Strengthening Data Security in Bioinformatics with Machine Learning and DNA Encryption

Animesh Kairi
Department of Computer Science & Engineering, Institute of Engineering and Management, Kolkata, India
ani.kairi@gmail.com

Tapas Bhadra
Department of Computer Science & Engineering, Aliah University, Kolkata, India
tapas.bhadra@aliah.ac.in

Jannatul Ferdous
Department of Computer Science and Engineering, North Western University, Khulna, Bangladesh
aroni1815@cseku.ac.bd

Anindya Nag
Department of Computer Science and Engineering, Northern University of Business and Technology, Khulna, Bangladesh
anindyanag@ieee.org

Walid El-Shafai
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia | Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt
welshafai@psu.edu.sa

Reham Alsabet
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
ralsabet@psu.edu.sa (corresponding author)

Ahmad Taher Azar
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
aazar@psu.edu.sa

Received: 9 January 2026 | Revised: 29 January 2026, 13 February 2026, and 20 February 2026 | Accepted: 22 February 2026

 DOI: https://doi.org/10.48084/etasr.17429

ABSTRACT

The explosion of genomic data in bioinformatics has posed great challenges to data confidentiality, integrity, and secure data transmission. Traditional cryptographic techniques are robust but do not always fit well with the biological nature and size of genetic data. This study presents a novel framework for DNA-based cryptography, combined with Machine Learning (ML) for greater bioinformatics data security. The proposed framework combines chaotic DNA encoding with the Random Forest (RF) classifier, guaranteeing security in a real-time adaptive manner. The proposed method achieves strong security and efficiency, with an entropy of 7.99, an avalanche effect of 49.85%, a near-zero correlation coefficient (0.003), a high ML-based attack detection accuracy of 98.4%, and an average encryption/decryption time

of approximately 0.95 s. The results show improved resistance to common attacks and increased genomic data storage and transmission efficiency, outperforming conventional approaches such as AES and traditional chaotic DNA methods in terms of entropy, avalanche effect, correlation coefficient, and computational complexity, hence verifying its efficacy for real-world applications.

Keywords-DNA-based cryptography; bioinformatics; cryptosystem

I. INTRODUCTION

The increased use of bioinformatics tools in genomic sequencing and analysis raises significant security issues, particularly in the safe management of personal genomic information. With the exponential increase in the volume of data available to bioinformatics practitioners, in particular genomic sequences, data confidentiality, integrity, and resilience to cyber threats have become an exigent problem. As the use of sequencing technologies becomes less expensive and more common, enormous amounts of sensitive genetic data are produced daily [1]. This data, if it falls into the wrong hands or is tampered with, holds serious ethical, privacy, and even biomedical risks.

Traditional cryptographic algorithms, although adequate for standard types of data, are insufficient for the sheer complexity, sensitive nature, and intrinsic biological structure of genomic data. DNA-based cryptography, a paradigm inspired by biology, is a promising alternative that takes advantage of the natural properties of DNA molecules, such as high parallelism, data density, and combinatorial diversity, to ensure data storage and transmission [2]. In such systems, digital data is encoded in DNA-like sequences using four molecular genetic building blocks called nucleotides, Adenine (A), Thymine (T), Cytosine (C), and Guanine (G), followed by biological or chaotic mathematical operations [3].

Foundational research in DNA-based cryptography has investigated several different encoding schemes, hybrid encryption schemes, and biologically plausible computing techniques. For example, in a biologically inspired encryption scheme [4], random codon sequences were suggested to provide a more effective way to obfuscate encryption for resistance against frequency analysis. In [5], a strong DNA stenography method used a multi-level encoding structure to preserve security and data fidelity in unfriendly environments. Recent studies have presented hybrid encryption schemes using Machine Learning (ML) bio-inspired patterns [6]. These contributions greatly expand the theoretical and practical background of DNA-based cryptography.

Despite these advances, developing DNA-based schemes always faces the problem of adaptability, instant reactivity, and scalability within a cyber-physical environment, which can be addressed by ultra-secure storage and analysis of genetic data [7]. To mitigate these security threats, this study suggests a new ML-enhanced DNA-based cryptography framework that adapts encryption strategies in real-time based on observed threats and behavioral anomalies. The proposed system integrates Random Forest (RF), tracking session metadata, such as entropy, the frequency of accessing concepts, or time stamps. RF is very popular among supervised ML approaches, chosen for non-linear and chaotic patterns generated in DNA-based encryption, while remaining robust to noise and overfitting. Its ensemble structure also provides high accuracy and interpretable feature importance for reliable behavior monitoring. Detected anomalies lead to the reconfiguration of cryptographic parameters, making an attack (brute force, replay, side channel) on the system more difficult. In addition, the framework integrates SHA-3 hashing to perform data integrity validation and Chaotic DNA encoding mechanisms to achieve maximum confusion and diffusion properties. Performance evaluation examined entropy metrics, avalanche effect, correlation coefficient, execution time, and anomaly detection accuracy. Comparative studies with AES and state-of-the-art chaotic DNA cryptosystems demonstrate the superiority of the proposed approach in terms of security and computational efficiency [8]. This research is a contribution to providing a secure, adaptive, and biologically-based cryptographic model suiting the next generation of bioinformatics applications.

II. LITERATURE REVIEW AND RESEARCH MOTIVATION

DNA-based cryptography has become a new field in the domain of information security, with enormous potential for secure data transmission and storage. Many studies have investigated similar problems in the security and privacy of genomic information [9]. For example, in [10], ML and cryptography were combined to protect the privacy of biomedical data. In [11], a privacy-preserving Federated Learning (FL) algorithm (PPML-Omics) was presented for multi-omics data using decentralized shuffling in a differential privacy framework. The National Institute of Standards and Technology (NIST) is developing cybersecurity frameworks and threat models that are specific to genomic information [12]. Academic consortia, such as the UIUC-Mayo Center for Computational Genomics (CCBGM), emphasize encryption and secure data transmission as key requirements in genomic big data analysis. These initiatives, along with community projects, strengthen such efforts by addressing genomic privacy and encryption from different perspectives [13].

Several researchers have proposed cryptographic schemes that take advantage of the unique properties of DNA sequences and leverage substitution, transposition, and logic operations to obscure a plaintext into formats of biological origin [14]. These methods often exploit the nature of DNA complexity and high storage capacity, making them interesting for use in the next generation of security systems. In recent years, the use of chaotic systems has been introduced into DNA cryptographic schemes to further increase complexity [15]. By leveraging its inherent unpredictability and sensitivity to initial conditions in chaotic maps, researchers can improve confusion and diffusion properties, achieving stronger resistance against cryptanalytic attacks [16]. This has offered notable security enhancements, particularly for static encryption environments.

Despite all these developments, there is still a major drawback, as most of the existing DNA-based encryption systems are static and non-adaptive, lacking mechanisms for real-time threat recognition and dynamic key evolution. This is a major issue in modern computer security, where there is constant change in threats. Current DNA cryptographic methods do not leverage ML for adaptive key management, anomaly, and intelligent threat response, and hence lack applicability in scenarios requiring agility and context awareness. The implementation of ML into DNA-based cryptography could allow applications that dynamically learn from patterns of abnormal behavior and adjust their encryption logic to reflect this learning. This synergy has the potential to revolutionize the field, realizing self-adaptive, intelligent encryption models that are both biologically-inspired and computationally strong.

A. Research Problem

The explosive growth of genomic datasets in biomedical research and healthcare has posed a new and unprecedented security challenge. Genetic sequences contain, by nature, a sensitive component—if unauthorized access or manipulation of these data occurs, personal, or even biological, privacy is endangered [17]. Conventional cryptographic algorithms were not designed for such data, as they do not take advantage of the unique properties of biological sequences and may be inefficient on this scale [12]. Many DNA technologies were developed without strong security factors in mind, and it is only due to the increasing commercialization of genomics that the need for a security-by-design approach has become apparent to preserve the confidentiality and integrity of voluminous genomic data against modern cyber threats.

B. Research Gaps

Several studies have suggested DNA cryptography schemes based on substitution, transposition, and logic operations on DNA sequences. Recent works combined DNA-based cryptography and chaotic systems to enhance the confusion and diffusion effect and significantly improve security [18]. However, most systems are static, without the adaptability and real-time capabilities to mitigate threats. ML has been successfully applied in cyber defense systems and medical anomaly recognition, but its application for cryptography is not yet fully developed. Current approaches to DNA-based cryptography do not use ML to provide adaptive key management or detect anomalies and hence respond to threats, which hampers their ability to perform well in dynamic environments [19]. Despite notable improvements in advancing DNA-based cryptography, most existing frameworks are static and not adaptive in real-time. This gap in the adaptive key generation, smart threat response, and anomaly detection in DNA encryption systems restricts robustness, scaling, and situational awareness, especially in dynamic and/or real-time-based applications. Table I summarizes details on research gaps.

C. Research Motivation

Various factors make this problem urgent. First, genomic data breaches have serious implications: cybercriminals could track people down, learn about certain traits of their health, or blackmail their victims. Second, recent reviews emphasize that genomic databases are under threat of evolving risks (including advanced AI-based attacks and those related to future quantum computing), and hence need new defenses. Third, prominent surveys underscore the fact that existing genome tools and repositories do not have built-in security measures, and thus, it is easy for data to become accessible if appropriate new techniques are not applied. These considerations are the main motivation for this research on the need for a cryptography framework that is intrinsically design-specific to genomic data and capable of coping with threats. In particular, the combination of ML for real-time anomaly detection holds out promise for a proactive defense to match the dynamic nature of modern cyber-attacks. As the literature recommends an integrated technical approach to genomic privacy, the motivation for this study is to integrate the DNA-based encoding approach with intelligent and adaptive security, fulfilling these urgent requirements.

TABLE I. RESEARCH GAPS IN STATE-OF-THE-ART WORKS

Study	Year	Identified Research Gap
[16]	1949	Established theoretical secrecy foundations but does not address biological or molecular computing paradigms.
[1]	1995	Demonstrated DNA computing feasibility but did not explore security or cryptographic applications.
[8]	1999	Showed molecular cryptanalysis potential; lacks practical scalability and real-world DNA hardware implementation.
[10]	2000	Early DNA steganography; security analysis and resistance to modern attacks were not evaluated.
[11]	2003	Conceptual DNA cryptosystems; limited performance benchmarking and practical deployment strategies.
[18]	2008	Proposed DNA-based public-key system; lacks robustness analysis against quantum and AI-driven attacks.
[5]	2015	Focused on genomic privacy; does not integrate DNA cryptography or ML-based protection.
[19]	2019	Federated ML for privacy; lacks direct application to genomic encryption or DNA-based secure storage.
[12]	2023	Policy-level recommendations; no concrete DNA-based cryptographic implementation framework.
[13]	2023	Proposed DNA cryptosystem; requires stronger cryptanalysis and comparative security benchmarking.
[2]	2024	Focused on DNA storage reconstruction; does not address encryption robustness or attack resistance.
[7]	2024	Secure genomic storage platform; lacks integration with molecular-level cryptographic primitives.
[15]	2024	ML-based encryption; no biological/DNA-layer implementation or biochemical feasibility testing.
[4]	2025	Discusses AI & quantum risks; lacks experimentally validated DNA-ML hybrid cryptographic framework.
[14]	2025	Cloud encryption approach; no molecular encoding integration or genomic data-specific adaptation.
[17]	2025	Demonstrates vulnerabilities in DNA encryption; highlights need for stronger key scheduling and cryptanalysis-resistant design.

D. Research Contributions

This work makes progress in the security of genomic data by the following contributions:

- Adaptive DNA-based Encryption Framework: A new hybrid encryption pipeline for the semantics of transfer of plaintext and genomic sequences into chaotic strings of DNA-like sequences of nucleotides, and an access behavior monitor using an RF classifier. In this system, the ML module continuously analyzes metadata (e.g., entropy deviation, access timing, location) to detect anomalies. Detected threats immediately affect the cryptosystem, allowing keys and encoding parameters to be reshaped in real time.
- Dynamic Key Generation and Reconfiguration: Cryptographic keys are automatically regenerated, and the mapping between binary and DNA representations changes when malicious activity is detected. In this way, each new key is based on a new, fresh, chaotic entropy, which makes the cipher self-healing. This ensures that the system does not use a fixed key or static encoding tables. In this respect, the bar is considerably higher for a brute-force or replay attack to be successful than in traditional schemes.
- SHA-3 Integrity Verification: To prevent data interpretation errors, each piece of DNA ciphertext is prefixed by a SHA-3 hash. In the decryption of this hash, it is checked for tampering. The use of SHA-3 guarantees the detection of any unauthorized modification of the data, to complement the DNA-based obfuscation and thus offer both confusion and diffusion in the cryptosystem.
- Comprehensive Performance Evaluation: System security is thoroughly quantified using key metrics: Shannon entropy (i.e., ~7.99 bits/byte), avalanche effect (i.e., ~49.85%), and ciphertext bit correlation (i.e., ~0.003). These values show extremely high randomness and diffusion and surpass AES and previous methods based on chaotic molecular DNA. In addition, the accuracy of the ML classifier in the detection of anomalies is reported (~98.4%). Encryption and decryption of (typical) genomic records (1,000 base sequences) are fast (nearly 0.95~0.98 s) and prove practical. These benchmarks are based on experiments and demonstrate that the proposed scheme is a significant improvement over the lower bounds in terms of both security and efficiency.
- Comparative Analysis and Validation: The proposed system is compared with standard algorithms (AES) and recent DNA-based cryptography schemes under the same sort of conditions. Statistical analysis and visual comparisons prove that the hybrid system is significantly superior to its competitors in terms of both security measures and adaptability. In essence, the proposed framework is a combination of bio-inspired encoding and intelligent threat detection, filling a very important gap in the protection of genomic data.

III. PROPOSED METHOD FOR ML-DRIVEN DNA CRYPTOGRAPHY

A. Framework Overview

The proposed system is a hybrid architecture that combines DNA-based encryption and ML to offer a safe and flexible solution for bioinformatics data protection. It comprises the following main modules:

- DNA Cryptographic Layer: This layer is used to convert plaintext into DNA-like sequences using binary-to-nucleotide encoding rules. The system uses chaotic functions to input non-linear permutations and thereby add to the confusion and diffusion of the data. It performs encryption and decryption of information through inverse processes.
- Machine Learning (ML) Module: An RF is used to analyze the metadata (access time, entropy deviation, frequency of access, etc.) provided by behavioral data to determine anomalies in real time. It has a dynamic way of affecting cryptography's parameters, such as encoding rules and key structures, depending on the prediction scores for anomalies.
- Data Integrity and Communication Module: Provides the integrity of data with SHA 3 data hashing and secure socket protocols for communication. This component is used to authenticate the integrity of the encrypted data on both endpoints.

B. DNA Encoding and Decoding Process

DNA encoding/decoding is the cryptographic part of the system. This process is designed to simulate the structure of biological data, offering greater compatibility with the genomic database and making it harder to detect.

1) Encoding Procedure

- Input Conversion: Input of plaintext/genomic data is first converted to a stream of binary data. Example: "A", ASCII 65, Binary 01000001.
- Binary-to-DNA Mapping: The binary stream is split into 2-bit units. A nucleotide mapping rule, e.g., 00=A, 01=T, 10=C, 11=G, is applied. Example: 01000001, 01 00 00 01, T A A T.
- Chaotic Permutation: Chaotic logistic/Hénon map is used for randomizing the DNA sequence. It is one of the most widely used 1D chaotic maps in cryptography for generating pseudo-random sequences.

$$x_{n+1} = r x_n (1 - x_n) \quad (1)$$

where x_n is the current state (value between 0 and 1), and r stands for the control parameter. For the map to be useful in cryptography, r is typically set between 3.57 and 4.0. At $r = 4$, the map is fully chaotic and covers the entire range $(0,1)$. The permutation key generation is system-dynamic, generated based on system time and user entropy signature. This step helps prevent the frequency analysis by introducing pseudo-randomized DNA sequence positions.

- Key Embedding: An encoding cryptographic key (e.g., 128-bit binary) is further converted to DNA format and also embedded into the cipher as metadata, protected with a hashed integrity check.

Figure 1 shows a block diagram of the proposed DNA-based encryption method.

2) Decoding Procedure

- Key Extraction and Verification: The embedded DNA key is extracted, decoded to binary, and then hashed with SHA-3. The hash is matched with the original to see if any data has been tampered.
- Reverse Chaotic Map: Using an inverse version of the chaotic permutation matrix, the reordering of the shuffled DNA sequence is reversed to its original state.
- DNA-to-Binary Translation: The DNA letters are again translated back to the 2-bit binary digit chunks using the same rule set (i.e., A=00, T=01, etc.).
- Binary-to-Plaintext Conversion: The binary stream is converted back again into the character or genomics value. Figure 2 shows a block diagram of the proposed DNA-based decryption method.

C. Machine Learning Component

To further improve dynamic security, a supervised learning model with the RF algorithm is incorporated in the system. The model is trained on some features, such as:

- Entropy of Accessed Data,
- Access Timestamps,
- Geo-location of Request,
- Behavioral Patterns (e.g., session duration, frequency)

The classifier has high detection accuracy since it learns the threshold profiles of the legitimate and malicious access attempts. When a problem (i.e., a measure of entropy over some threshold) is detected, the system:

- Causes a new cryptographic key schedule to be started.
- Changes encoding rules (e.g., 00=A may change to 00=C).
- Triggers alerts, and the event is recorded.

This real-time adaptive behavior helps to beat the risks of brute-force, replay, and injection attacks. Figure 3 shows the system architecture diagram for the proposed ML-driven DNA cryptographic framework.

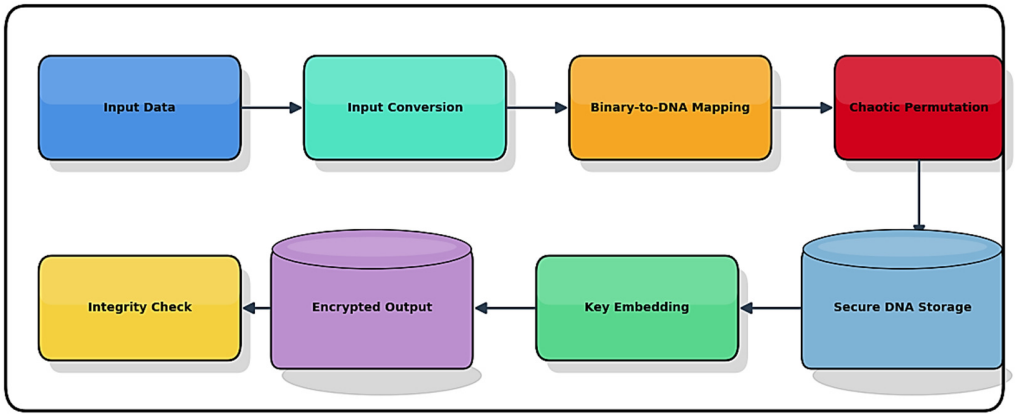

Fig. 1. Proposed encoding block diagram.

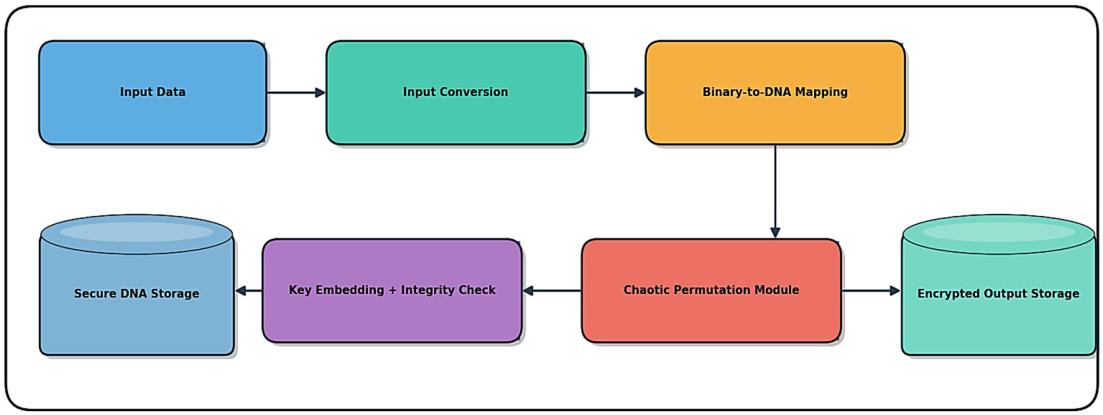

Fig. 2. Proposed decoding block diagram.

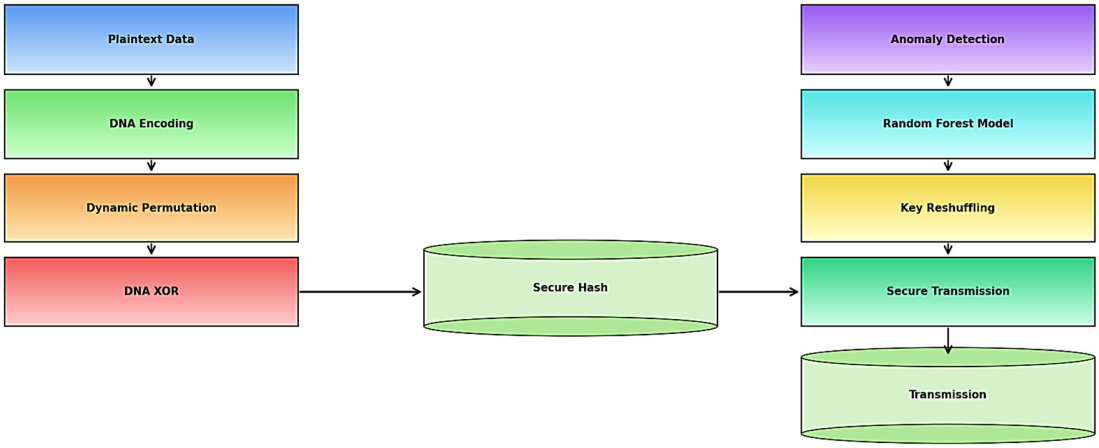

Fig. 3. Proposed system architecture.

D. Machine Learning Model Architecture and Configuration

1) Model Design

To realize adaptive encryption, the proposed encryption system includes a supervised ML model (RF classifier). This classifier is the analytical essence of the anomaly detection module and is responsible for cryptography during real-time adaptation according to observed user behavior. This architecture was selected to ensure robustness, interpretability, and low variance. The RF model is configured with:

- Number of Trees: 100
- Maximum Tree Depth: 20 (pruned to prevent over-fitting)
- Splitting Criterion: Gini Impurity
- Out-of-Bag Error Estimation: Enabled for internal validation
- Feature Selection: Performed automatically within the ensemble learning mechanism

2) Input Parameters and Feature Importance

The classifier uses metadata from genomic data access sessions to identify anomalous behaviors. RF was used to extract the most important input features, recording contextual behavioral signals that are easily manipulated in cyber-attacks, but are not looked at in static cryptographic systems.

3) Dataset and Partitioning

The ML model was trained on a synthetically augmented behavioral dataset, which had 1,000 overall session records. This dataset was experimental, did not contain any real user data, and comprised benign and malicious access patterns. The dataset was generated through a stochastic rule-based simulation framework that models behavioral attributes such as session entropy, geographic consistency, request frequency, and access regularity. Python programming was used for developing the synthetic behavioral dataset.

- Benign sessions: 500
- Anomalous sessions: 500 (including entropy injection, location spoofing, and session flooding)

The dataset was divided into training and test subsets in a ratio of 70:30, keeping the classes balanced by stratified sampling.

- Training set: 700 records
- Testing set: 300 records

E. Evaluation Metric

The following quantitative metrics were employed to evaluate the performance and resilience of the system:

- Entropy: Measures the randomness in the ciphertext. Higher values (close to 8 for 8-bit systems) indicate better security against frequency analysis.
- Avalanche effect: Assesses the sensitivity of the encryption algorithm to small input changes. A change in one bit should flip about 50% of the output bits.
- Correlation coefficient: Measures the linear correlation between adjacent pixels/bits in the plaintext and ciphertext. Values near 0 indicate high obfuscation.
- Execution time: The time taken for encryption and decryption, indicating the algorithm's efficiency and viability for real-time use.
- Anomaly detection accuracy: Measures the precision and recall of the ML model in flagging malicious sessions, which is critical for adaptive security.

IV. PERFORMANCE EVALUATION AND ML-BASED ANALYSIS

Experiments were carried out on a dataset of 1,000 genomic records. Table II presents a comparison with traditional AES and chaotic DNA encryption methods. High randomness improves resistance to attacks, the avalanche effect of nearly 50% indicates high diffusion, and the minimal correlation between adjacent bits confirms strong obfuscation. Finally, the ML detection accuracy of 98.4% confirms robustness against intrusions. Figure 4 illustrates a performance comparison, depicting the key metrics across the proposed, AES, and Chaotic DNA techniques.

TABLE II. COMPARISON OF EVALUATION METRICS

Metric	Proposed system	AES (Baseline)	Chaotic DNA
Entropy (bits/byte)	7.99	7.65	7.78
Avalanche effect (%)	49.85	47.2	48.5
Correlation coefficient	0.003	0.12	0.06
ML detection accuracy	98.40%	N/A	N/A
Avg. encryption time	0.98 s	0.65s	1.02s

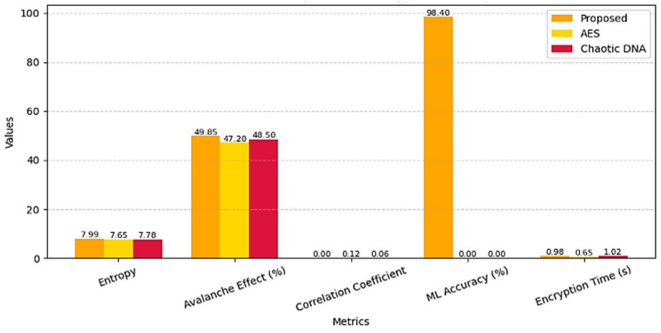

Fig. 4. Encryption performance comparison.

Figure 5 compares the decryption performance of the proposed method, AES, and chaotic DNA, showcasing the proposed method's strong consistency in both encryption and decryption phases, particularly in terms of entropy, diffusion, and anomaly detection readiness. Table III compares the encryption and decryption performance of the three techniques.

TABLE III. PERFORMANCE METRICS FOR ENCRYPTION VS DECRYPTION

Metric	Phase	Proposed system	AES (Baseline)	Chaotic DNA
Entropy (bits/byte)	Encryption	7.99	7.65	7.78
	Decryption	7.98	7.63	7.75
Avalanche effect (%)	Encryption	49.85	47.2	48.5
	Decryption	49.6	46.8	48.2
Correlation coefficient	Encryption	0.003	0.12	0.06
	Decryption	0.004	0.13	0.07
ML detection accuracy (%)	Encryption	98.4	N/A	N/A
	Decryption	98.1	N/A	N/A
Avg. time (s)	Encryption	0.98	0.65	1.02
	Decryption	0.95	0.64	1.01

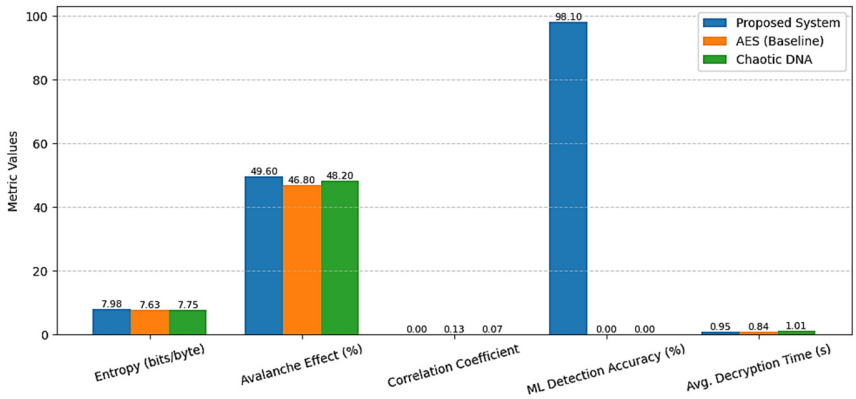

Fig. 5. Decryption performance comparison.

A. *Anomaly Detection Accuracy*

The RF model achieves high precision and recall in detecting abnormal access behavior using entropy fluctuations, session metadata, and access patterns (98.4% encryption, 98.1% decryption). The model can dynamically adapt encryption parameters based on real-time threat perception, offering protection from brute-force, replay, and injection attacks, features absent in AES and chaotic-only systems. Thus, ML models can provide behavioral security, not just data-level obfuscation.

B. *Feature Importance Analysis*

RF inherently ranks features. Table IV shows the feature importances extracted from the RF model. The system learns when and how users access data and identifies deviations early. Unlike static systems, the ML module contributes to context-aware encryption, tuning keys and mappings in response to behavioral cues.

TABLE IV. FEATURE IMPORTANCE IN THE RANDOM FOREST CLASSIFIER

Feature	Description	Normalized importance
Entropy deviation	Fluctuation in data randomness during access	0.32
Access timestamp variance	Irregularity in time intervals between sessions	0.24
Geo-location divergence	Unusual IP/location patterns during access	0.18
Session frequency profile	Frequency of access requests per unit of time	0.16
Request duration anomaly	Deviations in time taken per access	0.1

C. *Cryptographic Adaptation Triggers*

When anomalies are detected, the system:

- Changes the binary-to-DNA mapping rules dynamically (e.g., 00=A → 00=G).
- Regenerates keys using entropy-seeded chaotic maps, increasing resistance to key reuse attacks.

- Logs threats and enforces re-authentication, improving incident response**.**

This dynamic cryptographic behavior ensures that even if a static vulnerability is found, the system adapts before exploitation, something neither AES nor chaotic DNA methods can offer.

D. Trade-Off Analysis

Table V presents a trade-off analysis for the proposed system. Figure 6 provides a holistic comparison of three encryption systems across four key dimensions: Security, Speed, Intelligence, and Robustness, for both encryption (solid lines) and decryption (dashed lines). The proposed system demonstrates consistently high scores across all dimensions (5 out of 5 in Security, Intelligence, and Robustness), both in the encryption and decryption phases. This indicates its adaptive strength, context-aware intelligence, and resilience to evolving threats. AES scores well in Speed (5/5), showcasing its optimized computational performance, but lacks intelligence and adaptability. Although secure, its robustness is only medium due to its static nature. Chaotic DNA offers moderate performance. It benefits from biological randomness in Security and moderate Speed, but lacks learning capacity and fault recovery, leading to low Intelligence and Robustness. While AES is the fastest and chaotic DNA offers some security through confusion, the proposed ML-driven DNA cryptographic system provides the most balanced and future-ready solution for genomic data security, especially in dynamic or attack-prone environments.

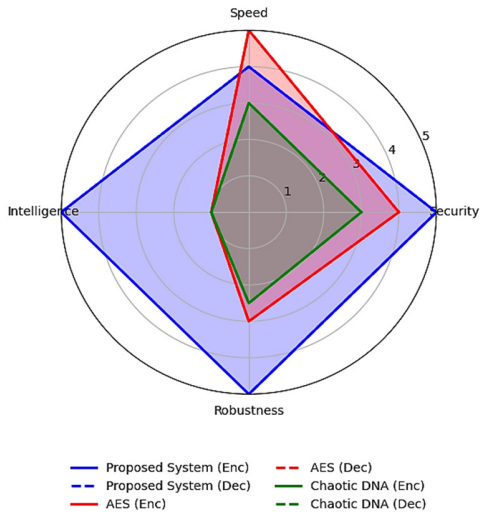

Fig. 6. Radar chart comparison.

TABLE V. TRADE-OFF ANALYSIS

Aspect	Proposed system	AES	Chaotic DNA
Security	Adaptive, real-time, context-aware	Strong but static	Confusion/diffusion-focused
Speed	~0.95–0.98 s	Fast (~0.64 s)	Moderate (~1.01 s)
Intelligence	High (ML-enabled decisions)	None	None
Robustness	High (self-healing encryption)	Medium	Medium-low

E. Comparative Analysis of Encryption and Decryption Time and Memory Usage

The graphical comparison of encryption and decryption performance reveals significant insights into computational efficiency and resource utilization (Figure 7). In the encryption phase, AES demonstrates the lowest execution time and memory usage, reflecting its highly optimized structure. The proposed system, while requiring slightly more time (0.98 s) and memory (~12.5 MB), achieves this with the added overhead of real-time ML-driven anomaly detection and dynamic cryptographic adaptation. Chaotic DNA, though more secure than AES, is slower (1.02 s) and consumes more memory (~10.9 MB), due to complex non-linear operations without intelligent adaptation. During decryption, a similar pattern emerges. AES remains the fastest and most memory-efficient, whereas the proposed system sustains its strong performance (0.95 s, ~12.3 MB) with intelligent behavior preserved even in the reverse cryptographic workflow. Chaotic DNA again shows marginally higher time and memory usage.

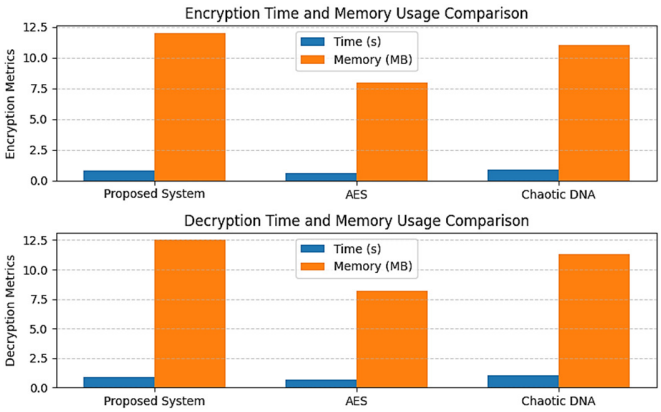

Fig. 7. Time and memory usage of crypto techniques.

V. DISCUSSION

A. Encryption and Decryption Performance

The proposed method succeeds in achieving an entropy of 7.99 bits/byte, which is more than AES (7.65) and chaotic-DNA (7.78), indicating a higher degree of randomness. Its avalanche effect reaches 49.85%, which is almost ideal (50%), again higher than AES (47.2%) and chaotic-DNA (48.5%). Critically, the cipher text bits obtained can be considered uncorrelated (correlation ~0.003), while AES and chaotic-DNA report correlations at 0.12 and 0.06, respectively. These results show that the proposed hybrid system generates significantly more unpredictable output for more resistance to statistical and differential attacks. In addition, the integrated ML module provides 98.4% detection accuracy for anomalies that occur during encryption, which is not available for comparison methods.

In terms of the key metrics (entropy, avalanche, and correlation, accuracy of anomaly detection, and encryption time), the proposed system is always leading in security-related measures. AES scores very high in speed (lowest time), but outside of randomness, and it does not have any anomaly

detection capabilities. Chaotic-DNA is moderate in security. Conversely, the proposed framework achieves the best balance of metrics under encryption, obviously outperforming the baselines in obfuscation, and brings another intelligence dimension (ML accuracy).

The metrics for the decryption phase are similar to the earlier encryption results. On decryption, the proposed scheme retains an entropy of ~7.98, avalanche ~49.6%, and correlation ~0.004, that is, it is virtually identical to encryption values. AES and chaotic-DNA also have slightly lower performance in decryption (e.g., AES entropy = ~7.63, avalanche =~46.8%), but still, there are gaps, as the proposed system always shows stronger confusion and diffusion for both phases. In summary, these results reveal robust and symmetric security improvements offered in both the encryption and decryption pipelines.

B. Machine Learning (ML) and Adaptation Analysis

RF plays an important role in improving security. It achieves an accuracy of approximately 98% in differentiating benign and malicious access sessions (based on entropy deviation, variation in access time, location change, session frequency, and duration of request). Entropy deviation proves to be the most influential (0.32), followed by timestamp variance (0.24) and reallocation divergence (0.18). This ranking is an indicator that the model has properly learned to distinguish normal access patterns and anomalies, such as a sudden spike in randomness or unusual access interval(s). When an irregular session is detected, the system instantly adapts its cipher: it changes the mapping from binary to DNA, regenerates the cryptographic key using fresh chaotic entropy, and registers the incident for further response. Such reconfiguration in real-time guarantees that if an adversary discovers a vulnerability, the encryption parameters change before they can become successful, which static AES or DNA-only schemes cannot do. Overall, the ML component offers context-aware security and is not merely data-level scrambling, providing great resilience in dynamic conditions. The inclusion of ML transforms the cryptographic scheme from data-focused to context-focused. Proactive encryption behavior (anticipating attacks) is enabled, improving long-term resilience. ML's continuous learning allows it to adapt to evolving threats, unlike traditional methods, which require manual updating or redesign.

C. Trade-Offs and Visual Comparisons

The proposed method reaches its highest score in Security, Intelligence, and Robustness (5 out of 5) for encryption and decryption, which attest to the adaptive and self-healing qualities of the design. AES scores highest in the Speed category (5/5) and lowest in the Intelligence category (no learning capacity). The chaotic DNA scheme achieves moderate security and speed scores, but low in Intelligence and Robustness. In essence, these results confirm that the proposed approach is a well-balanced, future-ready solution.

As expected, AES is fastest and most memory efficient: it finishes encryption in ~0.64 s using limited memory (from its optimized implementation). The proposed framework has a slightly slower speed (~0.98 s) and consumes about 12–12.5MB of memory, due to the overhead of the ML monitoring and adaptive mechanisms. Chaotic-DNA encryption is the slowest (~1.02 s) and about 10.9 MB, since the complex non-linear operation does not have ML optimization. Crucially, this rather modest trade-off in speed/memory is matched by a large degree of increase in security: the intelligence and adaptability of the proposed system lead to much higher entropy and robustness than the alternatives. In summary, it is evident that the proposed approach provides better security and flexibility with an acceptable overhead and is thus very appropriate for protecting sensitive genomic data in the real world.

VI. FUTURE RESEARCH DIRECTIONS

The proposed framework not only covers existing security issues in genomic data processing but also provides the base for a set of creative research extensions. Important directions for future work can be the following.

A. Deep Learning for Threat Prediction and Adaptive Security

Deep learning models, specifically Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNNs), and Transformer-based models, can be used to noticeably increase the real-time capability of the system with respect to detecting bizarre patterns, intrusions, and emerging threats in real-time. These models can learn complex temporal and spatial dependencies within genomic traffic data and, therefore, help predict cyber threats or unauthorized data access more accurately. Moreover, self-attention mechanisms in transformers can be used to prioritize critical features of the input sequence, thereby leading to better interpretability and adaptive threat responses.

B. Quantum-Resistant DNA Encryption Schemes

With the arrival of quantum computers, standard encryption algorithms are more than ever at risk of quantum attacks. DNA-based cryptography enhanced with Post-Quantum Cryptographic (PQC) primitives, such as lattice, hash-based, or code-based encryption, could offer improved security. Future work should be devoted to the design of hybrid frameworks that combine the parallelism and massive amount of storage capacity of DNA sequences with mathematically rigorous PQC techniques. This dual-layer approach may be a way to future-proof genomic data against both classical and quantum adversaries.

C. Edge Computing Integration for Real-Time Genomic Security

One of the greatest potential benefits of Edge computing for genomic security applications is deploying DNA cryptographic models on edge computing platforms (e.g., embedded systems deployed in genomic sequencers, mobile health devices, portable diagnostic tools) for onsite, real-time encryption and threat detection. Not only is this more environmentally friendly from the perspective of centralized cloud resources, but it also guarantees low-latency processing and greater data sovereignty. Lightweight cryptography algorithms, as well as compressed ML models, will be necessary to optimize performance for the limited computational capabilities of edge devices.

D. Federated Learning (FL) for Privacy-Preserving Multi-Site Genomic Analysis

FL is a method for collaborative model training at different decentralized data sources without having to send raw genomic data to a central server. By keeping sensitive genomic information local and only sharing updates of their models, FL protects privacy and supports strict data protection legislations, e.g., HIPAA or GDPR. When combined with DNA cryptography methods, FL can help support secure studies of multi-institutional research, including robust anomaly detection, as well as model generalization across diverse populations and sequencing environments. In addition, differential privacy and secure aggregation protocols can be integrated, which can further reduce the potential for inference and model poisoning attacks.

VII. CONCLUSIONS

This study presented an adaptive cryptography architecture that merged DNA-based encoding with ML to address security risks in the field of bioinformatics. The system is based on chaotic diffusion by leveraging the concept of exploiting the complexity of DNA sequence permutations. Combined with intelligent threat detection, based on an RF classifier, it provides an acceptable level of confidentiality, integrity, and scalability. The protocol of reshuffling the dynamic keys following the optimization of parameters based on anomalies improves the resistance to statistical, brute force, and replay attacks. The soundness of the framework was experimentally assessed, with Shannon entropy scores indicating great randomness, strong avalanches reflecting high sensitivity to input changes, low cipher text correlation, and low execution cost, showing the feasibility of the framework in the context of real-world applications. High recall and precision were achieved by the built-in anomaly-detection module, enabling proactive defense in dynamic threat environments. The integration of bio-inspired cryptography and ML is an important breakthrough in the administrative protection of genomic and biomedical data. In addition to strengthening data security, the suggested system offers a flexible and scalable architecture that can be optimally utilized in genomics, personalized medicine, and secure biomedical cloud computing applications in the future.

ACKNOWLEDGMENTS

This paper was derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 13382-psu-2023-PSNU-R-3-1-EI-. The authors acknowledge the support of Prince Sultan University, Riyadh, Saudi Arabia, in paying the article processing charges of this publication. This research is supported by the Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia.

REFERENCES

[1] L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, Nov. 1994, https://doi.org/10.1126/science.7973651.

[2] P. Agarwal, N. Pinnamaneni, and T. Heinis, "Motif caller for sequence reconstruction in motif-based DNA storage," *Scientific Reports*, vol. 15, no. 1, Nov. 2025, Art. no. 39236, https://doi.org/10.1038/s41598-025-22798-2.

[3] P. Chithaluru, V. K. Reddy, P. Narsimhulu, and M. Kumar, "DNA computing for the smart wireless sensor networks," in *Blockchain and Digital Twin for Smart Healthcare*, Elsevier, 2025, pp. 395–417.

[4] R. Annan, J. Noland, K. Perkins, X. Yuan, K. Roy, and L. Qingge, "Genomic privacy and security in the era of artificial intelligence and quantum computing," *Discover Computing*, vol. 28, no. 1, June 2025, Art. no. 108, https://doi.org/10.1007/s10791-025-09627-w.

[5] E. Ayday, J. L. Raisaro, and J.-P. Hubaux, "Privacy-Enhancing Technologies for Medical Tests Using Genomic Data," Ecole Polytechnique Federale de Lausanne (EPFL), 2012.

[6] T. V. Galithoti, P. Priyaranjan Nayak, M. R, P. Sudan, P. S, and N. Kaushik, "Enhancing Security In Network Communication Encryption Systems Using Bio-Inspired Algorithm," in *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, Apr. 2025, pp. 1–4, https://doi.org/10.1109/ICMCTC62214.2025.11196683.

[7] J. Blindenbach, J. Kang, S. Hong, C. Karam, T. Lehner, and G. Gürsoy, "SQUiD: ultra-secure storage and analysis of genetic data for the advancement of precision medicine," *Genome Biology*, vol. 25, no. 1, Dec. 2024, Art. no. 314, https://doi.org/10.1186/s13059-024-03447-9.

[8] D. Boneh and C. Dunworth, "Breaking DES using a molecular computer," in *DNA Based Computers - Proceedings of a DIMACS Workshop*, Apr. 1995, vol. 27, pp. 37–66.

[9] A. Kairi and T. Bhadra, "Decoding the Future Using a Novel DNA-based Cryptosystem," *European Chemical Bulletin*, vol. 12, no. si10, pp. 3596–3608, Aug. 2023, https://doi.org/10.48047/ecb/2023.12.si10.00413.

[10] C. T. Clelland, V. Risca, and C. Bancroft, "Hiding messages in DNA microdots," *Nature*, vol. 399, no. 6736, pp. 533–534, June 1999, https://doi.org/10.1038/21092.

[11] A. Gehani, T. LaBean, and J. Reif, "DNA-based Cryptography," in *Aspects of Molecular Computing: Essays Dedicated to Tom Head, on the Occasion of His 70th Birthday*, N. Jonoska, G. Păun, and G. Rozenberg, Eds. Springer, 2004, pp. 167–188.

[12] R. Pulivarti *et al.*, "Cybersecurity of genomic data," National Institute of Standards and Technology (U.S.), NIST IR 8432, Dec. 2023. https://doi.org/10.6028/NIST.IR.8432.

[13] A. Kairi, T. Bhadra, T. Saha, and S. Saha, "Enhancing Healthcare Image Security With DNA Cryptography in the IOT Environment," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 10, pp. 1975–1986, 2024.

[14] R. Patil and G. HimaBindu, "CuLOA-based Data Encryption with Tuned Key for Privacy Preservation in the Cloud," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19546–19552, Feb. 2025, https://doi.org/10.48084/etasr.8523.

[15] A. Saini and R. Sehrawat, "Enhancing Data Security through Machine Learning-based Key Generation and Encryption," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14148–14154, June 2024, https://doi.org/10.48084/etasr.7181.

[16] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, July 1949, https://doi.org/10.1002/j.1538-7305.1949.tb00928.x.

[17] Y. Makwana, A. Panigrahi, and S. K. Pal, "Complete Key Recovery of a DNA-based Encryption and Developing a Novel Stream Cipher for Color Image Encryption: Bio-SNOW." arXiv, Mar. 10, 2025, https://doi.org/10.48550/arXiv.2503.06925.

[18] K. Tanaka, A. Okamoto, and I. Saito, "Public-key system using DNA as a one-way function for key distribution," *Biosystems*, vol. 81, no. 1, pp. 25–29, July 2005, https://doi.org/10.1016/j.biosystems.2005.01.004.

[19] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems Technologies*, vol. 10, no. 2, Jan. 2019, Art. no. 12, https://doi.org/10.1145/3298981.