Journal of the Faculty of Engineering and Architecture of Gazi University 41:2 (2026) 841-853

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University
Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Transformer-based DNA encryption for genomic data security

Alev Kaya[1]*, İbrahim Türkoğlu[2]
[1]Department of Software Engineering, Institute of Science, Firat University, 23119, Elazığ, Türkiye
[2]Department of Software Engineering, Faculty of Technology, Firat University, 23119, Elazığ, Türkiye

Highlights:

- Transformer-based DNA-native pseudorandom keystream generation
- DNA-native encryption pipeline compatible with biophysical channel constraints
- Statistical quality assessment using NIST tests and scalable core-processing performance

Keywords:

- Genomic data security
- DNA-based encryption
- Transformer-based DNA PRNG
- Quaternary DNA-XOR
- NIST randomness tests

Article Info:

Research Article
Received: 13.01.2025
Accepted: 21.02.2026

DOI:

10.17341/gazimmfd.1619337

Acknowledgement:

This study was produced within the scope of the doctoral thesis entitled "New Artificial Intelligence-Based Approaches to Improve the Performance of Genomic Encryption Methods" conducted by Alev KAYA under the supervision of Prof. Dr. İbrahim TÜRKOĞLU at the Department of Software Engineering, Institute of Science, Firat University

Correspondence:

Author: Alev Kaya
e-mail: alev.kaya@firat.edu.tr
phone: +90 538 681 3981

Graphical/Tabular Abstract

Figure A presents the mean p-values obtained from NIST statistical tests over 10 independent DNA keystream sequences. The results provide a compact summary of the pseudorandom behavior of the proposed DNA-native encryption framework. All mean p-values remain above the conventional significance threshold, indicating statistically acceptable randomness behavior in the evaluated test setting.

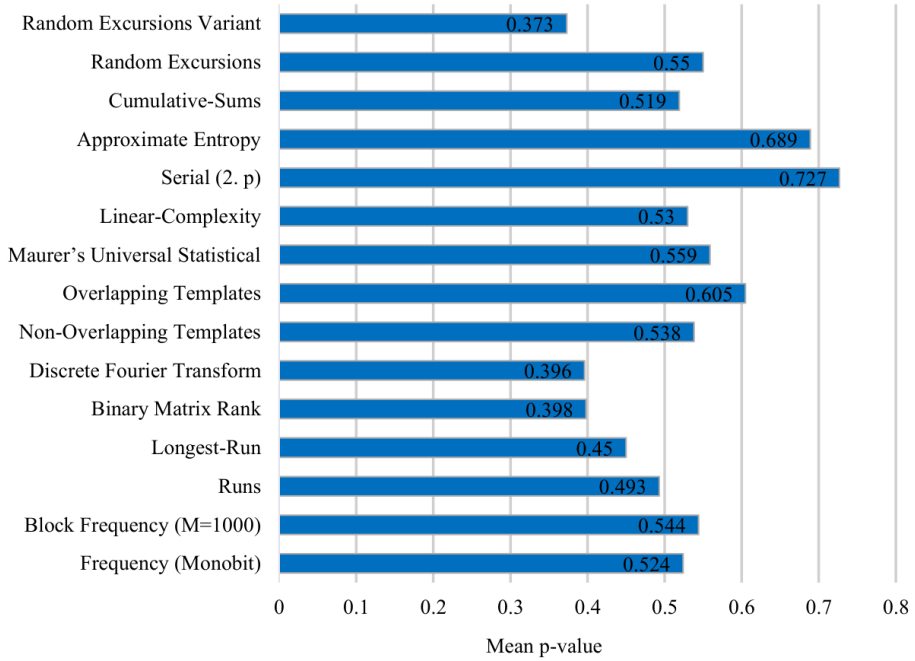

Figure A. Mean p-values obtained from NIST statistical tests over 10 independent sequences

Purpose: This study proposes a DNA-native encryption framework for genomic data security. The method operates directly on the {A, C, G, T} alphabet and avoids a separate bit-to-DNA conversion stage.

Theory and Methods: The method uses a no-reference Transformer-based DNA keystream generator to produce constrained pseudorandom DNA sequences. The generated keystream is synchronized with the plaintext genomic sequence and combined through a quaternary DNA-XOR operation. GC-content balance, homopolymer limitation and short-range correlation control are considered during keystream generation. Decryption applies the inverse operation to recover the original sequence.

Results: Experiments on genomic DNA sequences showed lossless reconstruction after decryption. The generated keystreams preserved DNA-channel constraints, and NIST-based randomness analyses over 10 independent DNA keystream sequences indicated statistically acceptable pseudorandom behavior. The proposed workflow also provided efficient encryption and decryption performance with low additional memory demand.

Conclusion: The results indicate that Transformer-based DNA-native keystream generation can be integrated with quaternary DNA-XOR encryption for genomic data protection. The framework provides a reproducible and constraint-aware basis for DNA-oriented secure data processing.

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University
Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Genomik verilerin güvenliği için Transformer tabanlı DNA şifreleme

Alev Kaya[1]*, İbrahim Türkoğlu[2]
[1]Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Anabilim Dalı, 23119, Elazığ, Türkiye
[2]Fırat Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, 23119, Elazığ, Türkiye

Ö N E Ç I K A N L A R

- Transformer tabanlı DNA-yerel sözde rastgele anahtar akışı üretimi
- Biyofiziksel kanal kısıtlarıyla uyumlu DNA-yerel şifreleme hattı
- NIST testleri ve ölçeklenebilir çekirdek işlem performansı ile istatistiksel kalite değerlendirmesi

Makale Bilgileri
Araştırma Makalesi
Geliş: 13.01.2025
Kabul: 21.02.2026

DOI:
10.17341/gazimmfd.1619337

Anahtar Kelimeler:
Genomik veri güvenliği,
DNA şifreleme,
Transformer tabanlı DNA-PRNG,
NIST testleri

ÖZ

Genomik verilerin hassas doğası ve hızla artan ölçeği, verimli ve biyofiziksel kısıtlarla uyumlu güvenli depolama çözümlerini zorunlu kılmaktadır. Mevcut kriptografik yaklaşımlar, DNA kanalına özgü GC içeriği ve homopolimer sınırlamalarını doğrudan karşılayamadığı için genomik veri iş akışlarında uyum sorunu yaratabilmektedir. Ayrıca şifreli çıktının açık biçimde ayırt edilebilmesi, steganografik açıdan ek bir sınırlılık oluşturmaktadır. Bu çalışma, çıktısı doğrudan {A, C, G, T} bazlarından oluşan Transformer tabanlı DNA-yerel bir sözde rastgele sayı üreteci (PRNG) ile kuaterner XOR çekirdeğini birleştiren uçtan uca bir şifreleme hattı sunmaktadır. PRNG süreci; GC ≈ 0,5 hedefi, homopolimer ≤ 5 sınırı ve lag-1 sönümleme ile optimize edilmiştir. Performans değerlendirmeleri, literatürdeki yöntemlere göre yaklaşık $3{,}6\times10^3$ kat hızlanma, O(N) zaman karmaşıklığı ve O(1) ek bellek kullanımı ile yüksek ölçeklenebilirlik göstermiştir. Rastgelelik değerlendirmelerinde NIST SP 800-22 ve NIST SP 800-90B testleri kullanılmış ve istatistiksel kalite açısından uyumlu sonuçlar elde edilmiştir. Bütünlük testlerinde 10^7 baza kadar kayıpsız geri kazanım sağlanmış ve sonuçlar SHA-256 özet eşitliği ile baz/bit düzeyi karşılaştırmalar üzerinden doğrulanmıştır. Güvenlik analizleri, PRNG düzleminde p ≈ 0,5 çığ etkisi ve 128 bazlık çekirdek anahtar bloğu altında anahtar duyarlılığı sergilemektedir. Sonuçlar, önerilen yaklaşımın DNA tabanlı veri ekosistemlerine doğrudan bütünleştirilmesi için elverişli olduğunu göstermektedir.

Transformer-based DNA encryption for genomic data security

H I G H L I G H T S

- Transformer-based DNA-native pseudorandom keystream generation
- DNA-native encryption pipeline compatible with biophysical channel constraints
- Statistical quality assessment using NIST tests and scalable core-processing performance

Article Info
Research Article
Received: 13.01.2025
Accepted: 21.02.2026

DOI:
10.17341/gazimmfd.1619337

Keywords:
Genomic data security,
DNA encryption,
Transformer-based DNA PRNG,
NIST tests

ABSTRACT

The sensitive nature and rapidly increasing scale of genomic data necessitate secure storage solutions that are efficient and compatible with biophysical constraints. Existing cryptographic approaches often encounter compatibility issues in genomic data workflows because they do not directly address DNA channel-specific GC-content and homopolymer limitations. The explicit distinguishability of encrypted outputs may also introduce an additional steganographic limitation. This study presents an end-to-end encryption pipeline that integrates a Transformer-based DNA-native pseudo-random number generator (PRNG), which directly outputs {A, C, G, T} bases, with a quaternary XOR kernel. The PRNG process is optimized using a GC ≈ 0.5 target, a homopolymer limit ≤ 5, and lag-1 autocorrelation suppression. Performance evaluations demonstrate high scalability, achieving approximately 3.6×10^3 speedup over existing methods with O(N) time complexity and O(1) auxiliary memory. Randomness assessments using NIST SP 800-22 and NIST SP 800-90B indicate statistically consistent results. Integrity tests yielded lossless recovery for sequences up to 10^7 bases, verified using SHA-256 hash equality and base-/bit-level comparisons. Security analyses exhibit an avalanche effect of p ≈ 0.5 at the PRNG level and key sensitivity under a 128-base core key block. These results indicate that the proposed approach is suitable for direct integration into DNA-based data ecosystems.

*Sorumlu Yazar / Yazarlar / *Corresponding Author / Authors: *alev.kaya@firat.edu.tr, iturkoglu@firat.edu.tr / Tel: +90 538 681 3981

1. Giriş (Introduction)

Genomik veriler, bir organizmanın genomuna ilişkin kapsamlı ve kalıcı biyolojik bilgi taşımaktadır. Özellikle insan genomu ve biyomedikal uygulamalar bağlamında bu veriler, klasik veri türlerine kıyasla daha yüksek mahremiyet ve bütünlük gereksinimi doğurmaktadır [1-3]. Kişiselleştirilmiş tıp, klinik karar destek ve biyobanka uygulamalarında genomik diziler uzun vadeli depolama ve paylaşım süreçlerinin konusudur [4-6] ve çoğu kez bulut tabanlı altyapılara bağımlıdır [7-9]. Bu nedenle genomik veri güvenliği yalnızca verinin şifrelenmesiyle sınırlı değildir; hesaplama maliyeti, geri kazanım doğruluğu, ölçeklenebilirlik ve sistem düzeyinde uygulanabilirlik de güvenli tasarımın temel bileşenleri hâline gelmektedir [10, 11]. Genomik bilginin yeniden kimliklendirme, yetkisiz erişim ve kötüye kullanım riskleri dikkate alındığında, bu alan bireysel mahremiyetin ötesinde toplumsal, ekonomik ve biyogüvenlik boyutları bulunan kritik bir araştırma konusu olarak değerlendirilmektedir [12-14].

Bu bağlamda genomik verinin deoksiribonükleik asit (Deoxyribonucleic acid, DNA) alfabesi olan {A, C, G, T} üzerinde temsil edilmesi, şifreleme yaklaşımlarının yalnızca klasik ikili veri düzleminde değil, DNA-özgü kanal özellikleriyle birlikte ele alınmasını gerektirmektedir. DNA tabanlı iş akışlarında Guanin (G) + Sitozin (C) dengesi, homopolimer uzunluğu ve yerel motif düzenlilikleri gibi özellikler hem biyolojik temsil hem de veri işleme uyumluluğu açısından önem taşımaktadır. Buna karşılık yaygın "önce bit sonra DNA" yaklaşımı, ikili şifreli akışı sabit eşleme kurallarıyla DNA alfabesine dönüştürmekte ve GC sürüklenmesi, uzun homopolimer koşuları veya yerel istatistiksel düzenlilikler gibi istenmeyen örüntüler oluşturabilmektedir [15-17]. Bu durum, şifreli çıktının DNA-özgü istatistiksel özelliklerini etkileyebildiği gibi, çıktının doğal DNA dizilerinden ayırt edilebilirliğini artırarak steganografik açıdan ek bir sınırlılık da oluşturabilmektedir [16].

Yapay zekâ (YZ) tabanlı sözde rastgele sayı üreteçleri (Pseudo-Random Number Generator, PRNG), kriptografik anahtar akışı üretimi için farklı bağlamlarda incelenmiştir [18-20]. Hopfield ağları [21, 22], uzun kısa süreli bellek (Long Short-Term Memory, LSTM) [23, 24], çekişmeli üretici ağlar (Generative Adversarial Networks, GAN) [25, 26], evrişimli sinir ağları (Convolutional Neural Networks, CNN) [27-29] ve varyasyonel otokodlayıcılar (Variational Autoencoders, VAE) [30] bu kapsamda kullanılan başlıca yaklaşımlar arasındadır. Bununla birlikte mevcut yöntemlerin önemli bir bölümü ikili uzayda anahtar üretimine odaklanmakta veya haricî veri kümeleri, ekstraktörler ve anahtar türetim fonksiyonları sonrasında doğrulama gerektirmektedir [31, 32]. Doğrudan DNA alfabesinde çalışan, dış eğitim verisine bağımlı olmayan ve üretim sırasında GC dengesi, homopolimer sınırı ve kısa gecikmeli otokorelasyon gibi DNA-özgü kısıtları gözeten PRNG tabanlı şifreleme yaklaşımları ise sınırlı düzeydedir [33, 34].

Bu çalışma, söz konusu boşluğu genomik FASTA dizileri üzerinde çalışan DNA-yerel bir şifreleme hattı ile ele almaktadır. Önerilen yaklaşımda, düz metin DNA dizisi eğitimsiz ve referanssız Transformer tabanlı DNA-PRNG tarafından üretilen anahtar akışı ile kuaterner DNA-XOR çekirdeği üzerinden birleştirilmektedir. Anahtar akışı doğrudan {A, C, G, T} alfabesi üzerinde üretilmekte; üretim sırasında GC $\approx 0{,}50 \pm 0{,}02$ hedefi, homopolimer ≤ 5 sınırı ve lag-1 otokorelasyon sönümlemesi uygulanmaktadır. Her mesaj için benzersiz başlatma vektörü (Initialization Vector, IV) / tek kullanımlık sayı (nonce) kullanımı ve anahtar akışının yeniden kullanılmaması, güvenli kullanım için temel ilke olarak benimsenmiştir. Bu çerçevede çalışmanın temel katkıları aşağıdaki şekilde özetlenebilir:

- Birincisi, dış eğitim verisine ihtiyaç duymadan doğrudan DNA alfabesinde çalışan Transformer tabanlı ve referanssız bir PRNG ile DNA-yerel anahtar akışı üretimi sunulmuştur.
- İkincisi, GC dengesi, homopolimer sınırı ve kısa gecikmeli otokorelasyon sönümlemesini üretim sırasında gözeten kuaterner DNA-XOR tabanlı bir şifreleme/şifre çözme hattı geliştirilmiştir.
- Üçüncüsü, önerilen yapı performans, rastgelelik, bütünlük ve güvenlik odaklı deneylerle değerlendirilmiş; farklı ölçeklerde kayıpsız geri kazanım, rastgelelik sonuçları ve ölçeklenebilir çekirdek işlem performansı raporlanmıştır.

Böylece önerilen yöntem, genomik veri iş akışlarında DNA-uyumlu, tekrarlanabilir ve ölçeklenebilir bir şifreleme yaklaşımı olarak konumlandırılmıştır.

1.1. İlgili Çalışmalar (Related Work)

Kriptografik şemaların temel hedefleri gizlilik, bütünlük ve kimlik doğrulamadır. Mükemmel gizlilik kuramında tek kullanımlık anahtar yaklaşımı, anahtarın tam rastgele, bağımsız ve mesajla en az eş uzunlukta olmasını gerektirir [35, 36]. Pratik akış şifrelerinde ise anahtar akışı genellikle sonlu uzunlukta gizli bir anahtar ve her mesaj için benzersiz IV/nonce bilgisinden türetilir. Aynı anahtar akışının yeniden kullanılmaması temel bir güvenlik ilkesi olarak kabul edilir [35]. Bu çalışma, ilgili literatürü üç eksende konumlandırmaktadır: genomik verilerin kriptografik korunması, DNA tabanlı ve YZ destekli kriptografi yaklaşımları ve YZ tabanlı PRNG üretimi. Bu ayrım, önerilen yöntemin genomik veri güvenliği, DNA-yerel şifreleme ve referanssız anahtar akışı üretimi bağlamındaki konumunu açık biçimde göstermektedir.

1.1.1. Kriptografik tekniklerle genomik verilerin korunması (Protecting genomic data with cryptographic techniques)

Bulut tabanlı ortamlarda genomik verilerin güvenli saklanması, paylaşılması ve sorgulanması; gizlilik koruyucu protokoller, homomorfik şifreleme (Homomorphic Encryption, HE) ve güvenli sorgu modelleri üzerinden ele alınmaktadır [1-4]. Bu kapsamda HE, veriyi çözmeden belirli işlemlerin yapılabilmesini sağladığı için genomik veri güvenliği çalışmalarında önemli bir yer tutmaktadır [2]. Kantarcioğlu vd. [6], genomik dizilerin güvenli paylaşımı ve sorgulanması için kriptografik bir çerçeve önermiştir. Canim vd. [7] ise donanım destekli güvenli biyomedikal veri yönetimiyle performans-güvenlik dengesini farklı bir açıdan ele almıştır. Daha güncel çalışmalarda Schneider ve Tkachenko [37], dış kaynaklı genomik veri tabanlarında benzerlik sorgularını gizlilik koruyucu biçimde gerçekleştiren verimli bir protokol sunmuştur. Ghasemi vd. [38] dış kaynaklı genomik veri tabanlarında sorgu işleme verimini artırmaya odaklanmıştır. Nassar vd. [12] DNA veri tabanlarında toplulaştırılmış sorguların korunmasını incelemiştir. Ayrıca büyük genomik veri dosyalarının şifrelenmesi için Shangyong Mima (SM) algoritma serisini temel alan yaklaşımlar da raporlanmıştır [5]. Çok kaynaklı veya dinamik DNA veri kümeleri üzerinde doğrulanabilir gizlilik koruyucu sorgulara yönelik çalışmalar da literatürde yer almaktadır [8].

Bununla birlikte HE ve benzeri kriptografik yaklaşımların önemli bir bölümü ikili veri temsili ve genel amaçlı işlem katmanları üzerinde konumlanmaktadır. Tam HE geniş işlem sınıfını desteklese de hesaplama maliyeti ve gecikme nedeniyle pratikte sınırlıdır. Kısmi HE ise yalnız belirli işlem türlerini desteklediğinden ifade gücü daha dardır [2]. Ayrıca bu yöntemler çoğu senaryoda GC dengesi, homopolimer sınırı ve DNA-özgü istatistiksel uyumluluk gibi kanal kısıtlarını şifreli çıktı üretimi veya DNA-uyumlu temsil sırasında doğrudan hedeflememektedir [16]. Bu nedenle genomik veri

güvenliği bağlamında, DNA alfabesinde çalışan ve kanal kısıtlarını şifreleme hattının doğrudan parçası olarak gözeten yaklaşımlara ihtiyaç devam etmektedir.

1.1.2. DNA tabanlı ve YZ destekli şifreleme yaklaşımları (DNA-based and AI-supported encryption approaches)

DNA kriptografisi, verinin DNA alfabesi üzerinden temsil edilmesi, kodlanması veya şifreleme sürecine dâhil edilmesi yoluyla güvenliği artırmayı amaçlayan bir araştırma alanıdır [28, 31, 33]. Kuaterner alfabenin alternatif eşleme, kompakt temsil ve doğal paralellik gibi özellikleri, DNA tabanlı kriptografik tasarımlarda farklı biçimlerde kullanılmaktadır [33, 39, 40]. Bu kapsamda DNA kodlama, DNA-XOR, biyolojik hizalama, kaotik sistemler ve YZ destekli anahtar üretimi farklı bileşimlerle ele alınmıştır. Kalsi vd. [18], DNA tabanlı XOR ve dinamik anahtar üretimini Genetik Algoritma (Genetic Algorithm, GA) ve Needleman–Wunsch hizalamasıyla birlikte kullanmıştır. Basu vd. [19], Huffman-protein eşlemesi ve çift yönlü ilişkisel bellek ağı (Bidirectional Associative Memory Neural Network, BAMNN) ile hızlı bir şifreleme yaklaşımı önermiştir. Görüntü şifreleme odaklı çalışmalarda Patel vd. [15], hibrit kaos haritaları ve yapay sinir ağı (Artificial Neural Network, ANN) tabanlı PRNG kullanmıştır. Maddodi vd. [20] ise DNA ve kaotik sistemleri birleştiren bir PRNG yapısı raporlamıştır. Teng vd. [31], ENIGMA adı verdiği yöntemle ile çok formatlı verileri DNA dizileri üzerinden gizlemiş ve BAMNN ile ölçeklenebilirliği değerlendirmiştir. Taj vd. [14], DNA ve GA hibriti ile Diffie–Hellman anahtar değişimini ele almış; Alloun vd. [32] Lorenz kaosu tabanlı ANN-PRNG yapısını şifreleme bağlamında incelemiştir. Sakr vd. [33] ise amino asit tabanlı model, GA ile başlangıç anahtarı ve DNA-protein dönüşümüyle ikinci anahtar üretimini içeren hibrit bir anahtar üretim hattı önermiştir.

Bununla birlikte mevcut çalışmaların önemli bir bölümü "önce bit sonra DNA" dönüşümüne veya DNA temsillerinin şifreleme hattına sonradan eklenmesine dayanmaktadır. Bu nedenle GC dengesi, homopolimer sınırı ve yerel motif düzenlilikleri gibi DNA kanal kısıtları çoğu kez üretim sırasında değil, çıktı sonrası uyarlama veya değerlendirme adımlarında ele alınmaktadır [16]. Ayrıca bazı çalışmalar haricî veri kümelerine, ek doğrulama adımlarına veya karmaşık hibrit bileşenlere dayandığından, yeniden üretilebilirlik ve protokol yalınlığı açısından sınırlılıklar ortaya çıkabilmektedir. IV/nonce yönetimi, anahtar akışı tekrar kullanımının engellenmesi ve DNA↔bit kural çizelgesi gibi uygulama düzeyi ayrıntıların her çalışmada aynı açıklıkta raporlanmaması da karşılaştırılabilirliği zorlaştırmaktadır [31-34]. Bu sınırlılıklar, DNA tabanlı şifreleme hattında anahtar akışının doğrudan DNA alfabesinde üretilmesi ve kanal kısıtlarının üretim sürecine çevrim içi olarak dâhil edilmesi ihtiyacını ortaya koymaktadır.

1.1.3. YZ tabanlı PRNG üretimi (AI-based PRNG generation)

YZ tabanlı PRNG çalışmaları, öğrenilebilir rastgelelik üretimi, cihaz-üstü verimlilik ve istatistiksel test bataryalarında yüksek geçiş oranları nedeniyle anahtar akışı üretimi açısından önemli bir araştırma çizgisi oluşturmuştur. Bu kapsamda Jeong vd. [23], irrasyonel sayı dizilerini taklit eden LSTM-RNN tabanlı bir PRNG ile yüksek doğruluk bildirmiştir. Pasqualini vd. [24], pekiştirmeli öğrenme çerçevesinde Markov karar süreci (Markov Decision Process, MDP) ile zaman bağımlılıklarını öğrenen bir PRNG önermiş; ancak ölçeklenebilirlik sınırlılıklarını vurgulamıştır. Kim vd. [25], GAN tabanlı PRNG'yi gömülü cihazlara uyarlayarak TensorFlow Lite üzerinde uygulanabilirliğini göstermiştir. De Bernardi vd. [26], kaotik olmayan bir entropi kaynağını taklit eden GAN yaklaşımıyla Ulusal Standartlar ve Teknoloji Enstitüsü (National Institute of Standards and Technology, NIST) testlerinde yüksek başarı raporlamıştır. Okada vd. [29] ise Wasserstein GAN (WGAN) tabanlı PRNG ile kapalı formül tasarımlarına ihtiyaç duymadan NIST bataryasında başarılı sonuçlar elde etmiştir. Mukherjee vd. [34] GA temelli bir yöntemle DNA anahtar israfını azaltmayı hedeflemiştir. Bu çalışmalar, LSTM-RNN, pekiştirmeli öğrenme, GAN, WGAN ve GA tabanlı yaklaşımların PRNG üretiminde uygulanabilirliğini göstermektedir.

Tablo 1. Önerilen yöntemin mevcut tekniklerle kıyaslanması (Comparison of the proposed method with existing techniques)

Kaynak	Anahtar üretim yöntemi	Çıktı yapısı	Veri türü	Güçlü yönler	Temel sınırlılıklar
[5]	ECC + SM	İkili / 128-bit statik	Genomik veri	Standart kriptografik protokollerle uyumlu; genomik dosya şifrelemeye uygulanabilir	DNA-yerel değildir; bit→DNA dönüşümü gerekir; kanal kısıtları üretim sırasında gözetilmez
[8]	HE + kaos haritası	İkili / statik	Genomik veri	Düşük donanım kaynağıyla uygulanabilir; gizlilik koruyucu işlem sağlar	Büyük veride performans düşebilir; GC ve homopolimer kısıtları doğrudan hedeflenmez
[12]	Paillier HE	İkili / 1024-bit statik	Genomik veri	Veriyi çözmeden işlem yapmayı destekler	Hesaplama maliyeti ve veri genişlemesi yüksektir; DNA-yerel entegrasyon yoktur
[37]	Hibrit HE	İkili / 128-bit statik	Genomik veri	Gizlilik koruyucu benzerlik sorguları için ölçek odaklıdır	Kapsam sorgu protokolüyle sınırlıdır; DNA kanal kısıtları üretim sürecine dâhil değildir
[38]	Paillier tabanlı proxy	İkili / statik	Genomik veri	Dış kaynaklı sunucuda güvenli sorgu/işleme sağlar	İşlem süresi yüksektir; çıktı ikili formattadır; DNA-yerel entegrasyon sunmaz
[14]	GA + Mealy machine	İkili / 48-bit dinamik	Metin verisi	Küçük ölçekli senaryolarda hızlı anahtar üretimi sağlar	Anahtar boyutu sınırlıdır; genomik veri ve büyük ölçekli değerlendirme yoktur
[18]	GA + Needleman–Wunsch	İkili / 64-bit dinamik	Metin verisi	DNA-XOR ve biyolojik hizalama tabanlı hibrit tasarım sunar	Anahtar uzayı sınırlıdır; büyük veri ve kanal kısıt değerlendirmesi sınırlıdır
[19]	YSA (BAMNN)	İkili / 16-bit dinamik	Metin verisi	YSA tabanlı anahtar üretimiyle hızlı şifreleme sağlar	Anahtar boyutu düşüktür; IV/nonce ve tekrar kullanım politikası açık değildir
[33]	GA + biyolojik dönüşüm	İkili / statik	Metin verisi	Biyolojik dönüşüm ve GA tabanlı hibrit anahtar üretimi sunar	Anahtar üretim hattı karmaşıktır; IV/nonce politikası ve tekrar kullanım yasağı açık değildir
Bu çalışma	Referanssız Transformer tabanlı DNA-PRNG + DNA-XOR	DNA / 128 bazlık çekirdek anahtar bloğu	Genomik FASTA verisi	DNA-yerel anahtar akışı üretir; GC, homopolimer ve lag-1 kısıtlarını çevrim içi gözetir; kayıpsız geri kazanım sağlar	Başlatma maliyeti ve Transformer mimarisine bağlı işlem yükü vardır; güvenli kullanım için benzersiz IV/nonce ve anahtar akışı tekrar yasağı gerekir

Bununla birlikte mevcut YZ tabanlı PRNG literatürünün ağırlıklı kısmı anahtar akışını ikili uzayda üretmekte veya DNA tabanlı şifreleme hattına doğrudan bağlanmamaktadır. DNA temelli anahtar dizileri üreten sınırlı sayıdaki çalışmada ise GC dengesi, homopolimer sınırı ve kısa gecikmeli otokorelasyon gibi kanal kısıtlarının üretim sırasında sistematik olarak gözetilmesi çoğu kez sınırlı kalmaktadır [16, 34]. Ayrıca IV/nonce yönetimi, anahtar akışının tekrar kullanımının engellenmesi, kural çizelgesi ve deneylerin yeniden üretilebilirliği gibi uygulama düzeyi ayrıntılar her çalışmada aynı açıklıkta raporlanmamaktadır. Bu durum, mevcut yöntemlerin karşılaştırılabilirliğini sınırlamaktadır. Transformer mimarisi uzun menzilli bağımlılıkları temsil etme kapasitesine sahip olsa da, bu kapasitenin referanssız DNA-yerel anahtar akışı üretimine sistematik biçimde taşınması hâlen açık bir araştırma alanıdır. Aynı şekilde, bu üretimin DNA-XOR tabanlı uçtan uca bir şifreleme hattıyla bütünleştirilmesi de literatürde sınırlı düzeyde ele alınmıştır. Tablo 1, bu boşlukları ve önerilen yöntemin mevcut yaklaşımlara göre konumunu özetlemektedir.

Tablo 1'de görüldüğü üzere mevcut çalışmaların önemli bir bölümü ikili veri temsiline, sorgu güvenliğine, görüntü/metin şifrelemeye veya DNA dönüşümünü sonradan kullanan hibrit yapılara odaklanmaktadır. Önerilen çalışma ise genomik FASTA verisi üzerinde doğrudan DNA alfabesinde üretilen referanssız Transformer tabanlı anahtar akışını kuaterner DNA-XOR çekirdeği ile birleştirmekte ve GC dengesi, homopolimer sınırı ve kısa gecikmeli otokorelasyon gibi DNA-özgü kısıtları üretim sırasında gözetmektedir. Çalışmanın geri kalanında Bölüm 2'de yöntem ve ölçütler, Bölüm 3'te deneysel bulgular, Bölüm 4'te ise sonuçlar ve gelecek çalışmalar sunulmaktadır.

2. Deneysel Metot (Experimental Method)

Bu çalışmada, eğitimsiz (untrained) Transformer tabanlı DNA-PRNG ile anahtar akışı üretimi ve DNA tabanlı şifreleme/şifre çözme süreci beş aşamada kurgulanmıştır. Süreç, veri seti hazırlığıyla başlamaktadır. Ulusal Biyoteknoloji Bilgi Merkezi'nden (National Center for Biotechnology Information, NCBI) temin edilen FASTA DNA dizileri [41] normalize edilerek yalnızca {A, C, G, T} bazları tutulmuştur. İkinci adım, anahtar akışı üretimidir. Şifreleme/şifre çözme uygulamasında, eğitimsiz ve referanssız Transformer DNA-PRNG ile üretilen 128 bazlık çekirdek anahtar bloğu kullanılmış; veri uzunluğunu karşılamak üzere anahtar akışı periyodik hizalama ile genişletilmiştir. Üçüncü adım şifreleme sürecidir. Düz metin DNA dizisi ile hizalanmış anahtar akışı, kuaterner uzayda mod-4 DNA-XOR işlemiyle birleştirilerek şifreli metin elde edilmiştir. Dördüncü adım şifre çözme sürecidir. Bu aşamada, şifreleme ile aynı koşullarda elde edilen anahtar akışı kullanılmış; şifreli metne mod-4 ters işlem uygulanarak özgün DNA dizisi kayıpsız biçimde geri kazanılmıştır. Beşinci ve son adım olan değerlendirme sürecinde sistemin başarımı; performans, istatistiksel rastgelelik, veri bütünlüğü ve güvenlik/savunma analizleri ile ölçülmüştür.

2.1. Veri Seti Hazırlığı (Dataset Preparation)

Genomik verilerin yetkisiz erişim, yeniden kimliklendirme (re-identification) ve kötüye kullanım risklerine karşı korunması çok katmanlı bir güvenlik yaklaşımı gerektirmektedir [42]. Bu bağlamda yapılan çalışmada, NCBI veritabanından temin edilen [41] ve FASTA formatında sunulan kamuya açık bir DNA referans dizisi, yaklaşık 11,79 MB dosya boyutuyla düz metin kaynağı olarak kullanılmıştır. Önerilen şifreleme şemasının kuaterner alfabe yapısıyla tam uyumlu olması için ham veri deterministik bir ön işleme sürecinden geçirilmiştir. Bu kapsamda FASTA başlık satırları (">...") ayıklanmış, karakterler büyük harfe dönüştürülmüş, satır sonları standartlaştırılmış ve boşluklar temizlenmiştir. Ayrıca Uluslararası Temel ve Uygulamalı Kimya Birliği (International Union of Pure and Applied Chemistry, IUPAC) belirsiz baz kodları (N, R, Y, W, S, K, M, B, D, H, V) dışarıda bırakılmış ve yalnızca kesin nükleotid çağrıları olan {A, C, G, T} bazları korunmuştur. Normalizasyon sonrası olası giriş/çıkış farklılıklarının bütünlük ölçümlerini etkilemesini önlemek için toplam baz sayısı, GC oranı ve SHA-256 özet değeri kayıt altına alınmıştır. Ölçeklenebilirlik analizi amacıyla ana veri setinden deterministik seçim/kırpma ile 1×10^3, 1×10^4, 1×10^5, 1×10^6 ve 1×10^7 nükleotid (nt) uzunluklarında alt veri kümeleri türetilmiştir. Böylece kriptografik akışın farklı dizi uzunluklarındaki tutarlılığı incelenmiştir.

2.2. Transformer ile Anahtar Akışı Üretimi (Keystream Generation with Transformer)

Kerckhoffs ilkesi uyarınca bir kriptosistemin güvenliği algoritmanın gizliliğine değil, anahtarın gizliliğine dayanmalıdır [35]. Bu nedenle anahtar akışı üretiminde her mesaj için benzersiz IV/nonce kullanımı, anahtar akışının yeniden kullanılmaması ve çıktı dizisinin istatistiksel olarak yeterli kalite göstermesi temel gereklilikler arasındadır [26-28]. Bu çalışmada, dış eğitim verisine ihtiyaç duymayan eğitimsiz ve referanssız bir üretim senaryosu benimsenmiş; yalnızca {A, C, G, T} alfabesi üzerinde çalışan Transformer tabanlı bir DNA-PRNG ile DNA anahtar akışı üretilmiştir [43-46]. Üretim süreci, nedensel maskeleme altında otoregresif örnekleme ile yürütülmüş; çıktı karakteristikleri öğrenme tabanlı ağırlık güncellemesi yerine çevrim içi örnekleme kısıtlarıyla kontrol edilmiştir [47].

Önerilen üretici yapı, gömme katmanı ve decoder bloklarından oluşan hafif bir Transformer mimarisine dayanmaktadır [44]. Her decoder bloğu; çok başlıklı öz-dikkat (Multi-Head Self-Attention, MHSA), ileri beslemeli ağ (Feed-Forward Network, FFN) ve katman normu (Layer Norm, LN) bileşenlerini içermektedir. Bu mimari, önceki bazlara koşullu olarak yeni baz olasılıklarını üretmekte ve her adımda {A, C, G, T} alfabesi üzerinde bir sonraki bazın seçilmesini sağlamaktadır. Bu çalışmada bağlam penceresi W = 128 olarak belirlenmiştir. Referanssız DNA-PRNG kod yolu farklı uzunluklarda DNA çıktısı üretebilecek şekilde tasarlanmıştır; PRNG değerlendirme koşusunda 500.000 bazlık çıktı üretimi desteklenmiştir. Ancak bu çalışmadaki şifreleme deneylerinde çekirdek anahtar bloğu $L_k = 128$ baz olarak alınmış ve veri uzunluğunu karşılamak üzere anahtar akışı periyodik hizalama ile genişletilmiştir. Hafif mimari yapılandırmada decoder katman sayısı 3, gömme boyutu 128, dikkat başlığı sayısı 4 ve ileri beslemeli ağ genişliği 512 olarak ayarlanmıştır. Aşırı uyum veya dış veri bağımlılığı oluşturabilecek bir eğitim süreci kullanılmamış; model parametreleri deterministik bir başlatma tohumu altında oluşturulmuştur. Konumsal bilgi, döndürmeli konumsal kodlama (Rotary Positional Encoding, RoPE) ile sağlanmış ve uzun üretimlerde hesaplama verimliliğini korumak amacıyla önbellek tabanlı çıkarım kullanılmıştır.

Üretim sırasında DNA-uyumluluğunu korumak ve kısa menzilli bağımlılıkları sınırlandırmak için çevrim içi kısıtlar uygulanmıştır. Bu kapsamda GC oranı hedef değere yakın tutulmuş (GC $\approx$ 0,50), homopolimer koşuları $\leq$ 5 ile sınırlandırılmış ve ardışık baz tekrarlarını azaltmak amacıyla lag-1 olasılığı zayıflatılmıştır. Bu kısıtlar, çıktı sonrasında baz dağılımı, k-mer dağılımı, homopolimer uzunluğu ve otokorelasyon ölçütleriyle değerlendirilmiştir. Böylece Transformer mimarisi anahtar akışının otoregresif üretimini sağlarken, çevrim içi kısıtlar DNA-uyumluluk ve istatistiksel davranışı yönlendirmiştir. Anahtar akışı üretim süreci aşağıdaki adımlarla yürütülmüştür:

1. Tohumlama: Her üretim koşusu için tekil bir koşu tohumu belirlenmiş ve deterministik örnekleme süreci başlatılmıştır.
2. Başlatma dizisi: Uzunluğu W olan başlangıç dizisi {A, C, G, T} alfabesi üzerinde oluşturulmuş ve modele giriş olarak verilmiştir.

3. İleri geçiş: Başlangıç dizisi üzerinden bir sonraki baza ilişkin olasılık dağılımı hesaplanmıştır.
4. GC yönlendirme: Belirli aralıklarla baz olasılıkları hedef GC oranına yaklaştırılmıştır.
5. Kısıt maskeleme: Homopolimer sınırı, yumuşak dengeleme ve lag-1 sönümleme uygulanarak olasılık dağılımı yeniden düzenlenmiştir.
6. Örnekleme: Güncellenen ve normalize edilen dağılımdan tohum kontrollü deterministik örnekleme ile yeni baz seçilmiş ve anahtar akışına eklenmiştir.
7. Kural seçimi ve ikili çıktı kaydı: Seçilen baz için sekiz DNA↔bit eşleme kuralından biri belirlenmiş; ikili değerlendirme ve tekrarlanabilirlik amacıyla 2 bitlik karşılık ve kural bilgisi kaydedilmiştir.
8. Önbellek güncelleme: Yalnızca son seçilen baz üzerinden yeni ileri geçiş yapılmış; bağlam penceresi aşıldığında KV önbelleği budanmış ve konumsal süreklilik korunmuştur.

Bu yapı sayesinde anahtar akışı, haricî eğitim verisine veya çıktı sonrası ayrı bir anahtar türetim fonksiyonuna ihtiyaç duyulmadan doğrudan DNA alfabesinde üretilmektedir. Üretim sırasında uygulanan GC dengesi, homopolimer sınırı ve lag-1 sönümleme, anahtar akışının DNA tabanlı şifreleme hattıyla uyumlu ve istatistiksel olarak dengeli bir yapı sergilemesine katkıda bulunmaktadır.

2.3. DNA Şifreleme (DNA Encryption)

Bu çalışmada şifreleme, literatürde yaygın olan "ikili → DNA → kripto" hattından ayrılarak doğrudan DNA dizileri üzerinde gerçekleştirilmektedir. Düz metin DNA dizisi $P \in \{A,C,G,T\}^N$ Transformer tabanlı PRNG tarafından üretilen çekirdek anahtar dizisi ise $K_0 \in \{A,C,G,T\}^{L_k}$ olarak tanımlanmıştır. Bu çalışmada $L_k = 128$ baz olup $4^{128} = 2^{256}$ olası çekirdek anahtar bloğuna karşılık gelmektedir. Çekirdek anahtar dizisi, veri uzunluğunu karşılamak üzere periyodik hizalama ile N uzunluğuna genişletilmiş ve hizalanmış anahtar akışı K elde edilmiştir. Böylece ikili dönüşüme dayalı ek ara adımlar kullanılmadan DNA→DNA alanında işlemsel sadelik korunmakta ve moleküler/veri-işleme iş akışlarıyla doğrudan uyum sağlanmaktadır [36, 48]. Kuaterner XOR işlemi, DNA bazlarının ikili uzaya birebir eşlendiği φ dönüşümü üzerinden tanımlanmıştır. Bu çalışmada bazlar {A,C,G,T} kümesinden seçilmiş ve ikili temsil $\varphi(A) = 00,\ \varphi(C) = 01,\ \varphi(G) = 10,\ \varphi(T) = 11$ olarak alınmıştır. Kuaterner XOR işlemi Eş. 1'de verilmiştir.

$$x \oplus_4 y = \varphi^{-1}(\varphi(x) \oplus \varphi(y)) \quad (1)$$

Burada $x, y \in \{A,C,G,T\}$ olup işlem değişmeli ve öz-ters özellik göstermektedir. Bu nedenle aynı işlem şifreleme ve şifre çözme için kullanılabilmektedir. Şifreleme ve şifre çözme işlemleri Eş. 2'de verilmiştir.

$$C = P \oplus_4 K, \quad P = C \oplus_4 K \quad (2)$$

Uygulamada $\oplus_4$ işlemi Tablo 2'de verilen doğruluk tablosuna göre gerçekleştirilmiştir. Kullanılan kuaterner XOR işlemi öz-ters yapıda olduğundan, şifreleme ve şifre çözme aynı tablo üzerinden yürütülmektedir [14, 15, 36, 48].

Tablo 2. DNA dizisi için XOR ekleme ve çıkarma işlemleri
(XOR addition and subtraction operations for DNA sequence)

Ekleme (XOR) ve çıkarma (E-XOR)	A	G	C	T
A	A	G	C	T
G	G	A	T	C
C	C	T	A	G
T	T	C	G	A

2.4. DNA Şifre Çözme (DNA Decryption)

Şifre çözme aşamasında amaç, şifreli DNA dizisi $C \in \{A,C,G,T\}^N$'den özgün düz metin dizisi $P \in \{A,C,G,T\}^N$'yi kayıpsız olarak geri elde etmektir. Kuaterner XOR işleminin öz-ters yapısı nedeniyle şifre çözme, şifreleme ile aynı işlem çekirdeği kullanılarak gerçekleştirilmiştir. Bu aşamada, şifreleme ile aynı koşullarda elde edilen anahtar akışı K, şifreli DNA dizisiyle Eş. 2'de verilen işlem üzerinden birleştirilmiştir. Ardından bileşen bazında $P = C \oplus_4 K$ işlemi uygulanmış ve özgün DNA dizisi geri kazanılmıştır. Kullanılan işlem öz-ters olduğu için ayrı bir çıkarma tablosu tanımlanmamış; Tablo 2'de verilen XOR ekleme/çıkarma doğruluk tablosu hem şifreleme hem de şifre çözme için kullanılmıştır. Bu yapı, FASTA normalizasyonu sonrasında şifre çözme sürecinin doğrusal zamanda yürütülmesini ve yalnızca sabit ek bellek gerektirmesini sağlamaktadır. Bütünlük doğrulaması için geri kazanılan DNA dizisi ile özgün düz metin dizisi karşılaştırılmış; gerekli durumlarda SHA-256 özet eşitliği kontrol edilmiştir.

2.5. Testler (Tests)

Bu çalışmanın başarımı iki ana değerlendirme hattı üzerinden incelenmiştir. Birinci hat, eğitimsiz Transformer tabanlı DNA-PRNG tarafından üretilen anahtar akışının istatistiksel kalitesine ve bağımsızlık özelliklerine odaklanmaktadır. Bu kapsamda PRNG çıktıları; NIST SP 800-22 testleri, NIST SP 800-90B kapsamında min-entropi değerlendirmeleri, bağımsız ve özdeş dağılım (independent and identically distributed, IID) varsayımına yönelik kontroller, çevrim içi sağlık testleri ve DNA-özgü ölçütler ile değerlendirilmiştir. DNA-özgü ölçütler arasında GC dengesi, k-mer dağılımı, homopolimer uzunluğu ve kısa gecikmeli otokorelasyon yer almaktadır.

İkinci hat, DNA kanalında yürütülen şifreleme ve şifre çözme sürecinin doğruluğunu, verimliliğini ve güvenlik senaryoları karşısındaki davranışını değerlendirmektedir. Bu kapsamda kuaterner XOR ($\oplus_4$) çekirdeğiyle üretilen şifreli DNA akışları; bütünlük doğrulaması, şifreli-DNA istatistikleri, performans ölçümleri ve saldırı/savunma senaryoları açısından incelenmiştir. Bütünlük değerlendirmesinde SHA-256 özet eşitliği ile baz ve bit düzeyinde eşleşme birlikte kullanılmıştır. Güvenlik analizlerinde çığ etkisi, anahtar duyarlılığı, seçilmiş düz metin saldırısı (Chosen-Plaintext Attack, CPA), bilinen düz metin saldırısı (Known-Plaintext Attack, KPA) ve hata yayılımı senaryoları ele alınmıştır. Performans ölçümleri yalnızca çekirdek işlemleri, yani anahtar hizalama ve $\oplus_4$ tabanlı şifreleme/şifre çözme adımlarını kapsayacak şekilde yapılmış; disk I/O, serileştirme ve kayıt tutma/loglama süreleri bu ölçümlere dâhil edilmemiştir [36, 48-50].

2.5.1. Performans analiz testleri (Performance analysis tests)

Bu bölümde, 128 bazlık çekirdek anahtar bloğu kullanılarak farklı dizi uzunluklarında şifreleme ve şifre çözme çekirdeğinin hız, kaynak tüketimi ve verimlilik davranışı değerlendirilmiştir [48, 51-53]. Performans ölçümü yalnızca kriptografik çekirdek işlemlerini kapsamaktadır. Bu nedenle anahtar akışının veri uzunluğuna hizalanması ve kuaterner XOR ($\oplus_4$) işlemi ölçüme dâhil edilmiş; disk I/O, serileştirme, dosya yazma/okuma ve loglama süreleri dışarıda bırakılmıştır. Transformer tabanlı PRNG ile 128 bazlık çekirdek anahtar bloğunun üretim süresi yaklaşık 0,326 s sürmüş, ancak çekirdek şifreleme/şifre çözme performansına dâhil edilmemiştir. Büyük dizi uzunluklarında bu başlatma maliyetinin mesaj başına etkisi düşük düzeye inmektedir. Tüm deneyler FASTA-normalize DNA dizileri üzerinde, sabit koşullar altında yürütülmüştür. Şifreleme ve şifre çözme işlemlerinde aynı IV/nonce yapısı, aynı çekirdek anahtar bloğu ve tek iş parçacıklı yürütme protokolü kullanılmıştır.

Kuaterner XOR işleminin öz-ters yapısı nedeniyle şifreleme ve şifre çözme süreçleri hesaplama bakımından izomorfiktir. Bu nedenle iki yöndeki çalışma sürelerinin pratikte birbirine yakın olması beklenmektedir. Çekirdek işlem hattının zaman karmaşıklığı $O(N)$ olarak değerlendirilmiştir. Burada N, FASTA-normalize DNA dizisinde işlenen baz sayısını göstermektedir. Ölçümler 1×10^3, 1×10^4, 1×10^5, 1×10^6 ve 1×10^7 baz uzunluklarında yürütülmüş; bu uzunluklar ASCII tabanlı FASTA temsili için yaklaşık 1 KB–10 MB giriş aralığına karşılık gelmektedir. Raporlanan metrikler; duvar saati çalışma zamanı Δt (ms), süreç-yerel CPU zamanı (s), uygulama düzeyi ek RAM kullanımı (MB) ve verimlilik $\eta = N/\Delta t$ (baz/ms) olarak tanımlanmıştır. RAM ölçümleri, Python uygulamasında bellekte tutulan çıktı tamponları ve geçici çalışma yapıları da dâhil edilerek uygulama düzeyi bellek kullanımı olarak değerlendirilmiştir. Elde edilen sonuçlar Tablo 3'te verilmiştir.

Tablo 3'te verilen bulgular, dizi uzunluğu arttıkça çekirdek çalışma zamanının yaklaşık doğrusal biçimde ölçeklendiğini göstermektedir. 1×10^3–1×10^7 aralığında şifreleme ve şifre çözme süreleri N ile birlikte artmış, iki yön arasındaki fark genel olarak düşük düzeyde kalmıştır. Küçük veri boyutlarında başlatma ve ölçüm sabit giderleri baskın olduğundan verimlilik daha değişken görünmektedir. 1×10^5 düzeyinde verimlilik en yüksek değere ulaşmış, 1×10^6 - 1×10^7 aralığında ise yaklaşık 7,9 –8,2 × 10^3 baz/ms bandında kararlı bir plato sergilemiştir. Ek RAM kullanımı, çalışma dizileri ve geçici kopyalar nedeniyle giriş boyutuyla birlikte artmış; 1×10^7 girişte yaklaşık 20 MB ek bellek kullanımı gözlenmiştir. Bu değer, uygulama düzeyinde ölçülen süreç belleğini göstermekte olup kuramsal çekirdek işlem belleğinden ayrı değerlendirilmelidir. Süreç-yerel CPU zamanı, duvar saati ölçümleriyle uyumlu seyretmiştir.

2.5.2. Rastgelelik analiz testleri (Randomness analysis tests)

Rastgelelik, hem anahtar akışının istatistiksel kalitesi hem de şifreli çıktının güvenlik davranışı açısından temel bir ölçüttür. İstatistiksel sapmalar, anahtar akışının tahmin edilebilirliğini artırarak saldırı yüzeyini genişletebilir [35, 40]. Bu çalışmada rastgelelik değerlendirmesi iki hatta yürütülmüştür. İlk hatta, Transformer tabanlı DNA-PRNG tarafından üretilen anahtar akışları NIST SP 800-22, NIST SP 800-90B ve DNA-özgü metriklerle değerlendirilmiştir [54 - 56]. İkinci hatta ise şifreli DNA çıktıları GC dengesi, k-mer dağılımı, homopolimer uzunluğu ve kısa gecikmeli otokorelasyon açısından incelenmiştir.

NIST SP 800-22 testleri için PRNG hattında üretilen $\{A, C, G, T\}$ dizileri, sabit ve birebir bir eşleme ile ikili uzaya dönüştürülmüştür $(A \rightarrow 00, C \rightarrow 01, G \rightarrow 10, T \rightarrow 11)$. Toplamda on bağımsız akışta üretimler yapılmış ve her PRNG koşusunda yaklaşık 500.000 bazlık DNA anahtar akışı üretilmiş ve bu akış yaklaşık 1 Mbit uzunluğunda ikili test dizisine dönüştürülmüştür. NIST SP 800-22 kapsamında 15 test α=0,01 anlamlılık düzeyinde yürütülmüş ve p-değerlerinin beklenen aralıkta kalıp kalmadığı incelenmiştir [54]. On bağımsız akış için test başına ortalama p-değeri, standart sapma ve minimum–maksimum aralıkları Tablo 4 verilmiştir.

Çoklu sonuç üreten NIST alt testlerinde (Test 11: 2, Test 14: 8, Test 15: 18 alt sonuçlar alt çıktılar üzerinden değerlendirilmiş, tabloda ise karşılaştırılabilirliği korumak amacıyla özet değerler raporlanmıştır. Tablo 4'te görüldüğü üzere, ortalama p-değerleri genel olarak beklenen aralıkta dağılmış ve testlerin reddedilmesine yol açacak sistematik bir sapma gözlenmemiştir. Ortalama p-değerlerinin yaklaşık 0,37–0,73 bandında yer alması, farklı NIST alt testleri arasında tek yönlü bir yoğunlaşma olmadığını göstermektedir. Bu sonuçlar, PRNG çıktılarının NIST SP 800-22 testleri bakımından kabul edilebilir istatistiksel davranış sergilediğini desteklemektedir.

NIST SP 800-90B hattında min-entropi alt sınırları, IID kestirimi ve çevrim içi sağlık testleri değerlendirilmiştir. Bu kapsamda MCV, t-tuple ve Markov tabanlı kestirimler ile RCT/APT sağlık testleri uygulanmıştır [55, 56]. Eğitimsiz ve referanssız Transformer

Tablo 3. Farklı dizi uzunluklarında şifreleme ve şifre çözme performansı
(Encryption and decryption performance for different sequence lengths)

Dizi uzunluğu, N (baz)	Çalışma zamanı (ms)		Ek RAM kullanımı (MB)		CPU tüketimi (s)		Verimlilik (baz/ms)	
	Şifreleme	Çözme	Şifreleme	Çözme	Şifreleme	Çözme	Şifreleme	Çözme
1×10^3	0,152	0,147	~ 0	~ 0	~ 0	~ 0	6578,948	6793,478
1×10^4	1,184	1,166	~ 0	~ 0	~ 0	~ 0	8449,514	8573,388
1×10^5	10,650	10,611	0,844	0,018	0,016	0,016	9390,112	9424,538
1×10^6	122,231	121,740	1,012	0,975	0,125	0,125	8181,237	8214,234
1×10^7	1269,867	1259,818	20,664	20,005	1,266	1,266	7874,839	7937,656

Tablo 4. Transformer tabanlı PRNG ile üretilen on bağımsız akış için NIST SP 800-22 p-değeri özeti
(Summary of NIST SP 800-22 p-values for ten independent streams generated by the Transformer-based PRNG)

Test Adı	Ortalama p (n=10)	Standart sapma	Minimum–Maksimum (p)
Frequency (Monobit)	0,523751	0,312607	0,121142 – 0,971282
Block Frequency (M=1000)	0,544016	0,320997	0,049074 – 0,964374
Runs	0,492654	0,332557	0,189520 – 0,990399
Longest-Run	0,449677	0,328651	0,018382 – 0,879432
Binary Matrix Rank	0,398072	0,276545	0,024931 – 0,813178
Discrete Fourier Transform	0,396079	0,340455	0,026368 – 0,948782
Non-Overlapping Templates	0,538426	0,298955	0,155757 – 0,965278
Overlapping Templates	0,604615	0,267372	0,059439 – 0,895141
Maurer's Universal Statistical	0,558861	0,311027	0,129861 – 0,969391
Linear-Complexity	0,529599	0,307287	0,010353 – 0,949881
Serial (2. p)	0,726974	0,247343	0,212012 – 0,980831
Approximate Entropy	0,689301	0,247056	0,232086 – 0,990761
Cumulative-Sums	0,518816	0,298147	0,185525 – 0,900263
Random Excursions	0,549990	0,283216	0,223034 – 0,981868
Random Excursions Variant	0,373028	0,325800	0,025401 – 0,972074

PRNG'nin on bağımsız akışında min-entropi alt sınırı $H_{min} > 0{,}9932$ bulunmuş; ortalama değer 0,9958 ± 0,0012 olarak ölçülmüştür. IID, RCT ve APT testleri tüm akışlarda başarılıdır. Markov temelli bağımlılık ölçümleri ise düşük düzeyde kalmıştır. Bu sonuçlar, PRNG çıktılarının istatistiksel kalite ve çevrim içi sağlık testleri bakımından hedeflenen koşullarla uyumlu olduğunu göstermektedir.

DNA-özgü metrikler kapsamında, yaklaşık 500.000 baz uzunluğundaki anahtar dizileri için on bağımsız akış değerlendirilmiştir. Toplam ve yerel GC oranları yaklaşık 0,50 düzeyinde bulunmuş; 10.000 ve 50.000 bazlık pencere medyanları yaklaşık 0,498–0,501 aralığında gözlenmiştir. Homopolimer koşuları en fazla 5 ile sınırlı kalmış ve kısa gecikmeli otokorelasyon zayıf düzeyde bulunmuştur. Lag-1 otokorelasyon için $|\,r\,| < 0{,}01|$ düzeyi elde edilmiştir. Yüksek düzeyli ölçütlerde 6-mer entropisi $H(6) \approx 12$ ile kuramsal üst sınıra yakın bulunmuş, normalize Lempel–Ziv karmaşıklığı ise yaklaşık 0,985–0,987 aralığında ölçülmüştür. Bu bulgular, anahtar akışlarının DNA-özgü dağılım ve kısa menzilli bağımlılık ölçütleri açısından dengeli bir davranış sergilediğini göstermektedir.

Şifreli DNA akışları için kuaterner metrikler Tablo 5'te özetlenmiştir. Küçük örnek hacminde, özellikle 1×10³ baz düzeyinde GC oranı, k-mer ve lag-1 otokorelasyon göstergelerinde örnekleme kaynaklı uyarılar gözlenmiştir. Ancak örnek hacmi 1×10⁵ baz ve üzerine çıktığında GC ve k-mer göstergeleri belirgin biçimde yakınsamış, lag-1 otokorelasyon değerleri eşik altına inmiş ve metrikler kararlı bir rejime geçmiştir.

Tablo 5'teki bulgular birlikte değerlendirildiğinde, PRNG kaynaklı küresel bir sapma gözlenmemiştir. GC, k-mer ve ACF göstergeleri örnek hacmi arttıkça eşik altına yakınsamış; şifreli akışta düzenli biçimde artan başlıca gösterge L_{max} olmuştur. Bu artış, yerel $\bigoplus_4$ işleminin çıktı uzayını {A, C, G, T} içinde koruması altında ortaya çıkan beklenen ekstrem-değer etkisiyle tutarlıdır ve tek başına kriptografik bir zafiyet göstergesi olarak yorumlanmamıştır. Tablo 5'te "EV-beklenen" ifadesi, bağımsız ve eş-olasılıklı dört harfli dizilerde en uzun homopolimer koşusunun örnek hacmiyle birlikte beklenen ekstrem-değer davranışını göstermektedir. Bu nedenle $homopolimer \leq 5$ sınırı yalnızca PRNG kaynaklı anahtar akışı üretiminde uygulanmıştır. Şifreli DNA akışlarında ise L_{max} değeri, sabit bir üst sınırla değil, dizi uzunluğuna bağlı beklenen ekstrem-değer davranışıyla birlikte değerlendirilmiştir.

2.5.3. Güvenlik /savunma analiz testleri
(Security / defense analysis tests)

Önerilen yöntem, DNA-yerel senkron bir akış şifreleme şemasıdır. Şifreli çıktı C, düz metin DNA dizisi P ile Transformer tabanlı DNA-PRNG tarafından elde edilen hizalanmış anahtar akışı K'nın kuaterner XOR $(\bigoplus_4)$ işlemiyle birleştirilmesi sonucunda üretilmektedir. Anahtar akışı, her şifreleme oturumunda benzersiz IV/nonce ve koşu tohumu ile belirlenen 128 bazlık çekirdek anahtar bloğu K_0 üzerinden elde edilmekte ve mesaj uzunluğuna hizalanmaktadır. Bu değerlendirmede, anahtar akışının mesajlar arasında yeniden kullanılmaması temel güvenlik ilkesi olarak kabul edilmiştir. Güvenlik/savunma analizleri; çığ etkisi, anahtar duyarlılığı, seçilmiş düz metin saldırısı (Chosen-Plaintext Attack, CPA), bilinen düz metin saldırısı (Known-Plaintext Attack, KPA) ve hata yayılımı davranışı üzerinden yürütülmüştür. Bu metrikler, kriptografik ispat yerine savunma odaklı deneysel değerlendirme sunmaktadır. Beklenen kuramsal aralıklar Tablo 6'da, hata yayılımı referans bantları Tablo 7'de, ölçülen deneysel sonuçlar ise Tablo 8'de sunulmuştur.

- *Çığ Etkisi (Avalanche)*: Çığ etkisi analizinde amaç, düz metindeki en küçük değişimin şifreli çıktıya nasıl yansıdığını incelemektir. Kuaterner XOR işlemi yerel ve pozisyon koruyucu olduğundan, düz metindeki tek bazlık değişim şifreli çıktıda yalnızca aynı konumu etkilemektedir. Bu nedenle baz düzeyindeki beklenen çığ etkisi N ile ters orantılıdır. Baz ve bit düzeyindeki beklenen aralıklar Eş. 3 ve Eş. 4'te verilmiştir.

$$Avalanche_{base} = \frac{2}{2N} \times 100 = \frac{1}{N} \times 100, \quad (3)$$

$$Avalanche_{bit} \in \left[\frac{1}{2N} \times 100,\ \ \frac{1}{N} \times 100\right] \quad (4)$$

Tablo 6'da verilen çığ etkisi değerleri bu beklenen davranışla uyumludur. Düşük çığ etkisi, kullanılan yerel $\bigoplus_4$ çekirdeğinin doğal sonucudur ve tek başına kriptografik zafiyet göstergesi olarak yorumlanmamıştır. Buna karşılık anahtar akışı düzleminde, koşu tohumu üzerinde yapılan tek bit değişikliklerinde bit düzeyi fark oranı yaklaşık $p \approx 0{,}5$ olarak ölçülmüş ve bu durum anahtar akışı tarafındaki duyarlılığı desteklemiştir.

- *Anahtar Duyarlılığı (Key Sensitivity):* Anahtar duyarlılığı analizinde, 128 bazlık çekirdek anahtar bloğundaki küçük bir değişimin şifre çözme çıktısını ne ölçüde bozduğu incelenmiştir. Periyodik hizalama varsayımı altında çekirdek anahtar bloğundaki tek bazlık değişim, mesaj içinde aynı konuma karşılık gelen tekrar pozisyonlarını etkilemektedir. Bu nedenle etkilenen pozisyon sayısı yaklaşık $\lfloor N/L_k \rfloor$ ile $\lceil N/L_k \rceil$ arasındadır. Buna göre baz düzeyindeki etkilenme oranı $\frac{\lfloor N/L_k \rfloor}{N} \times 100$ ile $\frac{\lceil N/L_k \rceil}{N} \times 100$ arasında beklenmektedir. Bit düzeyindeki etkilenme oranı ise her etkilenen bazda 1–2 bit değişebileceği için $\frac{\lfloor N/L_k \rfloor}{2N} \times 100$ ile $\frac{\lceil N/L_k \rceil}{N} \times 100$ arasında beklenmektedir. $L_k = 128$ için bu değerler baz düzeyinde yaklaşık %0,781, bit düzeyinde ise yaklaşık %0,391–%0,781 bandına karşılık gelmektedir. Fiilî kullanımda çekirdek anahtar bloğu K_0, her oturumda benzersiz IV/nonce ve koşu tohumu ile elde edildiğinden anahtar akışının mesajlar arasında yeniden kullanımı önlenmektedir.

Tablo 5. Şifreli DNA akışları için kuaterner metrik özeti (Summary of quaternary metrics for encrypted DNA streams)

Dizi uzunluğu, N (baz)	GC	k-mer (k=1–3)	ACF lag-1	Homopolimer maksimumu	Genel not
1×10³	0,521 (Uyarı)	0,023 / 0,017 / 0,010 (Uyarı)	C: 0,050 (Uyarı)	5 (Geçti)	Küçük-N örnekleme sapması; çoklu uyarı
1×10⁴	0,504 (Geçti)	0,003 / 0,007 / 0,003 (Geçti)	C: 0,026 (Uyarı)	7 (EV-beklenen)	ACF-C sınıra yakın; L_{max} log-N
1×10⁵	0,500 (Geçti)	0,003 / 0,003 / 0,002 (Geçti)	≤ 0,014 (Geçti)	9 (EV-beklenen)	Mükemmel yakınsama; yalnız L_{max}
1×10⁶	0,500 (Geçti)	0,000 / 0,002 / 0,001 (Geçti)	≤ 0,008 (Geçti)	13 (EV-beklenen)	İstatistikler ideal; L_{max} tipik
1×10⁷	0,500 (Geçti)	0,000 / 0,001 / 0,0010 (Geçti)	≤ 0,007 (Geçti)	13 (EV-beklenen)	Büyük-N'de plato; L_{max} tipik

Tablo 6'daki sonuçlar, yanlış anahtar altında şifre çözme çıktısının kararlı biçimde bozulduğunu ve anahtar duyarlılığının beklenen aralıkta gerçekleştiğini göstermektedir.

- *Seçilmiş Düz-Metin Saldırısı (Chosen-Plaintext Attack, CPA):* CPA değerlendirmesinde amaç, aynı anahtar akışının yeniden kullanılması durumunda oluşabilecek bilgi sızıntısını ölçmek ve anahtar akışı yeniden kullanım yasağının önemini deneysel olarak göstermektir. Aynı hizalanmış anahtar akışı K iki farklı düz metin için kullanıldığında $C_1 = P_1 \oplus_4 K$ ve $C_2 = P_2 \oplus_4 K$ elde edilir. Bu durumda iki şifreli çıktı arasındaki kuaterner XOR ilişkisi, Eş. 5'te verildiği gibi doğrudan iki düz metin arasındaki ilişkiye indirgenmektedir.

$$C_1 \oplus_4 C_2 = P_1 \oplus_4 P_2 \tag{5}$$

Dolayısıyla sızan bilgi, doğrudan iki düz metin arasındaki $P_1 \oplus_4 P_2$ fark yapısına bağlıdır. Deneysel CPA senaryosunda P_2, P_1 dizisinde bazların ρ oranında değiştirilmesiyle elde edilmiştir. Bu kapsamda iki metrik raporlanmıştır: baz düzeyinde fark oranı $\rho \times 100$ ve bit düzeyinde fark oranı $(2/3)\rho \times 100$. Buradaki $2/3$ katsayısı, kuaterner eşlemede tek baz değişiminin 1 bit farkı oluşturma olasılığının $2/3$, 2 bit farkı oluşturma olasılığının ise $1/3$ olmasından kaynaklanmaktadır. Bu çalışmada deneysel senaryo için $\rho = 0{,}10$ alınmış; buna karşılık baz düzeyi fark oranı yaklaşık %10, bit düzeyi fark oranı ise yaklaşık %6,667 olarak beklenmiştir. $1{\times}10^3$ – $1{\times}10^7$ baz aralığında yapılan uygulamada gözlenen ortalama bit farkı %6,66–%6,68 bandında olup kuramsal değerle uyumludur [33, 34, 55]. Tablo 6'daki CPA sonuçları, aynı anahtar akışının iki farklı düz metin için kullanılması durumunda şifreli çıktılar arasındaki ilişkinin $P_1 \oplus_4 P_2$ fark yapısını yansıttığını göstermektedir. Bu bulgu, anahtar akışı yeniden kullanımının deneysel olarak gözlenebilir bir sızıntı oluşturduğunu ve benzersiz IV/nonce kullanımının gerekliliğini desteklemektedir. Bu alt analiz, IND-CPA güvenlik kanıtı olarak değil, yeniden kullanım riskini gösteren savunma odaklı bir deneysel değerlendirmedir.

- *Bilinen Düz-Metin Saldırısı (Known-Plaintext Attack, KPA):* KPA değerlendirmesinde amaç, aynı düz metnin bağımsız iki anahtar akışı ile şifrelenmesi durumunda şifreli çıktılar arasındaki ayrışmayı ölçmek ve anahtar akışı bağımsızlığını değerlendirmektir. Aynı düz metin P, iki bağımsız anahtar akışı K_1 ve K_2 ile şifrelendiğinde $C_1 = P \oplus_4 K_1$ ve $C_2 = P \oplus_4 K_2$ elde edilir. K_1 ve K_2 bağımsız ve yaklaşık eş-olasılıklı kabul edildiğinde, her pozisyonda iki şifreli bazın aynı kalma olasılığı $1/4$, farklı olma olasılığı ise $3/4$'tür. Bu nedenle baz düzeyinde beklenen Hamming oranı yaklaşık %75'tir. Sınırlı uzunluk nedeniyle oluşabilecek sapma binom varyansı üzerinden değerlendirilmiş ve standart hata Eş. 6'da verilmiştir.

$$\sigma_{baz} = \sqrt{\frac{0{,}75 \times 0{,}25}{N}} \times 100 \tag{6}$$

Burada N, normalize edilmiş DNA dizisindeki baz sayısını göstermektedir. %95 güven aralığı yaklaşık olarak $75\% \pm 1{,}96\sigma_{baz}$ biçiminde hesaplanmıştır.

Kuaterner–ikili eşleme ($A = 00, C = 01, G = 10, T = 11$) altında bu beklenti bit düzeyinde yaklaşık %50'ye karşılık gelmektedir; ancak ana raporlama baz düzeyinde yapılmıştır. Tablo 6'da verilen beklenen KPA aralıkları, bağımsız anahtar akışları altında yaklaşık %75 baz düzeyi ve yaklaşık %50 bit düzeyi ayrışma davranışını göstermektedir. Bu aralıklar, benzersiz IV/nonce ile elde edilen bağımsız anahtar akışı varsayımını deneysel sonuçları yorumlamak için bir referans çerçevesi olarak desteklemektedir. Anlamlı sapmalar ise anahtar akışı yeniden kullanımı veya akışlar arası korelasyon açısından uyarı göstergesi olarak yorumlanmalıdır.

- *Hata Yayılımı (Fault Propagation):* Hata yayılımı analizinde amaç, iletim/kanal hatası ya da kasıtlı oynama durumunda şifreli DNA dizisinde oluşan yerel bir bozulmanın şifre çözme sonrası düz metin üzerindeki etkisini incelemektir. Kuaterner XOR tabanlı yapı yerel ve pozisyon koruyucu olduğundan, şifreli dizideki tek bazlık hata, şifre çözme sonrasında yalnızca aynı pozisyonu etkilemektedir. Bu nedenle hata etkisi küresel biçimde yayılmamakta ve blok şifrelerdeki difüzyon temelli hata yayılımından farklı bir davranış sergilemektedir. Toplam uzunluk N, şifreli dizide bozulan baz sayısı n ve hata oranı $\rho = n/N$ olmak üzere, baz düzeyindeki benzerlik yaklaşık $(1-\rho) \times 100$ olarak hesaplanmıştır. Bit düzeyinde ise tek baz değişimi 1–2 bitlik fark üretebildiğinden, beklenen benzerlik aralığı yaklaşık $[(1-\rho) \times 100,\ (1-\rho/2) \times 100]$ olarak değerlendirilmiştir. Sabit hata oranlarına göre beklenen

Tablo 6. Güvenlik ve savunma metrikleri için beklenen baz ve bit düzeyi aralıklar
(Expected base- and bit-level ranges for security and defense metrics)

Dizi uzunluğu, N (baz)	Çığ etkisi baz %	Çığ etkisi bit %	Anahtar duyarlılığı baz %	Anahtar duyarlılığı bit %	CPA bit %	KPA baz % (%95 güven aralığı)	KPA bit % (%95 güven aralığı)	Hata benzerliği baz %	Hata benzerliği bit %
1×10^3	0,100	0,050 – 0,100	0,700 –0,800	0,350 – 0,800	6,800	72,316 – 77,684	47,809 – 52,191	99,50	99,50– 99,75
1×10^4	0,010	0,005 – 0,010	0,780 –0,790	0,390 – 0,790	6,615	74,151 – 75,849	49,307 – 50,693	99,95	99,95– 99,975
1×10^5	0,001	0,0005 – 0,001	0,781 –0,782	0,3905 – 0,782	6,675	74,732 – 75,268	49,781 – 50,219	99,995	99,995– 99,9975
1×10^6	0,0001	0,00005 – 0,0001	0,7812 –0,7813	0,3906 – 0,7813	6,667	74,915 – 75,085	49,931 – 50,069	99,9995	99,9995– 99,99975
1×10^7	0,00001	0,000005 – 0,00001	0,78125 –0,78125	0,390625 –0,78125	6,668	74,973 – 75,027	49,978 – 50,022	99,99995	99,99995– 99,999975

Tablo 7. Hata yayılımı için hata oranına bağlı beklenen benzerlik aralıkları
(Expected similarity ranges for fault propagation according to error rate)

ρ	Baz-benzerliği %	Bit-benzerliği % (alt–üst)
%0,1	99,9	99,9 – 99,95
%0,5	99,5	99,5 – 99,75
%1,0	99,0	99,0 – 99,5
%2,0	98,0	98,0 – 99,0

baz ve bit düzeyi benzerlik aralıkları Tablo 7'de verilmiştir. Bu tablo, hata yayılımı analizinde kullanılan referans bantları göstermekte; kriptografik güvenlik hakkında doğrudan hüküm vermemektedir. Blok şifrelerdeki tur/S-kutusu yapısı bulunmadığından, klasik Diferansiyel Hata Analizi (Differential Fault Analysis, DFA) modeli bu yapıda doğrudan uygulanmamıştır. Yerel $\bigoplus_4$ veri düzlemi nedeniyle tek sembol hatasının etkisi yerel kalmakta ve bu davranış difüzyon temelli blok şifre hata senaryolarından ayrılmaktadır [51-53].

Önerilen sistemin güvenlik/savunma davranışı, DNA dizilerinden türetilen 128 bazlık çekirdek anahtar bloğu, benzersiz IV/nonce kullanımı ve üretilen anahtar akışının istatistiksel özellikleri çerçevesinde deneysel olarak incelenmiştir. Çekirdek anahtar bloğu K_0, 128 baz uzunluğu nedeniyle $4^{128} = 2^{256}$ büyüklüğünde bir kuramsal anahtar uzayına karşılık gelmektedir. Her mesaj için benzersiz IV/nonce ve koşu tohumu kullanılması, aynı hizalanmış anahtar akışının mesajlar arasında yeniden kullanılmasını önlemeye yöneliktir. Bu tasarım ilkesi, CPA/KPA senaryolarında ortaya çıkabilecek yeniden kullanım kaynaklı sızıntıların sınırlandırılması açısından önem taşımaktadır. Tablo 6, güvenlik/savunma metrikleri için beklenen kuramsal aralıkları; Tablo 8 ise aynı metriklerin önerilen sistem üzerinde ölçülen deneysel karşılıklarını göstermektedir. Deneylerde periyodik hizalama senaryosu da ayrıca ölçülmüş ve sonuçlar kuaterner XOR mimarisinin beklenen yerel davranışıyla birlikte değerlendirilmiştir. Çığ etkisi $O(1/N)$ ölçeğinde gerçekleşmiş; N büyüdükçe tek bazlık bozulmanın göreli etkisi azalmıştır. Bu sonuç, yerel $\bigoplus_4$ çekirdeğinin zincirleme iz/bozulma üretmeyen yapısıyla uyumludur.

Tablo 8'deki bulgular birlikte değerlendirildiğinde, 1×10^3–1×10^7 baz ölçeğinde çığ etkisi ve anahtar duyarlılığı beklenen bantta gerçekleşmiştir. Anahtar duyarlılığı sonuçları, yanlış anahtar altında şifre çözme çıktısının kararlı biçimde bozulduğunu göstermektedir. CPA sonuçları, $\rho = 0{,}10$ senaryosu için beklenen yaklaşık %6,667 bit farkına yakın değerler üretmiş ve anahtar akışı yeniden kullanımının neden sakıncalı olduğunu desteklemiştir. KPA sonuçları ise ideal bağımsız ve eş-olasılıklı anahtar akışı varsayımı için beklenen %75 baz düzeyi ayrışmaya tam olarak ulaşmamış; özellikle orta ölçekli veri uzunluklarında daha düşük ayrışma değerleri göstermiştir. Bu nedenle KPA bulguları, ideal bağımsız anahtar akışı varsayımını kanıtlayan sonuçlar olarak değil, seçilen deneysel koşullar altında anahtar akışları arasındaki ayrışma davranışını gösteren tanısal ölçümler olarak değerlendirilmiştir. Hata yayılımı ise tasarım gereği yerel kalmış ve küresel zincirleme bozulma üretmemiştir. Genel olarak bu bulgular, önerilen sistemin istatistiksel ve savunma odaklı davranışını desteklemekte; ancak formel kriptografik güvenlik ispatı yerine deneysel savunma analizi olarak yorumlanmalıdır.

2.5.4. Bütünlük testleri (Integrity tests)

Bu bölümde amaç, doğru anahtar ve aynı oturum bilgileri kullanıldığında şifreleme–şifre çözme hattının kayıpsız çalıştığını doğrulamaktır. Diğer bir ifadeyle, şifre çözme sonunda elde edilen dizinin giriş dizisiyle birebir aynı olup olmadığı incelenmiştir. Tüm bütünlük kontrolleri, Bölüm 2.1'de tanımlanan FASTA-normalizasyon hattı uygulanmış diziler üzerinde yürütülmüştür. Bu hatta başlık satırları ayıklanmış, yalnızca $\{A, C, G, T\}$ bazları tutulmuş, karakterler büyük harfe dönüştürülmüş ve satır sonları standartlaştırılmıştır. Böylece satır genişliği, işletim sistemi farklılıkları ve I/O kaynaklı biçimsel sapmaların sonuçları etkilemesi önlenmiştir.

Her dizi uzunluğu için normalize edilmiş düz metin DNA dizisi P, aynı IV/nonce, koşu tohumu ve kural çizelgesi altında elde edilen anahtar akışı K ile şifrelenmiştir. Ardından aynı oturum bilgileri kullanılarak anahtar akışı deterministik olarak yeniden elde edilmiş ve şifreli DNA dizisi çözülmüştür. Şifreleme ve şifre çözme işlemleri sırasıyla $C = P \bigoplus_4 K$ ve $\hat{P} = C \bigoplus_4 K$ biçiminde yürütülmüştür. Bütünlük doğrulaması üç ölçüt üzerinden gerçekleştirilmiştir: SHA-256 özet eşitliği, baz düzeyi Hamming oranı ve bit düzeyi Hamming oranı. İlk olarak, normalize edilmiş düz metin dizisi ile şifre çözme sonunda elde edilen dizinin SHA-256 özetlerinin eşit olması beklenmiştir. Bu koşul Eş. 7'de verilmiştir.

$$SHA-256(P) = SHA-256(\hat{P}) \tag{7}$$

İkinci olarak, baz düzeyinde Hamming oranı hesaplanmıştır. Bu ölçüt, her pozisyonda özgün DNA bazı ile şifre çözme sonunda elde edilen DNA bazının aynı olup olmadığını doğrudan sınamaktadır. Baz düzeyi Hamming oranı Eş. 8'de verilmiştir.

$$H_{baz}(P, \hat{P}) = \frac{1}{N}\sum_{i=1}^{N} 1\left[P_i \neq \hat{P}_i\right] \times 100\% \tag{8}$$

Üçüncü olarak, kuaterner bazların ikili temsile dönüştürülmesiyle bit düzeyi Hamming oranı hesaplanmıştır. Bu amaçla $A = 00, C = 01, G = 10\ ve\ T = 11$ eşlemesi kullanılmıştır. Bit düzeyi Hamming oranı Eş. 9'da verilmiştir.

$$H_{bit}(P, \hat{P}) = \frac{1}{2N}\sum_{i=1}^{N} popcount\ (\varphi(P_i) \oplus \varphi(\hat{P}_i)) \times 100\% \tag{9}$$

Kuaterner XOR tabanlı şema öz-ters yapıdadır. Bu nedenle doğru anahtar akışı ve aynı oturum bilgileri kullanıldığında teorik olarak tam geri kazanım beklenmektedir. Bununla birlikte hatalı normalizasyon, satır sonu farklılıkları veya alfabe dışı semboller pratikte yanlış negatif sonuçlara yol açabileceğinden, bütünlük değerlendirmesi yalnızca tek bir ölçüte dayandırılmamıştır. SHA-256 özet eşitliği, baz düzeyi Hamming oranı ve bit düzeyi Hamming oranı birlikte kullanılmıştır. Deneylerde 1×10^3, 1×10^4, 1×10^5, 1×10^6 ve 1×10^7 baz uzunluklarının tamamında $SHA-256(P) = SHA-256(\hat{P})$ koşulu sağlanmış, $H_{baz} = 0{,}00\%$ ve $H_{bit} = 0{,}00\%$ elde edilmiştir. Ek kontrol olarak eş-karakter oranı %100 ve bit düzeyi doğruluk oranı %100,0000 bulunmuştur. Bu sonuçlar, şifreleme–şifre çözme hattının incelenen dizi uzunlukları boyunca kayıpsız ve tekrarlanabilir biçimde çalıştığını doğrulamaktadır.

3. Sonuçlar ve Tartışmalar (Results and Discussions)

Deneyler yerel bir sistemde yürütülmüştür: Python 3.12.3, Windows 11 Pro 64-bit, Intel Core i7-10750H @ 2.60 GHz ve 16 GB RAM. Tüm ölçümler aynı I/O politikası ve deterministik kipler altında alınmıştır. Zamanlama yalnızca çekirdek işlem yolunu, yani anahtar hizalama ve kuaterner XOR ($\bigoplus_4$) işlemini kapsayacak şekilde raporlanmıştır. Disk I/O, serileştirme ve dosya yazma/okuma süreleri ölçümlere dâhil edilmemiştir. PRNG tabanlı anahtar akışı üretimi ayrıca yürütülmüş ve çekirdek şifreleme/şifre çözme süresinden ayrıştırılmıştır. Böylece ölçümler, doğrudan DNA→DNA alanında çalışan şifreleme çekirdeğinin işlem maliyetini gösterecek şekilde sınırlandırılmıştır.

Performans açısından önerilen çekirdek, GC hedefi, homopolimer tavanı ve lag-1 sönümleme gibi DNA-özgü kısıtları üretim hattında gözeten anahtar akışıyla birlikte çalışmaktadır. Şifreleme ve şifre çözme aşamalarında ikili↔DNA dönüşüm maliyetleri devre dışı bırakılmış, işlem doğrudan kuaterner DNA uzayında gerçekleştirilmiştir. Ayrıntılı performans karşılaştırmaları Tablo 9'da verilmiştir. Tablo 9'daki sonuçlar, önerilen DNA-yerel $\bigoplus_4$ çekirdeğinin aynı veri boyutları üzerinde değerlendirilen literatür yöntemlerine kıyasla daha düşük çekirdek işlem süresi sunduğunu göstermektedir. Literatürde en iyi performans bildiren çalışmalar seçilmiş [8, 12, 37, 38] ve raporlanan veri boyutlarıyla uyumlu ölçeklerde yeniden karşılaştırma yapılmıştır. Küçük ve orta ölçekli verilerde belirgin hız farkları gözlenmiş; örneğin 400 karakter için yaklaşık $3{,}6\times10^3$, 20.000 karakter için ise yaklaşık $1{,}6\times10^3$ düzeyinde

Tablo 8. Eğitimsiz anahtar üretimi ile genomik şifreleme için ölçülen güvenlik ve savunma sonuçları
(Measured security and defense results for genomic encryption using untrained/reference-free key generation)

Dizi uzunluğu, N (baz)	Çığ etkisi baz %	Çığ etkisi bit %	Anahtar duyarlılığı baz %	Anahtar duyarlılığı bit %	CPA bit %	KPA baz %	KPA bit %	Hata benzerliği baz %	Hata benzerliği bit %
1×10^3	0,100	0,050	0,800	0,400	6,800	73,200	48,800	99,500	99,700
1×10^4	0,010	0,005	0,780	0,390	6,615	64,070	40,255	99,950	99,965
1×10^5	0,001	0,0005	0,781	0,3905	6,6745	69,533	45,7035	99,995	99,9965
1×10^6	0,0001	0,00005	0,7812	0,3906	6,66685	68,750	46,4843	99,9995	99,99975
1×10^7	0,00001	0,000005	0,78125	0,390625	6,667235	71,09375	48,828125	99,99995	99,999965

Tablo 9. Genomik verilerde şifreleme yöntemlerinin çekirdek işlem performansı açısından karşılaştırılması
(Core-processing performance comparison of encryption methods on genomic data)

Kaynak	Algoritma	Anahtar tipi	Anahtar yöntemi- sistemi	Anahtar uzunluğu-yapısı	Veri boyutu	Şifreleme süresi (s)	Şifre çözme süresi (s)
[5]	ECC	Asimetrik Simetrik	SM-İkili Sistem	128 bit, Statik	1 MB	25,2 s (şifreleme + şifre çözme)	
[8]	HE	Simetrik	Lojistik kaotik harita- İkili Sistem	Raporlanmamış, Statik	400 5000 10000 15000 20000 karakter	0,3 1,3 1,9 3,0 4,0	Raporlanmamış
[12]	Paillier HE	Simetrik	Paillier (GMPY2 kütüphanesi)- İkili Sistem	1024 bit, Statik	400 5000 10000 15000 20000 karakter	$1{,}8 \times 10^3$ $2{,}2 \times 10^4$ $4{,}5 \times 10^4$ $6{,}7 \times 10^4$ $9{,}1 \times 10^4$	Raporlanmamış
[37]	HE	Simetrik	Raporlanmamış	128 bit, Statik	1 MB	240 (4 min) saniye (şifreleme + şifre çözme)	
[38]	Paillier HE	Simetrik	Proxy, İkili Sistem	Raporlanmamış, Statik	400 5000 10000 15000 20000 karakter	3,2 38,6 75,1 114,7 156,5	Raporlanmamış
Bu çalışma	DNA	Simetrik	Transformer tabanlı DNA-PRNG + DNA-XOR	256 bit (128 baz), Dinamik	400 5000 10000 15000 20000 karakter 1 MB	0,0000831 0,0006139 0,0011835 0,0017555 0,0023273 0,1222309	0,0000808 0,0005956 0,0011664 0,0017445 0,0021715 0,1217399

Tablo 10. Güvenlik ve savunma analizlerinin karşılaştırılması (Comparison of security and defense analyses)

Kaynak	Anahtar duyarlılığı	CPA	KPA	Hata yayılımı
[5]	–	–	–	✓
[8]	–	–	✓	–
[12]	–	–	–	–
[37]	–	–	–	–
[38]	–	–	–	–
Bu çalışma	✓	✓	✓	✓

Not: "✓" ilgili analiz boyutunun çalışmada raporlandığını, "–" ise açık biçimde raporlanmadığını göstermektedir

hızlanma hesaplanmıştır. Bununla birlikte donanım, uygulama dili, kütüphane ve ölçüm protokollerindeki farklılıklar nedeniyle bu oranlar mutlak performans garantisi olarak değil, çekirdek işlem süresi açısından gösterge niteliğinde karşılaştırmalar olarak yorumlanmalıdır.

Bu performans farkının temel nedeni, önerilen yöntemde bit→DNA dönüşüm/kopyalama adımlarının devre dışı bırakılması ve işlemin doğrudan DNA→DNA alanında yürütülmesidir. Ayrıca düşük sabit maliyetli kuaterner XOR çekirdeği $O(N)$ zaman karmaşıklığıyla çalışmakta, PRNG tabanlı anahtar akışı üretimi ise çekirdek şifreleme/şifre çözme ölçümlerinden ayrı bir işlem yolu olarak ele alınmaktadır. Karşılaştırılan yöntemlerin bir kısmı, özellikle Paillier/HE tabanlı şemalar, ağır aritmetik işlemler ve yüksek sabit maliyetler içerdiğinden küçük ve orta boyutlu verilerde belirgin gecikme üretmektedir. Bununla birlikte donanım, uygulama dili, kütüphane ve ölçüm protokollerindeki farklılıklar nedeniyle hız oranları mutlak performans garantisi olarak değil, çekirdek işlem süresi açısından gösterge niteliğinde bir karşılaştırma olarak yorumlanmalıdır. Buna rağmen aynı veri boyutlarında farkın sistematik biçimde korunması, önerilen DNA-yerel çekirdeğin pratik açıdan belirgin bir verim üstünlüğü sunduğunu göstermektedir.

Bu çalışma, formel güvenlik ispatı iddia etmemekte; bunun yerine rastgelelik, bütünlük, anahtar duyarlılığı ve savunma metrikleri üzerinden sistem davranışını nicel olarak karakterize etmektedir. Önerilen yaklaşımda güvenli kullanım için gerekli temel ilkeler; benzersiz IV/nonce kullanımı, anahtar akışı yeniden kullanımının engellenmesi ve oturum başına yeniden üretilen anahtar akışı yönetimi olarak tanımlanmıştır. Literatürdeki çalışmalarla karşılaştırıldığında [5, 8, 12, 37, 38], önerilen yöntemde anahtar duyarlılığı, CPA, KPA ve hata yayılımı aynı değerlendirme çerçevesinde birlikte ele alınmıştır. Bu karşılaştırma Tablo 10'da özetlenmiştir.

Tablo 10'da görüldüğü üzere, karşılaştırılan çalışmaların önemli bir bölümü anahtar duyarlılığı, CPA, KPA ve hata yayılımı boyutlarını ya hiç ele almamakta ya da sınırlı biçimde değerlendirmektedir. Önerilen şema ise dört eksenin tamamında ölçüm sunmakta ve bulguları kuramsal beklentilerle birlikte raporlamaktadır. Çığ etkisinin $O(1/N)$ ölçeğinde azalması ve anahtar duyarlılığının yaklaşık $1/L_k$ bandında gerçekleşmesi, yerel kuaterner XOR çekirdeğinin beklenen davranışıyla uyumludur. CPA sonuçları, anahtar akışı yeniden kullanımının ölçülebilir sızıntı oluşturduğunu göstermiştir. KPA sonuçları ise ideal bağımsız ve eş-olasılıklı anahtar akışı beklentisine tam ulaşmadığından, kriptografik bağımsızlık kanıtı olarak değil, deneysel ayrışma davranışını gösteren tanısal ölçümler olarak değerlendirilmiştir. Bu nedenle önerilen yöntemin güvenlik değerlendirmesi, formel güvenlik ispatı yerine yeniden üretilebilir ve ölçülebilir deneysel savunma analizleri üzerinden sunulmuştur.

4. Sonuçlar (Conclusions)

Bu çalışmada, genomik verilerin doğrudan DNA alfabesi üzerinde şifrelenmesine yönelik DNA-yerel bir şifreleme ve şifre çözme hattı sunulmuştur. Önerilen yapı, eğitimsiz ve referanssız Transformer tabanlı bir DNA-PRNG ile kuaterner XOR çekirdeğini birleştirmektedir. Anahtar akışı doğrudan $\{A, C, G, T\}$ alfabesinde üretilmektedir. Bu yönüyle yöntem, ikili uzaydan DNA'ya sonradan dönüşüm yapan yaklaşımlardan ayrılmaktadır. Üretim sürecinde GC dengesi, homopolimer sınırı ve kısa gecikmeli otokorelasyon gibi DNA-özgü kısıtlar da dikkate alınmıştır. Deneysel değerlendirmeler, önerilen hattın istatistiksel kalite, bütünlük ve çekirdek işlem verimliliği açısından uygulanabilir olduğunu göstermiştir. PRNG çıktıları NIST SP 800-22, NIST SP 800-90B ve DNA-özgü ölçütlerle değerlendirilmiştir. Şifreleme–şifre çözme hattında incelenen tüm dizi uzunluklarında kayıpsız geri kazanım sağlanmıştır. Kuaterner XOR çekirdeği $O(N)$ zaman karmaşıklığıyla çalışmıştır. Bu yapı, özellikle küçük ve orta ölçekli verilerde düşük çekirdek işlem süresi elde edilmesine katkı sağlamıştır. Bununla birlikte performans bulguları mutlak bir üstünlük iddiası olarak yorumlanmamalıdır. Bu sonuçlar, donanım ve ölçüm protokolü farklılıkları dikkate alınarak çekirdek işlem düzeyinde deneysel göstergeler olarak değerlendirilmelidir. Güvenlik/savunma analizleri, önerilen yapının anahtar duyarlılığı, yeniden kullanım riski ve hata yayılımı açısından ölçülebilir sonuçlar sunduğunu göstermiştir. Çığ etkisi ve hata yayılımı, kuaterner XOR çekirdeğinin beklenen davranışıyla uyumludur. CPA sonuçları, anahtar akışı yeniden kullanımının sakıncalı olduğunu desteklemiştir. KPA bulguları ise kriptografik bağımsızlık kanıtı olarak değerlendirilmemiştir. Bu bulgular, deneysel ayrışma davranışını gösteren tanısal ölçümler olarak yorumlanmıştır. Bu nedenle çalışma, formel IND-CPA güvenlik kanıtı veya tam kriptografik güvenlik ispatı sunmamaktadır. Gelecek çalışmalarda, önerilen hattın kimlik doğrulamalı şifreleme yapılarıyla bütünleştirilmesi hedeflenmektedir. Ayrıca yöntemin formel güvenlik modeli altında incelenmesi, adaptif DNA-özgü kısıtların geliştirilmesi ve klinik/biyobanka veri akışlarında uçtan uca doğrulama yapılması da gelecek araştırma yönleri arasındadır.

Teşekkür (Acknowledgement)

Bu çalışma, Alev KAYA tarafından Fırat Üniversitesi Fen Bilimleri Enstitüsü Yazılım Mühendisliği Anabilim Dalı bünyesinde Prof. Dr. İbrahim TÜRKOĞLU danışmanlığında yürütülmekte olan "Genomik Şifreleme Yöntemlerinin Başarımını Artırmak İçin Yapay Zekâ Tabanlı Yeni Yaklaşımlar" başlıklı doktora tez çalışması kapsamında üretilmiştir.

Kaynaklar (References)

1. Al Aziz M.M., Sadat M.N., Alhadidi D., et al., Privacy-preserving techniques of genomic data-a survey, Brief Bioinform., 20, 887-895, 2019.
2. Ewejobi P., Okokpujie K., Adetiba E., et al., Homomorphic Encryption for Genomics Data Storage on a Federated Cloud: A Mini Review, 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), 1-13, 2024.
3. Bonomi L., Huang Y., Ohno-Machado L., Privacy challenges and research opportunities for genomic data sharing, Nat Genet., 52, 646-654, 2020.
4. Wan Z., Hazel J.W., Clayton E.W., et al., Sociotechnical safeguards for genomic data privacy, Nat Rev Genet., 23, 429-445, 2022.
5. Jiang Y., Shang T., Liu J., SM algorithms-based encryption scheme for large genomic data files, Digit Commun Netw., 7, 543-550, 2021.
6. Kantarcioglu M., Jiang W., Liu Y., et al., A cryptographic approach to securely share and query genomic sequences, IEEE Trans Inf Technol Biomed., 12, 606-617, 2008.
7. Canim M., Kantarcioglu M., Malin B., Secure management of biomedical data with cryptographic hardware, IEEE Trans Inf Technol Biomed., 16, 166-175, 2011.
8. Lu D., Li M., Liao Y., et al., Verifiable privacy-preserving queries on multi-source dynamic DNA datasets, IEEE Trans Cloud Comput., 11, 1927-1939, 2022.
9. Hasan M.Z., Mahdi M.S.R., Sadat M.N., Mohammed N., Secure count query on encrypted genomic data, J Biomed Inform., 81, 41-52, 2018.
10. Kızıl A., Karabulut K., Makespan and energy based virtual machine scheduling in cloud systems, Journal of the Faculty of Engineering and Architecture of Gazi University, 39 (3), 1661-1672, 2024.
11. İlgün E.G., Samet R., Increasing the performance of intrusion detection models developed using machine learning method with preprocessing applied to the dataset, Journal of the Faculty of Engineering and Architecture of Gazi University, 39 (2), 679-692, 2024.
12. Nassar M., Malluhi Q., Atallah M., et al., Securing aggregate queries for DNA databases, IEEE Trans Cloud Comput., 7, 827-837, 2019.
13. Thabit F., Alhomdy S., Jagtap S., A new data security algorithm for cloud computing based on genetics techniques and logical-mathematical functions, Int J Intell Netw., 2, 18-33, 2021.
14. Taj T.A., Hossain M.I., A multi-level random key cryptosystem based on DNA encoding and state-changing Mealy machine, J Inf Secur Appl., 83, 103760, 2024.
15. Patel S., Thanikaiselvan V., Pelusi D., et al., Colour image encryption based on customized neural network and DNA encoding, Neural Comput Appl., 33, 14533-14550, 2021.
16. Yang X., Lai L., Qiang X., et al., Towards Chinese text and DNA shift encoding scheme based on biomass plasmid storage, Front Bioinform., 3, 1276934, 2023.
17. Şatir E., Kendirli O., A symmetric DNA encryption process with a biotechnical hardware, J King Saud Univ Sci., 34, 101838, 2022.
18. Kalsi S., Kaur H., Chang V., DNA cryptography and deep learning using genetic algorithm with NW algorithm for key generation, J Med Syst., 42 (17), 2018.
19. Basu S., Karuppiah M., Nasipuri M., et al., Bio-inspired cryptosystem with DNA cryptography and neural networks, J Syst Archit., 94, 24-31, 2019.
20. Maddodi G., Awad A., Awad D., et al., A new image encryption algorithm based on heterogeneous chaotic neural network generator and DNA encoding, Multimed Tools Appl., 77, 24701-24725, 2018.
21. Tirdad K., Sadeghian A., Hopfield neural networks as pseudo random number generators, 2010 Annual Meeting of the North American Fuzzy Information Processing Society, IEEE, 1-6, 2010.
22. Wu Y., Zeng J., Dong W., et al., A novel color image encryption scheme based on hyperchaos and Hopfield chaotic neural network, Entropy, 24, 1474, 2022.
23. Jeong Y.S., Oh K.J., Cho C.K., et al., Pseudo-random number generation using LSTMs, J Supercomput., 76, 8324-8342, 2020.
24. Pasqualini L., Parton M., Pseudo random number generation through reinforcement learning and recurrent neural networks, Algorithms, 13, 307, 2020.
25. Kim H., Kwon Y., Sim M., et al., Generative adversarial networks-based pseudo-random number generator for embedded processors, Information Security and Cryptology – ICISC 2020, Cilt 12593, Editör: Hong D., Springer, Cham, 2021.
26. De Bernardi M., Khouzani M.H.R., Malacaria P., Pseudo-random number generation using generative adversarial networks, ECML PKDD 2018 Workshops, Cilt 11329, Editörler: Alzate C. ve diğerleri, Springer, Cham, 2019.
27. Chai X., Tian Y., Gan Z., et al., A robust compressed sensing image encryption algorithm based on GAN and CNN, J Mod Opt., 69, 103-120, 2022.

28. Raghuvanshi K.K., Kumar S., Kumar S., et al., Image encryption algorithm based on DNA encoding and CNN, Expert Syst Appl., 252, 124287, 2024.

29. Okada K., Endo K., Yasuoka K., et al., Learned pseudo-random number generator: WGAN-GP for generating statistically robust random numbers, PLoS One, 18 (6), 2023.

30. Duan X., Liu J., Zhang E., Efficient image encryption and compression based on VAE generative model, J Real-Time Image Process., 16, 765-773, 2019.

31. Teng Y., Yang S., Liu L., et al., Nanoscale storage encryption: Data storage in synthetic DNA using a cryptosystem with a neural network, Sci China Life Sci., 65, 1673-1676, 2022.

32. Alloun Y., Azzaz M.S., Kifouche A., A new approach based on artificial neural networks and chaos for designing deterministic random number generator and its application in image encryption, Multimed Tools Appl., 84 (9), 5825-5860, 2024.

33. Sakr A.S., Shams M.Y., Mahmoud A., et al., Amino acid encryption method using genetic algorithm for key generation, Comput Mater Continua., 70, 123-134, 2022.

34. Mukherjee P., Garg H., Pradhan C., et al., Best fit DNA-based cryptographic keys: The genetic algorithm approach, Sensors, 22, 7332, 2022.

35. Petitcolas F.A.P., Kerckhoffs' principle, Encyclopedia of cryptography, security and privacy, Editörler: Jajodia S., Samarati P., Yung M., Springer, Berlin, Heidelberg, 2023.

36. Shannon C.E., A Mathematical Theory of Cryptography, Bell System Technical Memo MM 45-110-02, 1945.

37. Schneider T., Tkachenko O., EPISODE: Efficient privacy-preserving similar sequence queries on outsourced genomic databases, Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, Auckland, New Zealand, 315-327, 2019.

38. Ghasemi R., Al Aziz M.M., Mohammed N., et al., Private and efficient query processing on outsourced genomic databases, IEEE J Biomed Health Inform., 21, 1466-1472, 2016.

39. Adleman L.M., Molecular computation of solutions to combinatorial problems, Science, 266, 1021-1024, 1994.

40. Sheela S.J., Suresh K.V., Tandur D., A novel audio cryptosystem using chaotic maps and DNA encoding, J Comput Netw Commun., 2017 (1), 2721910, 2017.

41. National Center for Biotechnology Information (NCBI). Tribolium castaneum strain Georgia GA2 linkage group LG4, whole genome shotgun sequence. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/nuccore/CM000279.2. Yayın tarihi 10 Mart 2016. Erişim tarihi 1 Ocak 2025.

42. Gymrek M., McGuire A.L., Golan D., et al., Identifying personal genomes by surname inference, Science, 339 (6117), 321-324, 2013.

43. Karaca A., Aydın Ö., Generating headlines for Turkish news texts with transformer architecture based deep learning method, Journal of the Faculty of Engineering and Architecture of Gazi University, 39 (1), 485-495, 2024.

44. Vaswani A., Shazeer N., Parmar N., et al., Attention is all you need, Adv Neural Inf Process Syst., 30, 5998-6008, 2017.

45. Lian Z., Liu B., Tao J., CTNet: Conversational transformer network for emotion recognition, IEEE/ACM Trans Audio Speech Lang Process., 29, 985-1000, 2021.

46. Han K., Xiao A., Wu E., et al., Transformer in transformer, Adv Neural Inf Process Syst., 34, 15908-15919, 2021.

47. Jiang J., Ke L., Chen L., et al., Transformer technology in molecular science, Wiley Interdiscip Rev Comput Mol Sci., 14 (4), e1725, 2024.

48. Sohal M., Sharma S., BDNA-A DNA inspired symmetric key cryptographic technique to secure cloud computing, J King Saud Univ Comput Inf Sci., 34, 1417-1425, 2022.

49. Nagy I., Suciu A., Randomness Testing with Neural Networks, 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), IEEE, 431-436, 2021.

50. Feng Y., Hao L., Testing randomness using artificial neural network, IEEE Access, 8, 163685-163693, 2020.

51. Gautam T., Jain A., Analysis of Brute Force Attack Using TG—Dataset, 2015 SAI Intelligent Systems Conference (IntelliSys), IEEE, 984-988, 2015.

52. Roy S., Shrivastava M., Pandey C.V., et al., IEVCA: An efficient image encryption technique for IoT applications using 2-D Von-Neumann cellular automata, Multimed Tools Appl., 80, 31529-31567, 2021.

53. Rajput S.K., Nishchal N.K., Known-plaintext attack on encryption domain independent optical asymmetric cryptosystem, Opt Commun., 309, 231-235, 2013.

54. Bassham L., Rukhin A., Soto J., et al., A statistical test suite for random and pseudorandom number generators for cryptographic applications (NIST Special Publication 800-22), National Institute of Standards and Technology, 2010.

55. Turan M.S., Barker E.B., Kelsey J.M., McKay K., et al., Recommendation for the entropy sources used for random bit generation (NIST Special Publication 800-90B), National Institute of Standards and Technology, 2018.

56. Buller D., Kaufer A., Roginsky A., Sönmez Turan M., Discussion on the full entropy assumption of the SP 800-90 series, NIST Interagency/Internal Report 8427, National Institute of Standards and Technology, 2023.