature Extraction: Identifying relevant attributes
4. Model Training: Applying ML algorithms
5. Evaluation: Measuring accuracy, precision, recall
6. Deployment: Using the model in real-world forensic systems

6. EXPERIMENTAL ANALYSIS

A sample experiment can be conducted using datasets like:

- KDD Cup dataset (for intrusion detection)
- Malware datasets

Evaluation metrics include:

- Accuracy
- Precision
- Recall
- F1-Score

Results show that ML-based systems outperform traditional rule-based systems.

7. ADVANTAGES OF MACHINE LEARNING IN DIGITAL FORENSICS

- Automation of forensic processes
- Faster analysis of large datasets
- Improved accuracy and efficiency
- Ability to detect unknown threats
- Scalability for big data environments

8. CHALLENGES AND LIMITATIONS

- Lack of standardized datasets
- High computational requirements
- Risk of false positives/negatives
- Legal issues in court admissibility
- Black-box nature of ML models

9. FUTURE SCOPE

- Integration of AI with forensic tools
- Real-time forensic analysis systems
- Use of blockchain for evidence integrity
- Development of explainable AI models
- Advanced deep learning techniques
-

Fig -1: Machine Learning Based Forensic Framework

Table -1: Comparison of ML Techniques

Technique	Advantage	Limitation
Supervised ML	High accuracy	Requires labeled data
Unsupervised ML	Detects unknown threats	Less accurate
Deep Learning	Handles complex data	High computation cost

10. CONCLUSIONS

Machine Learning has transformed digital forensic investigations by providing automated, accurate, and scalable solutions. It enhances the ability to detect cyber threats and analyze digital evidence efficiently. Despite challenges, ML continues to evolve and holds great promise for the future of digital forensics.

11. REFERENCES

1. Casey, E. (2011). Digital Evidence and Computer Crime.
2. Garfinkel, S. (2010). Digital Forensics Research.
3. Sommer, R., & Paxson, V. (2010). Machine Learning in Intrusion Detection.
4. Behl, A. (2017). Cybersecurity and Cyberwar.
5. NIST Digital Forensics Reports.
6. IEEE Papers on Machine Learning and Cybersecurity.