SJIF Impact Factor: 6.004

Date: 14-Jun-2026

International Journal of Advances in Engineering & Scientific Research

International Peer Reviewed & Refereed Journal as per UGC New Rules

Certificate of Publication

This is to certify that the paper

"A HYBRID ARTIFICIAL INTELLIGENCE AND CRYPTOGRAPHY FRAMEWORK FOR SECURE DOCUMENT MONITORING"

by

Dr. H. Jayamangala,
Assistant Professor, DCA PG, VISTAS, Chennai, India
Email: jayamangala.scs@vistas.ac.in

has been published in ARSEAM's Journal

"International Journal of Advances in Engineering & Scientific Research", Volume 13, Issue 01, -2026.

The Journal has been registered and assigned ISSN (Online): 2349 –3607 and ISSN (Print): 2349-4824

Managing & Publishing Editor

Academic Research in Science, Engineering, Art & Management (ARSEAM) Foundation

ARSEAM foundation aims to publish research paper, article, case study and book review of very high quality intended to serve as an outlet for theoretical and empirical research in all areas of Science, Engineering, Art and Management contributing to the advancement in the existing literature and are very significant in managerial implications.

For detail, please visit us at: www.arseam.com contact us at Email: info@arseam.com

International Journal of Advances in Engineering & Scientific Research
ISSN (O): 2349 –3607, ISSN (P): 2349 –4824
An International Open Access, Peer Reviewed, Refereed Journal
DOI: https://doi.org/ 10.5281/zenodo.22267474
Journal homepage: https://arseam.com/journal?title=IJAESR
Peer Reviewed Journal
open access

A HYBRID ARTIFICIAL INTELLIGENCE AND CRYPTOGRAPHY FRAMEWORK FOR SECURE DOCUMENT MONITORING

Dr. H. Jayamangala,
Assistant Professor, DCA PG, VISTAS, Chennai.
Email: jayamangala.scs@vistas.ac.in

Article History

Received on: 11 Mar 2026
Revised on: 18-Apr-2026
Accepted on: 30- May 2026
First Published: 14-Jun-2026

Cite this Article as:

H. Jayamangala (2026), "A Hybrid Artificial Intelligence and Cryptography Framework for Secure Document Monitoring", International Journal of Advances in Engineering & Scientific Research, Volume 13, Issue 1, 2026, pp 160-169. DOI : https://doi.org/10.5281/zenodo.22267474

ABSTRACT

As cyber threats, data security breaches and unauthorized access are becoming more prevalent in today's digital landscape, traditional security solutions are not capable of adequately protecting sensitive documents. To counteract the challenges faced by organizations, this proposal will develop a hybrid security solution with next-generation capabilities by combining Natural Language Processing (NLP), Elliptic Curve Cryptography (ECC), and Honeypot-based monitoring systems. Using NLP, a unique fingerprint for each document will be generated based upon intelligent analysis, classification, and the creation of a unique signature or fingerprint for each document. Using ECC will require strong encryption and digitally signed documents as part of the cryptographic process while also allowing for efficient key management. Using NLP to create realistic decoy documents that can be used as Honeypots will further provide an opportunity to mislead and trap prospective attackers. Because any interaction with the decoy document will log the intruder's IP address, access attempts, and other pertinent data through the use of an Intrusion Detection System (IDS). By integrating intelligent content analysis, proper encryption, and a deceptive defence strategy, the hybrid model will both protect unauthorized access to sensitive data as well as enable proactive threat detection and forensic investigation capabilities. Overall, this proposal will establish a scalable, flexible, and highly secure strategy for safeguarding critical digital documents in a high-risk environment

***Keywords**: NLP, Elliptic Curve Cryptography (ECC), Document Security, Honeypot Monitoring, Intrusion Detection, Data Encryption, Digital Fingerprinting.*

1. INTRODUCTION

As technology has increased rapidly, so too have threats associated with protecting sensitive documents from cyber-attacks, data breaches, and unauthorized access to sensitive data; therefore, traditional security systems (e.g., passwords, firewalls, basic encryption) no longer provide adequate protection against sophisticated attacks or changing types of attack. As a result, organizations in all industries including healthcare, finance, and defence require advanced solutions that are capable of protecting their information while at the same time

detecting and responding to organized crime activity in real-time. This necessity necessitated development of intelligent, multi-layer security frameworks that utilize the latest technologies together to provide enhanced protection. The proposed solution introduces a new hybrid approach using Natural Language Processing (NLP), Elliptic Curve Cryptography (ECC), and honeypot monitoring to create a new level of document security for businesses. NLP techniques will be utilized to analyse and classify document contents and create a unique set of fingerprints that will facilitate maintaining the integrity of data stored in the cloud or on premise, and will also allow for detection of unauthorized modifications to documents. Documents secured using ECC will be encrypted using strong encryption algorithms with the ability to easily verify digital signatures; this will provide secure storage of documents, as well as secure transmission of documents. Lastly, using decoy documents and honeypots (virtual deception) to mislead hackers will create greater awareness of the types of attacks hackers may use and provide more opportunities to capture their activity. Combining intelligent document analysis, strong encryption, and proactive intrusion detection will provide ultimate protection against modern cyber threats.

2. PROBLEM STATEMENT

Organizations are increasingly utilizing electronic documents to secure and communicate sensitive information due to the negative impacts of the digital environment, increasing the likelihood that they will be the target of cyber-attacks. The traditional means of securing these documents (e.g., static encryption, firewalls, and access control systems) do not adequately protect against modern cyber threats including insider attacks, data tampering, and advanced intrusion techniques; as such, they primarily focus on restricting unauthorized access, rather than on providing a mechanism to detect real-time changes at the content level or monitor malicious activity, or respond to an evolving threat landscape in real time. As a result of these shortcomings, sensitive information can still be modified, misappropriated, or leaked without being detected. Additionally, existing systems do not offer attack misdirection or entrapment methods that would impede an attacker's ability to directly access valuable information; therefore, there is no intelligence based method of detecting and tracing attackers or reviewing security incidents. Further, there is little or no integration between cryptographic protections and content analysis, which limits the ability of document security overall. A unified and comprehensive solution must be developed to provide secure storage and transmission of electronic documents, while incorporating intelligent document analysis, intrusion detection, and tracking of attackers. Enhancing data protection through improved means of identifying cyber threats is crucial for maintaining trust in digital systems.

3. PROPOSED SYSTEM

This hybrid Encryption System uses a combination of Natural Language Processing (NLP),Elliptic Curve Cryptography (ECC) and Honeypot Monitoring to create one of the MOs comprehensive and intelligently protected Document Protection Systems to date. This System is built using multiple layers of security in order to secure both the Document and the transmission of the Document to you as an end-user. Initially, by requiring the User to authenticate themselves and implement Role Based Access Control (RBAC) to restrict Accessibility to you as an authorized user only by checking your credentials for authenticity through the use of Digital Signature technology. After the Document has been confirmed as authentic by the Authentication System, the Document is sent to be processed using NLP, which Classifies what type of content there is in the Document and then generates Unique Fingerprints of the Document to ensure Document Integrity, and to identify any Unauthorized Modifications made to it. The System also

uses ECC for encrypting and securing the storage and transmission of Documents, and generating Digital Signatures to confirm that the Document is authentic and has not been altered in any way during the transmission process. Additionally, this Document Protection System generates realistic "fake" Documents (Decoy) by using the same type of NLP techniques, so that the attacker who tries to access a Sensitive Document will be misled by accessing a Decoy Document instead of the actual Sensitive Document. Whenever a hacker or intruder interacts with a Decoy Document, the System flags this activity as Suspicious and sends an Alert to the designated administrator. In addition to the above mentioned measures being employed by the Document Protection System, the intrusion detection module continuously monitors the Actions of the System, and logs pertinent information such as: the IP address, time and date of the "suspicious" Action, and any other Relevant User Activity.

KEY FEATURES:

- Multi-layered Security that includes Natural Language Processing, ECC and Honeypots · Intelligent Document Classification/Cognitive Identification (Fingerprinting).
- Real-time Detection of Unauthorized Document Modification
- Strong Encryption via Ecliptic Curve Cryptography (ECC)
- Digital Signatures provide Signature Verification = Authorization& Integrity
- Role Based User Authentication/Secure Access Control is Enforced
- Automated Generation of Decoy Documents to Deceive Attackers
- Honeypot Based Attack Detection and Monitoring
- Real-time Alerts for Intrusion Detection · Logging of Attacker IP Address / Activity for Forensic Investigation

4. SYSTEM ARCHITECTURE

The system architecture is designed as a multi layered framework that integrates security, intelligence, and monitoring components to ensure comprehensive document protection. It consists of five major layers working together in a structured flow.

At the first layer, the User Interface and Authentication Module handles user registration, login, and role-based access control. Only authorized users can access the system based on their credentials and permissions.

The second layer is the NLP Processing Layer, where uploaded documents are analysed. This layer performs document classification, keyword extraction, and generates unique fingerprints using text features and embedding to ensure content integrity.

The third layer is the Cryptographic Security Layer, which uses Elliptic Curve Cryptography (ECC) to encrypt documents and generate digital signatures. This ensures secure storage and safe transmission of data.

The fourth layer includes the Decoy and Honeypot Module, which creates realistic fake documents to trap attackers. These decoy files act as a defensive mechanism to detect malicious intent.

The final layer is the Intrusion Detection and Monitoring Module, which continuously tracks system activities. It logs attacker IP addresses, timestamps, and suspicious actions, enabling real time alerts and forensic analysis.

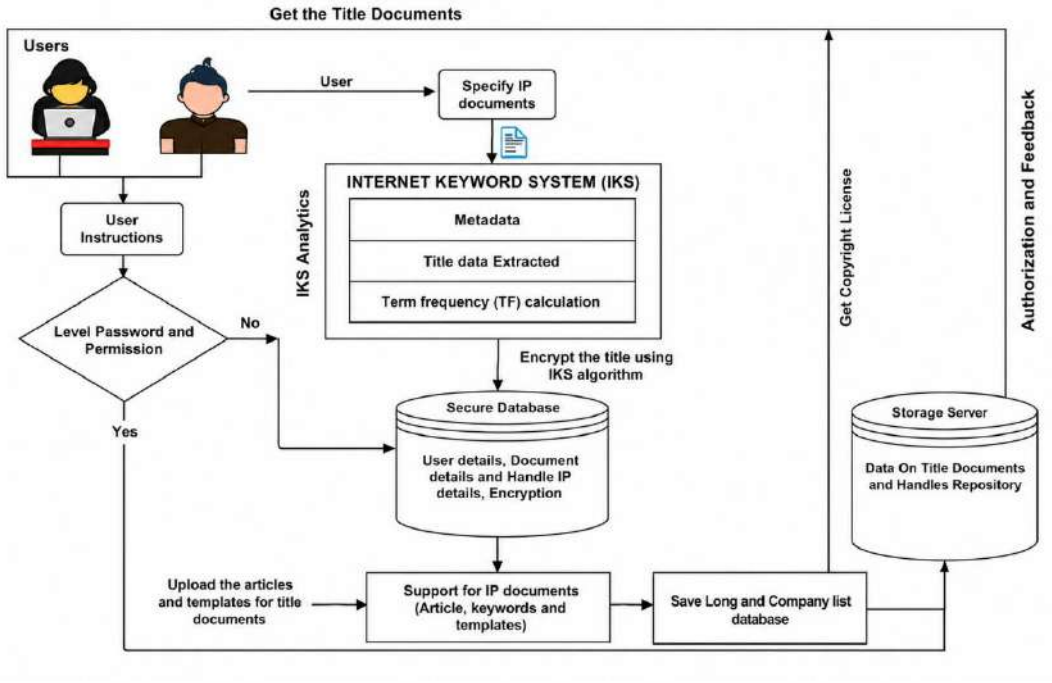

Figure 1: Diagram representation of the proposed methodology

5. MODULE ANALYSIS

USER AUTHENTICATION AND ACCESS CONTROL

This validates user identity when accessing a secure documents system. The module ensures that only users who have been authorized to access, modify, and manage sensitive documents are allowed to do so within the framework of the document system. Users must register and log in to the document system with valid credentials, and then the document system will verify the user's credentials before granting the user access to the document system. Furthermore, this module implements role-based access control to define different levels of access for different types of users (administrator vs. regular users). The modulealso implements session management so that unauthorized use does not occur during an active session. Password encryption and secure authentication further protect user accounts from unauthorized use. By monitoring and controlling user access to secure documents, this module provides the first line of defence in protecting confidential data.

NLP-BASED DOCUMENT CLASSIFICATION AND FINGERPRINTING

This employs Natural Language Processing techniques to document analysis and classifying based on that analysis. By extracting key terms, phrases, and context from documents, the module classifies each document as a different security level. As part of this process, the module creates a unique fingerprint for each document based on text features and word embedding. The fingerprints allow detection of any changes to the original document. If the content is changed, the module compares the changed content with the original document's fingerprint and identifies any discrepancy immediately. This process maintains the integrity of documents and preserves the integrity of sensitive data. This module enhances overall document management processes by enabling intelligent content analysis.

ELLIPTIC CURVE CRYPTOGRAPHY (ECC) ENCRYPTION AND DIGITAL SIGNATURE

Protecting documents with advanced cryptography techniques is the goal of this module. With ECC (Elliptic Curve Cryptography), documents will be encrypted before storing or transmitting them within the system. ECC offers a high level of security while using a smaller key size; therefore, ECC is an efficient and effective method of applying cryptography within today's modern secure computing systems. When documents are encrypted, they are only able to be decrypted by users who have been authorized to have access to that form of encryption through the use of a valid decryption key. When an encrypted document is signed digitally, the signature serves as a means to verify both the authenticity and integrity of the document. If there is any attempt to alter or tamper with the encrypted data, the digital signature will become invalid, which will alert the system that there has been a tampering incident. This module provides a means to store documents securely; communicate document information securely; and protect against unauthorized access to data.

DECOY DOCUMENT GENERATION AND HONEYPOT DEPLOYMENT

In this, misleading attackers will occur through the generating of believable decoy documents, which would consist of accurate contents and structure (i.e., the same physical appearance) and have no actual sensitized information. When an unauthorized individual accesses or digitises one of these files, the system would log such activity as suspicious. The use of a honeypot to trap and observe attackers allows for the preservation of the actual system data, therefore protecting the organisation from attack while diverting cybercriminals from legitimate documents or assets. Activities performed in these decoy files are logged for analysis and security; consequently, the use of deception as a means of prevention through this module will facilitate identifying and deflecting possible cyber threats.

INTRUSION DETECTION AND ATTACKER IP LOGGING

This module provides continuous monitoring of activity on systems to detect illegal or unauthorised access attempts. The module generates an intrusion alert whenever a user either interacts with decoy files or is found to be bypassing security controls. It automatically records specific information about the attack, such as the attacker's IP address, the time of access and any activity logs related to that user's actions. All of that recorded data is kept in a secure location for use in forensic investigations and analyses. Continuous monitoring

enables administrators to detect and respond to security incidents as soon as they occur. Data collected can also be analysed to help formulate strategies to improve future security measures. The Module is very important for tracking down the attackers and improving the overall cyber defence system.

6. DATA BASE ANALYSIS

In this proposal, the database serves as the main repository for all data needed for the proposed system. It is a secure location to hold and manage essential data such as users' identification, documents, fingerprints, encryption keys, and logs of any intrusion events. A structured relational database will be used for optimal organization, retrieval, and security of all data. For example, the User Table stores user credentials, roles, and authentication information, including users' encrypted passwords, which prevents unauthorized access. The Document Table maintains the metadata for each document stored in the system, such as document identification, owner, level of classification, and location; however, the actual document is stored in an encrypted format. The Fingerprint Table contains NLP-generated fingerprints for the verification of document integrity and to detect any unauthorized alterations. The Key Management Table contains encrypted ECC public and private keys that are used for performing the encryption and generating thedigital signatures. The Decoy Document Table provides a means for managing honeypot files as well as providing tracking of any type of interaction with those files. Lastly, the Intrusion Log Table maintains a record of any suspicious activities, including IP address, time/date stamped, and user activity, for use in the future for forensic analysis. Appropriate methods of data normalization, indexing, and access control have been implemented to improve the performance of the database and maintain the security of the data. The structure of the database will enable real-time monitoring and secure storage while also providing system efficiency.

7. FEASIBILITY ANALYSIS

Technical Feasibility

The proposed system is technically feasible as it utilizes well-established technologies such as Natural Language Processing (NLP), Elliptic Curve Cryptography (ECC), and database management systems. These technologies are supported by modern programming languages and frameworks, making development and integration easier. The system can be implemented using existing software tools without requiring specialized hardware. Additionally, the availability of libraries for NLP and cryptography ensures efficient system development and deployment.

Economic Feasibility

From an economic perspective, the system is cost effective since it mainly relies on open-source technologies and standard computing infrastructure. There is no need for expensive hardware or proprietary software, which reduces overall implementation costs. Moreover, the system helps in preventing data breaches and cyber-attacks, thereby avoiding potential financial losses for organizations. Maintenance and operational costs are also minimal.

Operational Feasibility

The system is operationally feasible as it is designed to be user-friendly and easy to manage. Users can easily upload, access, and manage documents through a simple interface.

Automated processes such as document classification, encryption, and intrusion detection reduce manual effort. The system requires minimal training for users and administrators, making it suitable for real-world deployment.

Legal Feasibility

The proposed system complies with standard data protection and cybersecurity regulations. It ensures secure handling of sensitive information through encryption and access control mechanisms. The system can be adapted to meet organizational and legal requirements, making it suitable for deployment in sectors like healthcare, finance, and government.

8. PERFORMANCE ANALYSIS

Classificati on Module	Precision (%)	Recall (%)	F1- Score (%)	Accuracy (%)
NLP-Based Classifier	92%	90%	91%	93%
Keyword Based Model	88%	85%	86%	89%
Hybrid Model	94%	92%	93%	95%

Table 1: Performance Evaluation of Classification Module

9. SECURITY ANALYSIS

A multi-layered strategy combining several security measures (including authentication, intelligent content analysis, encryption, and intrusion monitoring) provides high security. Authentication and role-based access control prevent unauthorized access by users to sensitive documents. A fingerprinting mechanism based on natural language processing (NLP) will ensure the integrity of documents while also identifying at all times any alterations to those documents made by unauthorized users. Documents are encrypted and keys are managed securely using elliptic curve cryptography (ECC). The use of an electronic signature gives confidence in the authenticity of documents and prevents tampering with them. Decoy documents and honeypot deployment provide a proactive means of defending against attacks by misleading the attacker and identifying the acts of an adversary as they occur.

10. ADVANTAGES OF PROPOSED SYSTEM

· Offers multi-layered protection using NLP, ECC and honeypot techniques.

· delivers robust data protection with sophisticated encryption solutions.

· Monitors for unauthorized changes to documents in real time.

· Enhances document integrity through the use of fingerprints.

· Creates confusion for attackers by using decoy documents.

· Provides real time intrusion detection and alerting systems.

· Collects information about attackers for forensic investigation.

11. FUTURE ENHANCEMENT

The proposed plan can include many additional technologies that would increase its performance, reliability, and security. Examples include adding Machine Learning (ML) and Deep Learning (DL) models into the system design to classify the documents more accurately and detect more complex patterns of attacks. Adding block chain technology for the decentralized and tamper-proof storage of document records would be beneficial. Introducing real-time behavioural analytics to monitor potential insider threats based on user behaviour patterns would also improve the overall performance of the system. Finally, expanding the system's ability to process documents in multiple languages through the use of state-of-the-art Natural Language Processing (NLP) technologies would increase the system's overall capabilities. Options for integrating the system with cloud platforms would increase its scalability and access to users. There are also a number of options for using different biometric methods (facial recognition and/or fingerprint scanning) to strengthen user verification. Finally, the system should include automated threat response mechanisms that can take immediate actions upon detection of potential attacks.

12. MATHEMATICAL MODEL

The system can be represented using a mathematical model as follows:

Let the system be defined as:

S = {U, D, F, E, H, I}

Where:

· U = Set of users

· D = Set of documents

· F = Set of document fingerprints

· E = Encryption function (ECC)

· H = Honeypot/decoy document set

· I = Intrusion detection system

Input:

User request (login, upload, access document)

Process:

1. Authentication:

Auth(U) → Valid / Invalid

2. Fingerprint Generation:

F(d) = NLP(d)

3. Encryption:

Enc(d) = E(d, key)

4. Integrity Check:

If F(d1) ≠ F(d2) → Modification Detected

5. Honeypot Trigger:

If access ∈ H → Intrusion Alert

6. Intrusion Logging:

I = {IP, Time, Activity}

Output:

- Secure document access
- Intrusion alerts
- Logged attacker information

13. PERFORMANCE COMPARISON

Metric	Existing System (%)	Proposed System (%)
Precision	78%	92%
Recall	75%	90%
F1-Score	76%	91%
Accuracy	80%	94%

Table 2: Performance Comparison betweenExisting Methods and Proposed System

PERFORMANCE IMPROVEMENT ANALYSIS

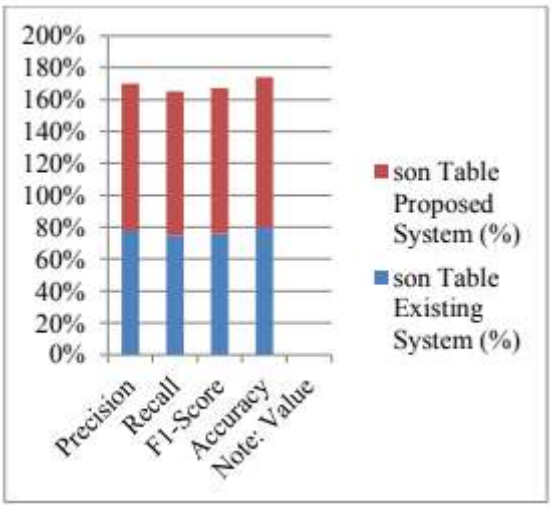

Figure 2: Performance comparison chart for existing and proposed values

14. CONCLUSION

The current system uses traditional security mechanisms which include simple types of encryption and access controls that do not offer complete protection from all types of attacks or modifications to content, as they do not have the ability to detect real-time threats or changes. The current system also does not support intelligent monitoring or tracking of attackers, leaving it open to advanced types of cyberattacks and data breaches. In contrast, the proposed system has significantly better performance features with the addition of NLP-based document classification methods and fingerprinting methods as well as ECC encryption, which allows it to detect unauthorized changes in real-time and provide greater integrity of data through improved key management than the existing configuration. The use of honeypot-based decoy documents, alongwith the intrusion detection module, will allow for the identification of and misdirection of attackers from the source of the real documents through real-time alerts and logged details to be analysed.

REFERENCES

[1] Pagnotta, Giulio, et al. "Dolos: A novel architecture for moving target defense." IEEE Transactions on Information Forensics and Security (2023).

[2] Sayeed, Sarwar, et al. "TRUSTEE: Towards the creation of secure, trustworthy and privacy preserving framework." Proceedings of the 18th International Conference on Availability, Reliability and Security. 2023.

[3] Ajmal, Abdul Basit, et al. "Toward effective evaluation of cyber defense: Threat based adversary emulation approach." IEEE Access 11 (2023): 70443-70458.

[4] Martínez, Antonio López, Manuel Gil Pérez, and Antonio Ruiz-Martínez. "A Comprehensive Model for Securing Sensitive Patient in a Clinical Scenario." IEEE Access 11 (2023): 137083-137098.

[5] Gambarelli, Gaia, Aldo Gangemi, and Rocco Tripodi. "Is your model sensitive? SPEDAC: A New resource for the automatic classification of sensitive personal data." IEEE Access 11 (2023): 10864-10880.

[6] Lansari, Mohammed, et al. "When federated learning meets watermarking: A comprehensive overview of techniques for intellectual property protection." Machine Learning and Knowledge Extraction 5.4 (2023): 1382-1406.

[7] Li, Mingjie, Zichi Wang, and Xinpeng Zhang. "An effective framework for intellectual property protection of NLG models." Symmetry 15.6 (2023): 1287.

[8] Lederer, Isabell, Rudolf Mayer, and Andreas Rauber. "Identifying appropriate intellectual property protection mechanisms for machine learning models: A systematization of watermarking, fingerprinting, model access, and attacks." IEEE Transactions on Neural Networks and Learning Systems (2023).

[9] Awadallah, Abeer, et al. "Artificial intelligence-based cybersecurity for the metaverse: research challenges and opportunities." IEEE Communications Surveys & Tutorials (2024).

[10] Elbasi, Ersin, et al. "Robust and Secure Watermarking Algorithm Based on High Frequencies of Integer Wavelet Transform." 2024, 47th International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2024.