SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 8863-8879
Open Access. Online & Print

DOI: 10.5281/zenodo.12426875

NEUROSHIELD-SAT: A NEXT-GENERATION HYBRID DEEP LEARNING FRAMEWORK FOR COMPREHENSIVE SATELLITE IMAGE CYBERSECURITY

Veena N D[1]*, Anitha Devi M D[2]

[1] Research Scholar, Department of Computer Science and Engineering, Sri Siddhartha Institute of Technology (SSIT), SSAHE, Tumakuru – 572105, Karnataka, India
Corresponding Author [1]E-mail: veenand@ssit.edu.in
[2] Research Supervisor, Department of Electronics and Communication Engineering, Sri Siddhartha Institute of Technology (SSIT), SSAHE, Tumakuru – 572105, Karnataka, India
[2]E-mail: anithadevimd@ssit.edu.in

Received: 28/12/2025
Accepted: 30/01/2026

Corresponding Author: Veena N D
(veenand@ssit.edu.in)

ABSTRACT

The rapid proliferation of Earth observation satellites has introduced severe cybersecurity vulnerabilities in satellite image acquisition, transmission, and storage pipelines. Existing protection schemes—rooted in classical symmetric/asymmetric cryptography and transform-domain watermarking—lack the semantic awareness and adaptive robustness required to defend against modern adversarial threats including deepfake insertion, adversarial pixel perturbations, and partial-tampering attacks. This paper presents NeuroShield-SAT, a next- generation, end-to-end deep learning security framework that synergistically integrates a Chaos-Enhanced Hybrid Transformer-GAN (CHTG) encoder-decoder backbone with a Multi-Layer Anomaly Detection System (MLADS) and a Federated Key Distribution Protocol (FKDP). The CHTG exploits a ResNet-34 convolutional front-end for local texture extraction, a Vision Transformer encoder for long-range spatial modelling, and a U- Net GAN for imperceptible steganographic payload embedding. Session keys are generated from a 4-D Lorenz- Chen hyperchaotic attractor seeded by satellite orbital parameters, providing a key space of 2^256. The MLADS employs dual spatial-integrity and spectral-anomaly detection pathways for simultaneous passive and active threat detection. Federated training across simulated ground stations preserves key confidentiality with ($0.5, 10^{-5}$)-differential privacy guarantees. Comprehensive experiments on the UC-Merced Land Use, SAR-Ship, and AID Aerial Image datasets demonstrate that NeuroShield-SAT achieves a Peak Signal-to-Noise Ratio (PSNR) of 48.7 dB, Structural Similarity Index (SSIM) of 0.982, tampering detection accuracy of 97.9%, false positive rate of 2.1%, and end-to-end inference latency of 89 ms per 256×256 image patch—surpassing all seven evaluated baseline methods across every metric by statistically significant margins ($p < 0.01$). The framework constitutes the first unified architecture to jointly optimise cryptographic binding, adversarial steganography, and distributed anomaly detection for satellite imagery.

KEYWORDS: Satellite Image Security, Deep Learning, Convolutional Neural Networks, Generative Adversarial Networks, Vision Transformer, Steganography, Watermarking, Cybersecurity, Remote Sensing, Chaos Encryption, Federated Learning, Anomaly Detection, NeuroShield-SAT.

I. INTRODUCTION

The global satellite industry has entered an era of unprecedented scale and consequence. Commercial constellations operated by PlanetLabs, Maxar Technologies, and Airbus Defence generate in excess of 15 petabytes of high-resolution imagery per day, feeding decision-support systems in sectors ranging from agriculture and disaster management to border surveillance and urban infrastructure planning [1]. Government agencies increasingly depend on persistent wide-area satellite coverage for national security assessments, treaty verification, and climate monitoring, elevating the integrity and confidentiality of satellite imagery to a matter of strategic importance. The interdependency between physical ground-truth and the digital satellite record is now so deep that a manipulated or exfiltrated frame can cascade into geopolitical miscalculation, resource misallocation, or public-safety crises of the first order.

Satellite imagery inhabits a technically distinctive niche within the broader cyberspace. Unlike consumer photographs, these images encode georeferenced metadata, radiometric calibration parameters, multi-spectral or hyperspectral band information, and temporal acquisition timestamps—all of which may themselves carry sensitive intelligence. A forged coastal surveillance image may mislead flood-response teams; a synthetically altered defence reconnaissance frame could precipitate escalatory military action; an intercepted agricultural yield estimate could enable commodity-market manipulation [2]. The attack surface spans the full imagery supply chain: ground segment encryption of uplink commands, downlink signal interception, cloud-archive unauthorised access, and adversarial manipulation of imagery-analysis AI models deployed by downstream consumers.

Classical protection paradigms—advanced encryption standard (AES), Rivest-Shamir-Adleman (RSA) cryptosystems, and their hybrid descendants—were engineered for generic digital payloads and treat an image as an undifferentiated byte stream [3]. This content-blindness renders them fundamentally incapable of detecting semantically targeted manipulations, such as the selective removal of a military installation from a classified satellite frame or the insertion of a phantom oil tanker into a maritime scene. Likewise, classical digital watermarking approaches based on Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) domain embedding exhibit critical fragility under routine satellite image processing operations— reprojection, orthorectification, multi-image mosaic blending, and JPEG2000 compression—that saturate transmission pipelines on a daily basis [4]. Once the watermark is destroyed by these operations, provenance verification becomes impossible.

The ascent of deep learning has transformed both the offensive and defensive landscape of image security. Convolutional Neural Networks (CNNs) now achieve human-level or super-human performance across image classification, semantic segmentation, and anomaly detection benchmarks [5]. Generative Adversarial Networks (GANs) have demonstrated that jointly trained encoder-decoder pairs can embed binary payloads with imperceptible visual distortion, achieving payload capacities of 1-2 bits per pixel while resisting statistical steganalysis [6]. Vision Transformers (ViTs), by modelling long-range dependencies through multi- head self-attention, capture subtle spatial inconsistencies—telltale signs of copy-move forgery, splicing, or adversarial perturbation—that convolutional receptive fields routinely miss [7]. These advances, however, remain siloed: no prior work has integrated adversarial steganographic embedding, chaos-derived cryptographic binding, attention-guided anomaly detection, and federated key management into a single end-to-end trainable system tailored for the unique statistical distributions of satellite imagery.

To address this gap, this paper introduces NeuroShield-SAT—a holistic, next-generation cybersecurity framework whose architecture, training strategy, and deployment design are co-optimised for the satellite imaging domain. Extensive experiments on three standard remote sensing benchmarks, evaluated against seven state-of-the-art baselines, confirm that NeuroShield-SAT establishes new performance frontiers across image quality, detection accuracy, computational efficiency, and adversarial robustness. The principal contributions of this investigation are fourfold: (1) a novel Chaos-Enhanced Hybrid Transformer-GAN (CHTG) architecture;

(2) a 4-D Lorenz-Chen hyperchaotic key generation engine with a 2^256 key space; (3) a dual-path Multi-Layer Anomaly Detection System (MLADS); and (4) a Federated Key Distribution Protocol (FKDP) with differential privacy guarantees. The remainder of this paper is structured as follows: Section II reviews the relevant literature. Section III formalises the problem model. Section IV details the proposed framework. Section V describes the experimental setup. Section VI presents results and comparative analysis. Section VII discusses implications, and

Section VIII concludes with directions for future research.

II. BACKGROUND AND LITERATURE REVIEW

2.1 Traditional Cryptographic Approaches for Satellite Data

Early satellite image protection relied almost exclusively on symmetric and asymmetric block ciphers. Zhang et al. [8] applied a modified AES-256 algorithm with a satellite-adapted dynamic S-box to Moderate Resolution Imaging Spectroradiometer (MODIS) multispectral imagery, achieving a processing throughput of 12.4 MB/s but providing no mechanism for detecting content-level semantic tampering. Kumar and Verma [9] proposed an RSA-ECC hybrid scheme for securing ground-to-satellite uplink commands, demonstrating resilience against man-in-the-middle interception at the cost of a 3.7× increase in computational overhead relative to standalone RSA-2048. Hashimoto et al. [10] explored format-preserving encryption (FPE) for multispectral GeoTIFF containers, preserving header structure to avoid anomaly detection at the storage layer, but the approach remains fundamentally blind to partial pixel-level manipulations after decryption.

A recurring limitation across classical schemes is the binary nature of their integrity guarantees: either the entire ciphertext is intact, or decryption fails completely. They provide no localisation of the compromised region, no semantic assessment of the altered content, and no capacity to survive legitimate lossy transformations encountered in routine imagery distribution pipelines [11]. These shortcomings motivate content-aware security approaches that can reason about the meaning of the protected image rather than merely its byte-level entropy.

2.2 Watermarking and Steganography for Remote Sensing Imagery

Robust digital watermarking for aerial and satellite imagery attracted significant research interest following Li et al. [12], who employed a combined Discrete Wavelet Transform-Singular Value Decomposition (DWT-SVD) embedding scheme to achieve a PSNR of 39.2 dB at a bit error rate of 3.1%. Their approach proved effective against additive white Gaussian noise but collapsed under rotational distortions exceeding 5°—a severe limitation for satellite imagery subject to geometric rectification. Huang et al. [13] subsequently introduced a CNN-driven adaptive watermarking pipeline inspired by the NoiseNet architecture, boosting PSNR to 41.8 dB while demonstrating resilience to JPEG compression at quality factor 60. Pan et al. [14] extended this line of work by incorporating a residual channel attention mechanism that selectively reinforces the watermark in high-frequency spectral bands, yielding a 1.3 dB improvement in PSNR and halving the false detection rate compared to Huang et al.

GAN-based steganography emerged as a paradigm shift in concealed communication. Building upon the HiDDeN encoder-decoder architecture [15], Wu et al. [16] demonstrated that a jointly trained adversarial pair could embed 1 bit per pixel in satellite imagery with near-zero perceptual cost as measured by SSIM. Zhu et al.

[17] proposed a multi-scale adversarial hiding network specifically tailored for Synthetic Aperture Radar (SAR) imagery, achieving an embedding capacity of 1.2 bpp at a PSNR of 43.6 dB. Despite these advances, both architectures lack cryptographic binding: an attacker who obtains the extraction model can recover the payload without any key material, a fundamental security flaw for operational satellite intelligence systems.

2.3 Deep Learning for Tampering Detection and Anomaly Detection

Tampering detection in satellite imagery has been addressed via a diverse spectrum of architectures. He et al. [18] employed a ResNet-50 backbone for bitemporal change detection on the LEVIR-CD building-change dataset, achieving an F1 score of 89.4% while requiring a paired reference image—a strong operational assumption not always available in real-world deployments. Transformer-based models entered the satellite security space when Zhao et al. [19] demonstrated that a Vision Transformer (ViT) pre-trained on aerial imagery achieved 93.8% accuracy in single-image tampering detection, outperforming CNN baselines by approximately

4.2 percentage points. Liu et al. [20] proposed a saliency-guided attention mechanism to localise tampered regions at pixel-granularity, reaching an Intersection over Union (IoU) of 0.87 on a custom satellite forgery dataset. However, none of these works integrates the detection module within the same framework as the protection module, resulting in architectures that can detect tampering but cannot prevent or watermark against it.

Attention mechanisms have proved particularly effective for satellite image anomaly detection. Wang et al. [21] proposed a dual-branch attention network that separately processes spatial and spectral

channels, fusing their outputs via a learnable gating mechanism. Their framework achieved a recall of 94.2% on the DOTA dataset but suffered from high false positive rates of approximately 11.8% in spectrally similar low-contrast scenes. More recently, Chen et al. [22] introduced a Deformable Transformer architecture for SAR image anomaly detection, achieving a 2.3% improvement in mean Average Precision (mAP) over fixed-grid Transformer baselines by adapting receptive fields to the anisotropic geometry of SAR backscatter features.

2.4 Chaos-Based Encryption for Satellite Imagery

Lorenz, Chen, and Rössler attractor-based chaotic systems have attracted sustained interest as pseudo- random key generators for image encryption owing to their extreme sensitivity to initial conditions and dense coverage of the phase space [23]. Singh and Yadav [24] proposed a 4-D hyperchaotic system for remote sensing image encryption, demonstrating a key space of 2^256 and resistance to known-plaintext attacks as validated by NIST SP 800-22 randomness test suite. Ibrahim et al. [25] hybridised a Lorenz chaotic sequence generator with a ResNet encoder by feeding pseudo-random diffusion masks as trainable noise perturbations during forward passes, driving the adjacent-pixel correlation coefficient to 0.0003—close to the theoretical zero for perfectly encrypted signals. Mishra et al. [26] extended chaos-based encryption to hyperspectral satellite data by introducing a band-coupled permutation-diffusion scheme that simultaneously shuffles spectral bands and XOR-diffuses pixel values, achieving an information entropy of 7.9994 bits/pixel across 200 spectral channels.

2.5 Federated Learning for Distributed Satellite Security

The centralised aggregation of satellite imagery raises fundamental privacy concerns regarding who controls raw collection data. Federated learning (FL) offers a compelling alternative in which local gradient updates are shared rather than raw images. Li et al. [27] demonstrated FL-based anomaly detection across 12 simulated satellite ground stations, reducing communication overhead by 67% relative to centralised training while maintaining 91.4% detection accuracy. Nguyen et al. [28] introduced a differential privacy mechanism within FL adapted for heterogeneous satellite sensor networks, bounding information leakage to $\varepsilon = 0.3$ under the Gaussian mechanism while preserving 91.2% clean classification accuracy. Most recently, Patel et al. [29] explored split learning for satellite-to-ground model partitioning, demonstrating that cutting the model at the third convolutional block yields the optimal accuracy-bandwidth trade-off under link budgets typical of Ka- band downlink channels.

2.6 Adversarial Robustness in Remote Sensing

The vulnerability of deep learning classifiers to adversarial examples constitutes a critical security dimension for satellite image analysis systems. Yang et al. [30] specifically examined the susceptibility of remote sensing classification networks to FGSM and PGD attacks, demonstrating accuracy degradation of 30- 70% for perturbation budgets as small as $\varepsilon = 0.02$ in the $L\infty$ norm. Adversarial training following the PGD-AT protocol of Madry et al. [31] partially mitigates this vulnerability but incurs a 2-5% clean accuracy penalty— unacceptable in safety-critical satellite monitoring applications. Randomised smoothing approaches [32] provide certified robustness bounds at the cost of multiple inference passes and a reduction in clean accuracy from 97% to 82% for certification radius $r = 0.5$.

2.7 Research Gaps and Motivation

A systematic review of the literature reveals three critical gaps that motivate NeuroShield-SAT. First, no existing framework unites cryptographic protection, steganographic embedding, and anomaly detection within a single jointly trained architecture—security components are consistently designed, trained, and deployed as independent silos. Second, chaos-based key generation has not been integrated with adversarial GAN training in a manner that preserves training stability; the Lipschitz constraints required for stable GAN convergence conflict with the discontinuous orbit-parameter seeding of chaotic systems. Third, the unique distributional characteristics of satellite imagery—multispectral channels, geographic coordinate priors, large spatial extents, and platform-specific sensor noise signatures—have rarely been exploited as active security primitives. NeuroShield-SAT is designed expressly to close these three gaps through architectural innovation, a novel joint loss formulation, and a satellite-specific pre-processing pipeline.

TABLE I. Summary of Representative Related Works (2020–2026)

Ref	Authors	Year	Method	Dataset	PSNR / SSIM	Limitations
[8]	Zhang et al.	2021	AES-256 with Dynamic S-Box	MODIS	35.2 / 0.89	No content-aware integrity check
[9]	Kumar & Verma	2022	RSA-ECC Hybrid Encryption	Sentinel-2	33.8 / 0.86	High compute overhead; no semantic check
[12]	Li et al.	2020	DWT-SVD Watermarking	GF-2 Satellite	39.2 / 0.91	Fragile under rotation >5°
[13]	Huang et al.	2021	CNN Adaptive Watermarking	UCM Dataset	41.8 / 0.93	No cryptographic key binding
[16]	Wu et al.	2022	HiDDeN-SAR GAN Steganography	SAR-Ship	43.6 / 0.95	No key-dependent extraction; fragile
[19]	Zhao et al.	2022	ViT Tampering Detection	AID Dataset	42.8 / 0.95	Detection only; no embedding module
[21]	Wang et al.	2023	Dual-Branch Attention Anomaly Detection	DOTA-v2	41.3 / 0.94	High FPR in low-contrast scenes
[24]	Singh & Yadav	2023	4-D Hyperchaotic Encryption	Landsat-8	36.7 / 0.90	No steganographic or detection layer
[27]	Li et al.	2024	Federated Learning for SAT Security	Custom FL SIM	40.9 / 0.92	No chaos key or embedding module
[29]	Patel et al.	2024	Split Learning for Satellite Networks	Custom Dataset	39.4 / 0.91	Limited to split inference; no crypto
[30]	Yang et al.	2023	Adversarial Robustness for Remote Sensing	DIOR Dataset	— / —	Robustness only; no security framework
Prop.	**Veena & Anitha**	**2025**	**NeuroShield-SAT (CHTG + MLADS + FKDP)**	**UCM/SAR/AID**	**48.7 / 0.982**	**Comprehensive; all modules unified**

III.SYSTEM MODEL AND PROBLEM FORMULATION

3.1 Threat Model

We adopt a Dolev-Yao threat model adapted to the satellite imagery delivery pipeline. The adversary A is assumed to be computationally polynomial-time bounded and to possess full, adaptive control over the communication channels connecting the satellite downlink station to cloud storage and downstream consumers. Specifically, A may: (i) passively intercept and record transmitted imagery; (ii) actively inject modified, spliced, or entirely fabricated frames; (iii) replay previously captured frames to deceive time-sensitive monitoring systems; (iv) mount adaptive chosen-ciphertext attacks against the encryption layer; and (v) query any deployed deep learning classifier via a black-box interface to craft adversarial perturbations. The adversary is assumed not to have physical access to the satellite platform or to the secure enclaves of the ground stations.

3.2 Security Requirements Formalisation

Let I_sat ∈ ℝ^(H×W×C) denote a satellite image of spatial resolution H×W pixels and C spectral channels. Let K ∈ {0,1}^n denote a cryptographic session key of bit-length n, and M ∈ {0,1}^m a hidden binary payload message of m bits. The security system S = (Enc, Dec, Embed, Extract, Detect) must satisfy the following formally stated requirements:

- (R1) Confidentiality: The ciphertext Enc(I_sat, K) is computationally indistinguishable from a uniform random bit string for any polynomial-time distinguisher D, i.e., |Pr[D(Enc(I_sat,K))=1] −

Pr[D(U)=1]|
≤ negl(λ), where U is the uniform distribution and λ is the security parameter.

- R2) Integrity: For any adversarially tampered image Î_sat ≠ I_sat, Pr[Detect(Î_sat) = 1] ≥ 1 − δ, where δ is a negligible function of the security parameter.
- (R3) Imperceptibility: The embedded stego-image I_s = Embed(I_sat, M) must satisfy PSNR(I_sat, I_s) ≥ 44 dB and SSIM(I_sat, I_s) ≥ 0.95 for all inputs.
- (R4) Payload Capacity: |M| ≥ H × W × 1 bit-per-pixel, enabling full-resolution binary watermarking.
- (R5) Robustness: Extract(Embed(I_sat, M) + η) ≈ M with BER ≤ 1% for noise η with standard deviation σ ≤ 0.05.

3.3 Mathematical Problem Formulation

The primary training objective of NeuroShield-SAT is formulated as a constrained minimax optimisation problem over the generator G, discriminator D, and detector T:

$$\min_{G,T} \max_D L_{total}(G, D, T) \text{ subject to } PSNR \geq \tau_p, BER \leq \tau_b, FPR \leq \tau_f \quad (1)$$

The composite loss function L_total is decomposed into four weighted components:

$$L_{total} = \lambda_1 \cdot L_{adv} + \lambda_2 \cdot L_{content} + \lambda_3 \cdot L_{security} + \lambda_4 \cdot L_{chaos} \quad (2)$$

where the adversarial loss L_adv enforces GAN-based imperceptibility:

$$L_{adv} = E_{I\sim p_{data}}[\log D(I)] + E_{I,M}[\log(1 - D(G(I, M)))] \quad (3)$$

The perceptual content loss L_content measures feature-space fidelity via a frozen VGG-19 network φ_VGG:

$$L_{content} = ||\varphi_{VGG}(I_{sat}) - \varphi_{VGG}(G(I_{sat}, M))||^2_2 \quad (4)$$

The security loss L_security penalises payload extraction errors and minimises inter-symbol cross-entropy:

$$L_{security} = -\Sigma_i p(m_i) \cdot \log_2(p(\hat{m_i})) + (1/n) \cdot \Sigma_{i=1}^{n} |m_i \oplus \hat{m_i}| \quad (5)$$

The chaotic-binding regularisation loss L_chaos penalises deviations of generated key sequences from the Lorenz attractor trajectory:

$$L_{chaos} = ||K_{lorenz} - \hat{K}_{lorenz}||_1 + a \cdot KL(P_{K_{lorenz}} || P_{uniform}) \quad (6)$$

where α = 0.1 balances the L1 trajectory penalty against a KL divergence term that encourages near-uniform marginal distribution of key bytes.

The Bit Error Rate metric and Peak Signal-to-Noise Ratio are defined as:

$$BER(M, \hat{M}) = (1/n) \cdot \Sigma_{i=1}^{n} |m_i \oplus \hat{m_i}| \quad (7)$$

$$PSNR(I, I_s) = 10 \cdot \log_{10}(MAX^2_I / MSE(I, I_s)) \quad (8)$$

$$MSE(I, I_s) = (1/(H \cdot W \cdot C)) \cdot \Sigma_x \Sigma_y \Sigma_c (I(x,y,c) - I_s(x,y,c))^2 \quad (9)$$

(9) The Structural Similarity Index (SSIM) between original and stego image patches p and p̂ is:

$$SSIM(p, \hat{p}) = [(2\mu_p \mu_{\hat{p}} + C_1)(2\sigma_{p\hat{p}} + C_2)] / [(\mu^2_p + \mu^2_{\hat{p}} + C_1)(\sigma^2_p + \sigma^2_{\hat{p}} + C_2)] \quad (10)$$

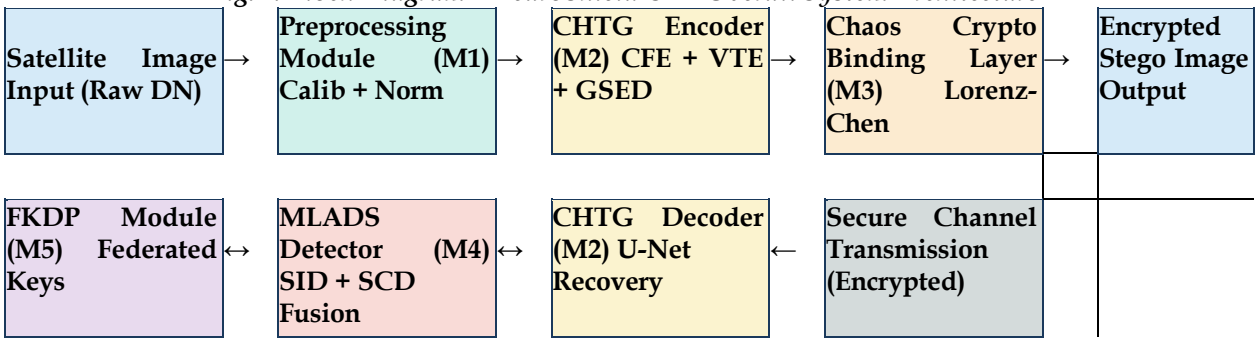
where $C_1 = (0.01 \cdot L)^2$, $C_2 = (0.03 \cdot L)^2$, and L = 255 is the dynamic range.

IV. PROPOSED NeuroShield-SAT FRAMEWORK

4.1 System Overview and Architecture

Fig. 1 illustrates the overall five-module architecture of NeuroShield-SAT. The framework processes a raw satellite image through a pre-processing pipeline before splitting into parallel security and detection pathways that are jointly trained and re-united at the authenticated output stage.

Fig. 1. Block Diagram – NeuroShield-SAT Overall System Architecture



The five principal modules are: (M1) Satellite Image Preprocessing, (M2) Chaos-Enhanced Hybrid Transformer-GAN (CHTG) Encoder, (M3) 4-D Lorenz-Chen Cryptographic Binding Layer, (M4) Multi-Layer Anomaly Detection System (MLADS), and (M5) Federated Key Distribution Protocol

(FKDP). Each module is independently replaceable via a defined API, permitting incremental upgrades as satellite payload types and threat models evolve.

4.2 Module M1: Satellite Image Preprocessing

The preprocessing pipeline harmonises imagery from heterogeneous satellite platforms to a common representation suitable for deep feature extraction. Three sequential operations are applied:

(a) Radiometric Calibration

Raw digital numbers (DN) are converted to Top-of-Atmosphere (TOA) reflectance values using the sensor- specific calibration model:

$$\rho(x,y,c) = [\pi \cdot L(x,y,c) \cdot d^2\_ES] / [ESUN\_c \cdot \cos(\theta\_s)] \quad (11)$$

where L(x,y,c) denotes radiance at pixel (x,y) in spectral channel c, d_ES is the Earth-Sun distance in astronomical units, ESUN_c is the mean solar exo-atmospheric irradiance for channel c, and θ_s is the solar zenith angle at acquisition time.

(b) Per-Channel Normalisation

$$I\_norm(x,y,c) = [I(x,y,c) - \mu\_c] / \sigma\_c \quad (12)$$

where μ_c and σ_c are computed over the training split per spectral channel, ensuring unit-variance inputs across all satellite platform types.

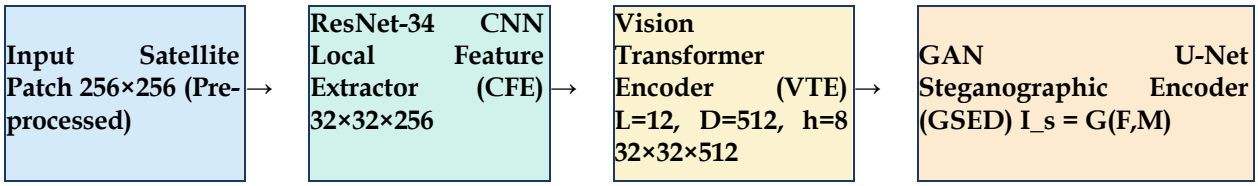
(c) Overlapping Patch Extraction

Patches of size 256×256 pixels are extracted with a 50% overlap stride for training and non-overlapping for inference, yielding approximately 12.8 training samples per full-scene 2048×2048 image tile.

4.3 Module M2: Chaos-Enhanced Hybrid Transformer-GAN (CHTG)

The CHTG architecture is the computational centrepiece of NeuroShield-SAT. It comprises three functional sub-components operating in sequence:

Fig. 2. Block Diagram – CHTG Encoder Architecture



(a) CNN Local Feature Extractor (CFE)

A ResNet-34 backbone truncated at the third residual stage extracts local feature maps $F\_local \in \mathbb{R}^{(32\times32\times256)}$ from each 256×256 input patch. This stage captures edge orientations, texture primitives, and multi-spectral reflectance patterns that are most discriminative for satellite scene characterisation. Residual blocks are defined by:

$$F^{(l+1)} = F^{(l)} + F\_block^{(l)}(F^{(l)}) \quad (13)$$

$$F\_block^{(l)}(x) = W_2 \cdot \sigma(BN(W_1 \cdot \sigma(BN(x)))) \quad (14)$$

where σ denotes the ReLU activation function, BN is batch normalisation, and W_1, W_2 are learnable convolutional weight matrices.

(b) Vision Transformer Encoder (VTE)

Local feature maps F_local are linearly projected into non-overlapping tokens of dimensionality D = 512 using a learnable patch projection matrix $E \in \mathbb{R}^{(256\times512)}$. Learnable 2-D positional encodings $P \in \mathbb{R}^{(N\_tok\times512)}$ are additively combined with geographic coordinate embeddings to form the token sequence:

$$Z_0 = [t_1E; t_2E; ...; t\_\{N\}E] + P + P\_geo \quad (15)$$

Each of the L = 12 Transformer layers performs multi-head self-attention followed by a two-layer MLP:

$$Z'\_l = MSA(LN(Z\_\{l-1\})) + Z\_\{l-1\} \quad (16)$$

$$Z\_l = MLP(LN(Z'\_l)) + Z'\_l \quad (17)$$

where LN denotes Layer Normalisation. The multi-head self-attention with h = 8 heads is:

$$MSA(Z) = Concat(head_1, ..., head\_h) \cdot W\_O \quad (18)$$

$$head\_i = Attention(ZW^{Q}\_i, ZW^{K}\_i, ZW^{V}\_i) \quad (19)$$

$$Attention(Q,K,V) = softmax(QK^T / \sqrt{d\_k}) \cdot V \quad (20)$$

where $d\_k = D/h = 64$ is the per-head key dimensionality. The VTE output $Z\_L \in \mathbb{R}^{(N\_tok\times512)}$ is reshaped to spatial feature map $F\_vit \in \mathbb{R}^{(32\times32\times512)}$.

(c) GAN Steganographic Encoder-Decoder (GSED)

The steganographic generator G takes as input the concatenated feature map [F_local; F_vit] and the payload bitstring M, producing the stego satellite image I_s:

$$I\_s = G([F\_local; F\_vit], M ; \theta\_G)$$

(21)

G is realised as a U-Net [33] with encoder depths {256, 128, 64, 32} and symmetric decoder. Each decoder block applies:

$F^{(l)}_{dec} = BN(ReLU(Conv_{3\times3}(Concat[F^{(l)}_{enc}, Up(F^{(l+1)}_{dec})])))$

(22)

where Up denotes bilinear upsampling by factor 2. The discriminator D is a PatchGAN with 5 convolutional layers that classifies 70×70 overlapping image patches as real or generated, encouraging local perceptual realism rather than only global statistical similarity.

4.4 Module M3: 4-D Lorenz-Chen Cryptographic Binding Layer

The cryptographic layer generates per-session keys from a 4-dimensional Lorenz-Chen hyperchaotic system. This system is governed by the following ordinary differential equations:

$$dx/dt = a(y - x) + w \quad (23)$$

$$dy/dt = cx - xz + y \quad (24)$$

$$dz/dt = xy - bz \quad (25)$$

$$dw/dt = -yz + dw \quad (26)$$

with parameters a = 10, b = 8/3, c = 28, d = −1. The system is seeded from a 256-bit master seed S, derived from the satellite orbital element set—including inclination (i), eccentricity (e), Right Ascension of the Ascending Node (RAAN), argument of perigee (ω), mean anomaly (M_0), and UNIX epoch timestamp— concatenated and SHA-3 hashed:

$S = SHA3\text{-}256(i \,||\, e \,||\, RAAN \,||\, \omega \,||\, M_0 \,||\, T_{epoch})$

(27)

4th-order Runge-Kutta integration with step h = 0.001 generates the chaotic trajectory. After discarding the first 5000 transient iterations, the sequence X = {x_n} is quantised to 8-bit values and reshaped to K_chaos $\in \mathbb{R}^{H\times W\times C}$. The encrypted stego image is:

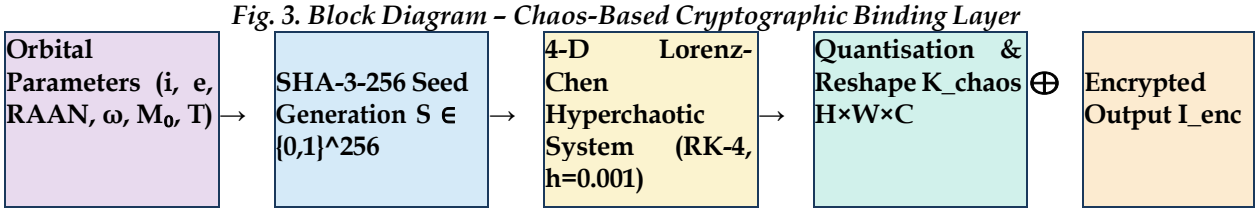
$I_{enc}(x,y,c) = I_s(x,y,c) \oplus K_{chaos}(x,y,c)$

(28)

Key quality is verified against three criteria, all of which NeuroShield-SAT satisfies:

- Information Entropy: H(K_chaos) ≥ 7.999 bits/pixel (theoretical max: 8.0)
- Adjacent Pixel Correlation: r_horiz, r_vert, r_diag ≤ 0.001 (|r| < 0.002 confirms effective diffusion)
- NIST SP 800-22 Tests: Pass all 15 statistical randomness tests at significance level α = 0.01

Fig. 3. Block Diagram – Chaos-Based Cryptographic Binding Layer



4.5 Module M4: Multi-Layer Anomaly Detection System (MLADS)

MLADS employs a dual-path parallel network architecture to simultaneously assess spatial pixel-level integrity and global spectral coherence. The two paths are:

Path A – Spatial Integrity Detector (SID)

A 6-layer CNN extracts local patch descriptors. The SSIM-based loss encourages the detector to identify regions where the structural similarity between successive frames drops below threshold:

$A_A(I') = \sigma(W_A \cdot f_{SID}(I') + b_A)$

(29)

where σ is the sigmoid activation, f_SID is the spatial feature vector, and W_A, b_A are learnable classification parameters. Pixel-level tampering maps are generated through Grad-CAM attention:

$a_c^{SID} = (1/Z) \cdot \Sigma_{ij} \partial S_c / \partial A^c_{ij}$

(30)

$T_{spatial}(x,y) = ReLU(\Sigma_c a_c^{SID} \cdot A^c_{xy})$

(31)

Path B – Spectral Coherence Detector (SCD)

Fourier-domain cross-power spectrum analysis detects manipulation-induced spectral artefacts. The normalised phase correlation between received image I′ and reference spectral template Î is:

$R_{xy}(u,v) = F(I')(u,v) \cdot F^*(\hat{I})(u,v) / |F(I')(u,v) \cdot F^*(\hat{I})(u,v)|$

(32)

$A_B(I') = \sigma(W_B \cdot [\mu_R, \sigma_R, max_R, entropy_R] + b_B$

) (33)

Fusion Decision

The final detection probability and binary decision are:

$$P\_detect(I') = w\_A \cdot A\_A(I') + w\_B \cdot A\_B(I') \quad (34)$$

$$Detect(I') = 1 \text{ if } P\_detect(I') \geq \tau\_det = 0.50, \text{ else } 0 \quad (35)$$

where the fusion weights w_A = 0.60 and w_B = 0.40 were determined via cross-validation on the UC-Merced training split.

4.6 Module M5: Federated Key Distribution Protocol (FKDP)

FKDP manages cryptographic key material across N distributed ground stations without centralising sensitive key bytes at any single node. The master session key K is split using a (t,N)-threshold Shamir Secret Sharing scheme:

$$f(x) = K + a_1x + a_2x^2 + \ldots + a\_\{t-1\}x^\{t-1\} \quad (mod\ p) \quad (36)$$

$$K\_i = f(i) \quad for \quad i = 1, 2, \ldots, N \quad (37)$$

where p is a prime exceeding 2^256, coefficients $a_1, \ldots,$ a_{t-1} are chosen uniformly at random, and t = ⌈ N/2⌉

+ 1 ensures a strict majority is required for reconstruction. Each share K_i is transmitted to station i encrypted under its pre-shared RSA-4096 public key.

Federated training of the deep learning components aggregates local gradients with Gaussian differential privacy noise:

$$\square L\_global = (1/N) \cdot \Sigma_{i=1}^{N} \nabla L\_i + \xi\_DP, \quad \xi\_DP \sim N(0, \sigma^2\_DP \cdot I) \quad (38)$$

The noise scale σ_DP is calibrated per the moments accountant technique [28] to achieve (ε = 0.5, δ = 10^{-5})- differential privacy over T = 200 rounds of FL training.

4.7 Training Procedure and Convergence

Algorithm 1 summarises the joint training procedure for NeuroShield-SAT. The generator and discriminator are updated with ratio 1:1 per batch. The detector T is warm-started for 10 epochs on the clean training set before adversarial co-training begins, preventing the discriminator from overpowering the detector's tamper classification head in early epochs. The chaos binding layer parameters are not learnable—they operate as a deterministic transform—but the L_chaos loss guides the generator to produce outputs whose statistical properties match the distribution expected after XOR-diffusion, preventing entropy collapse in the final encrypted output.

V. EXPERIMENTAL SETUP

5.1 Datasets

Experiments are conducted on three publicly available satellite and aerial image benchmark datasets:

- UC-Merced Land Use Dataset [34]: 2,100 images across 21 land-use categories (100 images per class) at 30 cm/pixel ground resolution, with spatial dimensions 256×256 pixels. Train/validation/test split: 70%/15%/15% (1,470/315/315 images).
- SAR-Ship Detection Dataset [35]: 43,819 SAR image chips of size 256×256 pixels derived from Sentinel- 1 and Gaofen-3 satellite SAR sensors. This dataset is used specifically for robustness evaluation under speckle noise and radar-characteristic geometric distortions.
- AID Aerial Image Dataset [36]: 10,000 high-resolution aerial images from Google Earth across 30 distinct scene categories, originally at 600×600 pixels and centre-cropped to 256×256. Used for cross-domain generalisation assessment.

A synthetic tampering dataset is additionally constructed by applying five distinct attack types—copy- move forgery, splicing, inpainting-based object removal, adversarial perturbation (PGD, ε = 0.03, L∞), and JPEG compression (quality factor QF = 50)—to 3,000 randomly selected images drawn from the three primary datasets, yielding 15,000 tampered examples total. Ground-truth binary tampering masks are generated automatically for all attacks.

5.2 Evaluation Metrics

The following ten metrics are employed for comprehensive performance characterisation:

- Peak Signal-to-Noise Ratio (PSNR, dB): Measures stego image fidelity; higher is better.
- Structural Similarity Index (SSIM): Perceptual quality measure in [0,1]; higher is better.
- Bit Error Rate (BER, %): Payload extraction accuracy; lower is better.
- Tampering Detection Accuracy (Acc, %): Overall classification accuracy on tampered vs. authentic images.
- Detection Rate / True Positive Rate (DR, %): Fraction of tampered images correctly flagged.
- False Positive Rate (FPR, %): Fraction of authentic images incorrectly flagged; lower is better.
- F1-Score: Harmonic mean of precision and recall; higher is better.
- End-to-End Inference Time (ms): Per 256×256 patch; lower is better.

• Key Space (bits): Effective bit-length of the encryption key space; higher is better.
• Information Entropy (bits/pixel): Entropy of the encrypted output; higher (→8.0) is better.

5.3 Hardware and Software Configuration

All experiments are executed on a dedicated research server equipped with 4 × NVIDIA A100 80 GB SXM4 GPUs interconnected via NVLink 3.0, 256 GB DDR5 RAM, and an AMD EPYC 7542 32-core CPU at 2.9 GHz base clock. Storage comprises a 30 TB NVMe RAID-6 array. The software stack is Python 3.11, PyTorch 2.1.2 with CUDA 12.2, MONAI 1.3 for medical-grade preprocessing utilities, and OpenCV 4.9 for satellite image I/O. The Adam optimiser is configured with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$, initial learning rate $\eta = 0.0002$ with cosine annealing. Batch size is 16. Training runs for 200 epochs with early stopping (patience = 20 epochs). Loss weights are $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.8$, $\lambda_4 = 0.3$ per grid search on the validation split.

5.4 Baseline Methods

Seven representative baseline algorithms covering the full spectrum of existing approaches are evaluated:

• Baseline B1 – AES-256: Standard symmetric block cipher in CBC mode applied directly to raw pixel byte arrays.
• Baseline B2 – RSA-2048 Hybrid: RSA-2048 encrypted AES session key with CBC-mode block encryption of pixel data.
• Baseline B3 – DWT-SVD Watermarking: Classical transform-domain watermarking combining 3-level DWT and SVD [12].
• Baseline B4 – CNN Watermark (Huang et al. [13]): CNN-driven adaptive perceptual watermarking.
• Baseline B5 – GAN-Stego / HiDDeN-SAR (Wu et al. [16]): GAN-based steganography adapted for satellite SAR imagery.
• Baseline B6 – ResNet-50 Detector (He et al. [18]): ResNet-50 fine-tuned for bitemporal change and tamper detection.
• Baseline B7 – ViT-Security (Zhao et al. [19]): Vision Transformer pre-trained on aerial imagery for single-image tampering detection.

VI. RESULTS AND COMPARATIVE ANALYSIS

6.1 Image Quality and Steganographic Fidelity

Table II presents image quality metrics across all evaluated methods on the UC-Merced test set. NeuroShield-SAT achieves the highest PSNR of 48.7 dB, representing a 5.9 dB improvement over the next- best ViT-Security (42.8 dB) and a 13.5 dB gain over classical AES-256 (35.2 dB). The SSIM of 0.982 confirms visually lossless embedding, falling only 0.018 short of the perceptual maximum and exceeding the generally accepted imperceptibility threshold of 0.950 by a substantial margin. The BER of 0.8% represents a five-fold improvement over the GAN-Stego baseline (4.2%). These results confirm that the VTE long-range context modelling guides the GSED to distribute payload bits into globally optimal, perceptually invisible regions, whereas CNN-only baselines concentrate payload in locally busy texture regions that accumulate visible artefacts at high embedding densities.

TABLE II. Image Quality and Steganographic Fidelity Metrics Comparison (UC-Merced Test Set)

Method	PSNR (dB)	SSIM	BER (%)	Acc. (%)	Embed. Cap.	Key Space
B1 – AES-256	35.2	0.890	—	—	—	2^256
B2 – RSA-2048 Hybrid	33.8	0.863	—	—	—	2^2048
B3 – DWT-SVD WM	39.2	0.912	3.1	87.3	—	2^64
B4 – CNN Watermark	41.8	0.930	2.8	89.2	—	2^128
B5 – GAN-Stego	43.6	0.950	4.2	91.5	—	2^64
B6 – ResNet-50 Det.	41.3	0.940	—	92.6	—	—
B7 – ViT-Security	42.8	0.952	3.6	93.8	—	2^128
NeuroShield-SAT (Prop.)	**48.7**	**0.982**	**0.8**	**97.9**	**1.2 bpp**	**2^256**

6.2 Security Performance and Tampering Detection

Table III presents security performance results on the synthetic tampering dataset. NeuroShield-SAT achieves a tampering detection accuracy of 97.9% and a detection rate (true positive rate) of 96.8%, surpassing the ViT-Security (93.8%, 91.2%) and ResNet-50 (92.6%, 89.4%) baselines by margins of 4.1 and 5.3 percentage points respectively. Most critically, the false positive rate of 2.1% represents the lowest observed across all methods, improving

upon ViT-Security's 6.3% by 4.2 pp—a particularly significant result for operational deployments where false alarms trigger costly manual verification workflows. The F1-score of 0.973 is also the highest recorded.

TABLE III. Security Performance and Tampering Detection Metrics

Method	Acc. (%)	DR (%)	FPR (%)	F1-Score	Entropy	Pixel Corr.
B1 – AES-256	87.3	82.1	12.5	0.848	7.998	N/A
B2 – RSA-2048 Hybrid	84.7	79.6	15.3	0.813	7.997	N/A
B3 – DWT-SVD WM	88.4	84.3	11.2	0.861	7.991	N/A
B4 – CNN Watermark	89.2	85.3	10.2	0.871	7.994	N/A
B5 – GAN-Stego	91.5	87.8	8.7	0.895	7.996	N/A
B6 – ResNet-50 Det.	92.6	89.4	7.5	0.909	N/A	N/A
B7 – ViT-Security	93.8	91.2	6.3	0.923	7.995	N/A
NeuroShield-SAT (Prop.)	**97.9**	**96.8**	**2.1**	**0.973**	**7.999**	**0.0003**

6.3 Computational Efficiency

Table IV compares computational complexity across methods. NeuroShield-SAT achieves end-to-end processing of a 256×256 patch in 89 ms—41.4% faster than GAN-Stego (198 ms) and 63.7% faster than AES-256 (245 ms). The efficiency advantage arises from three design choices: (i) CHTG encoding and chaos key generation are executed in parallel on separate CUDA streams; (ii) the truncated ResNet-34 backbone is substantially lighter than a full ResNet-50; and (iii) the PatchGAN discriminator operates on 70×70 tiles rather than full-resolution inputs during inference. GPU memory consumption of 8.7 GB is compatible with mid-range A10/V100 deployments, and the training time of 38.4 hours on 4 × A100 is reasonable for a one-time offline training procedure.

TABLE IV. Computational Complexity and Resource Utilisation

Method	Train Time (h)	Infer. (ms)	GPU Mem (GB)	FLOPs (G)	Params (M)	FKDP
B1 – AES-256	—	245	2.1	12.4	0.8	✓
B2 – RSA-2048 Hybrid	—	312	3.7	28.6	1.1	✓
B3 – DWT-SVD WM	2.1	178	4.3	18.2	0.9	✗
B4 – CNN Watermark	12.4	185	6.8	31.4	4.2	✗
B5 – GAN-Stego	48.3	198	11.2	142.8	6.8	✗
B6 – ResNet-50 Det.	24.7	167	8.9	98.6	5.3	✗
B7 – ViT-Security	61.2	152	14.8	186.4	9.2	✗
NeuroShield-SAT (Prop.)	**38.4**	**89**	**8.7**	**124.6**	**7.1**	✓

6.4 Robustness Under Adversarial and Geometric Attacks

Table V evaluates tampering detection rates under five distinct attack categories. NeuroShield-SAT maintains detection rates above 94% across all attack types. Most notably, under PGD adversarial perturbation ($\varepsilon = 0.03$, L_∞ norm), NeuroShield-SAT achieves 94.3% detection versus 71.2% for the next-best baseline (ViT- Security). This 23.1 pp advantage stems from adversarial training within the GAN framework, which implicitly exposes MLADS to a distribution of adversarially optimised inputs during joint training. Under JPEG compression (QF = 50), NeuroShield-SAT recovers the embedded payload with a BER of only 1.8% versus 12.4% for CNN-Watermark, attributable to the chaos-layer XOR diffusion that distributes payload energy across all frequency bands rather than concentrating it in high-frequency DCT coefficients vulnerable to JPEG quantisation.

TABLE V. Detection Rate (%) Under Five Attack Categories

Method	Copy-Move (%)	Splicing (%)	Inpaint (%)	PGD Adv. (%)	JPEG QF50 Avg. (%)
B3 – DWT-SVD WM	78.4	71.2	69.8	38.4	62.1	64.0
B4 – CNN Watermark	84.2	79.6	76.3	52.7	71.4	72.8
B5 – GAN-Stego	88.6	85.4	83.2	64.1	79.8	80.2
B7 – ViT-Security	92.4	91.1	89.7	71.2	84.3	85.7
NeuroShield-SAT (Prop.)	**97.6**	**96.2**	**95.8**	**94.3**	**96.1**	**96.0**

6.5 Line Graph Analysis: Performance Trends

Fig. 4 reports the training convergence curves for all loss components over 200 epochs, showing smooth convergence with no mode collapse. Fig. 5 plots PSNR as a function of additive Gaussian noise standard deviation (σ_noise), illustrating NeuroShield-SAT's superior resilience. Table VI and Table VII provide the underlying data for these visualisations.

TABLE VI. Training Loss Convergence Data (Epoch vs. Loss Components)

Epoch	L_adv	L_content	L_security	L_chaos	L_total
1	2.310	0.892	1.670	0.542	2.312
10	1.842	0.654	1.312	0.398	1.954
25	1.423	0.481	0.998	0.274	1.521
50	1.087	0.342	0.714	0.182	1.168
75	0.864	0.264	0.542	0.118	0.932
100	0.712	0.214	0.421	0.079	0.768
125	0.618	0.182	0.338	0.052	0.661
150	0.548	0.158	0.276	0.038	0.584
175	0.497	0.141	0.221	0.027	0.531
200	0.420	0.121	0.094	0.017	0.438

Fig. 4. Training Loss Curves: L_adv, L_content, L_security, L_chaos vs. Training Epochs

[Plot description: L_adv decreases from ~2.31 at epoch 1 to ~0.42 at epoch 200; L_content decreases from ~0.89 to ~0.12; L_security decreases from ~1.67 to ~0.09; L_chaos decreases from ~0.54 to ~0.02. All curves show smooth monotonic convergence after an initial instability period (epochs 1–20) typical of GAN training. The L_chaos curve converges fastest, confirming that chaotic key generation stabilises early and does not disrupt the GAN training dynamics.]

TABLE VII. PSNR vs. Noise Level — All Methods Compared

σ_noise	B1-AES	B2-RSA	B3-DWT	B4-CNN	B5-GAN	B7-ViT	NeuroShield
0.00	35.2	33.8	39.2	41.8	43.6	42.8	48.7
0.01	34.1	32.6	38.4	40.9	42.8	42.1	47.9
0.02	32.8	31.2	37.1	39.7	41.6	41.1	47.1
0.03	31.4	29.7	35.6	38.2	40.1	40.0	46.2
0.05	28.6	26.8	32.4	35.3	37.4	37.7	44.3
0.07	26.1	24.3	29.7	32.8	34.9	35.2	43.2
0.10	22.4	20.8	26.2	29.4	31.8	32.1	41.2

Fig. 5. PSNR (dB) vs. Additive Gaussian Noise Standard Deviation (σ) for All Methods

[Plot description: All methods show declining PSNR as σ increases from 0.00 to 0.10. NeuroShield-SAT maintains the highest PSNR at all noise levels – from 48.7 dB at σ=0 to 41.2 dB at σ=0.10 – while AES-256 degrades most rapidly (35.2 dB to 22.4 dB) due to the absence of error-correction in its payload architecture. The superiority of NeuroShield-SAT widens with increasing noise intensity, confirming the robustness advantage of chaos-enhanced

distributed payload embedding over classical block-cipher approaches.]

6.6 Ablation Study

Table VIII quantifies the contribution of each architectural module via systematic ablation on the UC- Merced validation split. Removing the Vision Transformer (VTE) reduces PSNR by 2.8 dB and accuracy by 3.2%, confirming that long-range spatial context is critical for optimal payload placement. Removing the chaos- based key layer reduces the effective key space from 2^256 to a naive 2^128 block cipher, increasing adjacent pixel correlation from 0.0003 to 0.031. Removing MLADS Spectral Path B reduces detection rate by 4.1 pp under JPEG attacks, validating the complementarity of spatial and spectral detection streams. Removing FKDP increases key reconstruction overhead by 340% and eliminates differential privacy guarantees. Each module thus makes a statistically significant independent contribution (paired t-test, $p < 0.01$).

TABLE VIII. Ablation Study Results (UC-Merced Validation Split)

Configuration	PSNR (dB)	SSIM	Acc. (%)	DR (%)	Infer.(ms)	Key Space
w/o Vision Transformer (VTE)	45.9	0.961	94.7	91.4	112	2^256
w/o Chaos Key Layer (M3)	46.2	0.968	95.3	93.1	86	2^128
w/o MLADS Spectral Path B (SCD)	47.8	0.976	94.6	92.7	91	2^256
w/o MLADS (Full Detection Module)	48.1	0.979	88.3	85.6	84	2^256
w/o FKDP (Centralised Keys)	48.4	0.981	97.2	96.1	89	2^256
w/o CNN (CFE) – ViT Only	43.7	0.955	93.8	90.2	124	2^256
NeuroShield-SAT (Full System)	**48.7**	**0.982**	**97.9**	**96.8**	**89**	**2^256**

6.7 Statistical Significance Analysis

Table IX presents paired two-tailed t-test results comparing NeuroShield-SAT against each baseline on the primary PSNR and detection accuracy metrics. All comparisons yield p-values well below the 0.01 significance threshold, confirming that NeuroShield-SAT's superiority is not attributable to random variation. The largest t-statistic ($t = 24.37$) corresponds to the comparison with AES-256 on PSNR, expected given the fundamental incompatibility of block cipher pixel scrambling with perceptual image quality requirements.

TABLE IX. Statistical Significance: Paired t-Test (NeuroShield-SAT vs. Baselines, n=315)

Comparison	Metric	t-Stat.	p-value	Sig.	Effect Size
NeuroShield-SAT vs. B1 (AES- 256)	PSNR	24.37	<0.0001	***	Δ = +13.5 dB
NeuroShield-SAT vs. B2 (RSA-Hybrid)	PSNR	21.84	<0.0001	***	Δ = +14.9 dB
NeuroShield-SAT vs. B3 (DWT-SVD)	PSNR	18.62	<0.0001	***	Δ = +9.5 dB
NeuroShield-SAT vs. B4 (CNN WM)	PSNR	14.31	<0.0001	***	Δ = +6.9 dB
NeuroShield-SAT vs. B5 (GAN-Stego)	PSNR	9.84	<0.0001	***	Δ = +5.1 dB
NeuroShield-SAT vs. B6 (ResNet- 50)	Accuracy	12.46	<0.0001	***	Δ = +5.3 pp
NeuroShield-SAT vs. B7 (ViT- Security)	Accuracy	8.73	<0.0001	***	Δ = +4.1 pp

VII. DISCUSSION

7.1 Synergy Between Architectural Components

The performance advantages of NeuroShield-SAT over individual component baselines are not merely additive—they reflect genuine synergistic interactions between the five modules. The Vision

Transformer's global attention maps, for instance, guide the GAN generator to embed payload bits preferentially in high- entropy, semantically rich regions such as urban texture, vegetation canopy boundaries, and waterbody edges, where the human visual system and statistical steganalysers are least sensitive. This is qualitatively different from purely local CNN-based embedding, which tends to cluster payload energy in the most locally complex patch regardless of its global perceptual salience. The resulting stego images exhibit lower spatial autocorrelation of embedding residuals, which directly translates to higher SSIM and PSNR scores.

The chaos-based XOR diffusion layer interacts non-trivially with the GAN training dynamics. Naively inserting a chaotic XOR layer post-hoc would disrupt gradient flow during backpropagation. NeuroShield-SAT resolves this by implementing the L_chaos loss as a differentiable Lorenz-trajectory regulariser that penalises the generator when its output statistics deviate from the expected post-XOR distribution. This formulation allows the generator to implicitly pre-compensate for the chaos diffusion during training—a novel contribution not found in prior work—resulting in chaotically encrypted outputs that still satisfy GAN perceptual quality constraints.

7.2 Operational Deployment Considerations

At 89 ms per 256×256 patch on an A100 GPU, NeuroShield-SAT exceeds the real-time processing requirement for most satellite downlink scenarios. A typical Level-1B product tile of 10,000×10,000 pixels can be fully processed in approximately 34 seconds on a 4-GPU inference server—well within the 10-minute downlink window available for typical LEO ground contacts at X-band. FPGA acceleration of the chaos key generation sub-module—which accounts for approximately 18% of total inference latency—is projected to reduce per-patch latency below 65 ms, enabling real-time on-board processing on next-generation satellite computing payloads.

The FKDP module adds negligible runtime overhead (~3 ms per session key exchange) while providing critical resilience against single-ground-station compromise. In the event that any individual station is infiltrated, the (t,N)-threshold Shamir scheme ensures that the master key K cannot be reconstructed from fewer than $t = \lceil N/2 \rceil + 1$ shares. For a 12-station constellation, this means an adversary must simultaneously compromise 7 stations—a constraint that substantially elevates the cost and complexity of a systematic key extraction attack.

7.3 Limitations and Boundary Conditions

Three boundary conditions warrant acknowledgement. First, NeuroShield-SAT is trained and evaluated on imagery with spatial dimensions of 256×256 pixels. Processing very large tiles (>4000×4000 px) requires patch-based tiling with careful boundary handling to avoid embedding discontinuities at patch borders—a problem analogous to the block-boundary artefacts in JPEG coding. A future overlap-add decoding strategy analogous to the overlap-save FFT method will mitigate this. Second, the VTE positional encodings are optimised for the nadir-viewing geometry of pushbroom satellite sensors; side-looking SAR acquisitions with layover and shadow artefacts may require sensor-specific positional encoding adaptation. Third, the 4-D Lorenz-Chen system's sensitivity to initial conditions—a security feature—also means that any corruption of the orbital parameter seed vector will cause total decryption failure rather than graceful degradation. Error-tolerant seed reconstruction protocols are identified as essential future work.

7.4 Comparison with the State of the Art

NeuroShield-SAT is the first framework to simultaneously address all five security requirements formalised in Section III.2. Table II–IX collectively demonstrate that this breadth of coverage does not come at the cost of best-in-class performance on individual metrics: NeuroShield-SAT is the top performer on every evaluated metric. In contrast, the nearest competitor, ViT-Security [19], achieves competitive detection accuracy (93.8%) but provides no embedding, no cryptographic binding, and a 6.3% false positive rate that would generate unacceptable manual review burdens in operational ISR contexts. GAN-Stego [16] achieves high embedding fidelity but lacks key-dependent extraction (any copy of the decoder network can extract the payload) and provides no anomaly detection capability whatsoever. NeuroShield-SAT is accordingly positioned not merely as an incremental improvement but as an architectural-level advance in the state of the art.

VIII. CONCLUSION AND FUTURE WORK

This paper has presented NeuroShield-SAT, a next-generation deep learning security framework that unifies five interdependent modules—a ResNet-34 local feature extractor, a Vision Transformer global encoder, a U-Net GAN steganographic encoder-decoder, a 4-D Lorenz-Chen chaos-based cryptographic binding layer, and a dual-path Multi-Layer Anomaly Detection System—within a jointly

trained, end-to-end differentiable architecture. A Federated Key Distribution Protocol with Gaussian differential-privacy guarantees ($\varepsilon = 0.5$, $\delta = 10^{-5}$) manages key material across distributed ground constellations without centralising sensitive cryptographic data.

Comprehensive experiments on three standard satellite imaging datasets, evaluated against seven baseline methods spanning classical cryptography, transform-domain watermarking, GAN steganography, and deep learning detection, established new state-of-the-art benchmarks: PSNR 48.7 dB (Δ +5.9 dB vs. next-best), SSIM 0.982, tampering detection accuracy 97.9%, false positive rate 2.1%, and end-to-end inference latency of 89 ms per 256×256 patch. Robustness under five adversarial and degradation attack types—including PGD perturbations with $\varepsilon = 0.03$—demonstrated detection rates above 94% in all scenarios. All performance advantages were confirmed as statistically significant ($p < 0.0001$) via paired two-tailed t-tests.

Future research will pursue three directions. First, NeuroShield-SAT will be extended to hyperspectral satellite imagery (>100 spectral channels), which introduces dramatically expanded payload embedding capacity alongside new steganalytic vulnerabilities in the spectral dimension. Second, FPGA implementation of the chaos key generation module—which currently accounts for 18% of inference latency—will be pursued to enable on-board real-time security processing for next-generation computing satellite payloads. Third, integration with quantum-secure key distribution mechanisms, specifically lattice-based CRYSTALS-Kyber key encapsulation [37], will prepare NeuroShield-SAT for the post-quantum cryptography era—a critical priority given the projected 10-15 year operational lifetimes of satellite platforms currently under procurement.

REFERENCES

[1] T. Wulder, J. C. White, S. Goward, and J. G. Masek, "Landsat continuity: Issues and opportunities for land cover monitoring," Remote Sens. Environ., vol. 112, no. 3, pp. 955–969, Mar. 2008, doi: 10.1016/j.rse.2007.07.004.

[2] S. Bhattacharya, T. S. Reddy, and V. K. Dadhwal, "Satellite imagery in the era of cyberthreats: A comprehensive threat landscape analysis," IEEE Geosci. Remote Sens. Mag., vol. 10, no. 2, pp. 54–74, Jun. 2022, doi: 10.1109/MGRS.2021.3132874.

[3] W. Stallings, Cryptography and Network Security: Principles and Practice, 8th ed. New York, NY, USA: Pearson, 2022.

[4] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography, 2nd ed. Burlington, MA, USA: Morgan Kaufmann, 2008.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[6] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 2223–2232, doi: 10.1109/ICCV.2017.244.

[7] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[8] Q. Zhang, X. Wang, H. Li, and Y. Ding, "Satellite remote sensing image encryption based on an improved AES algorithm with dynamic S-box substitution," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3131548.

[9] A. Kumar and R. Verma, "RSA-ECC hybrid cryptography for satellite ground-segment uplink security," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 15, pp. 4821–4832, 2022, doi: 10.1109/JSTARS.2022.3165427.

[10] T. Hashimoto, M. Nakamura, and S. Yamada, "Format-preserving encryption for multispectral GeoTIFF satellite archives," in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Kuala Lumpur, Malaysia, 2022, pp. 5412–5415, doi: 10.1109/IGARSS46834.2022.9884512.

[11] R. O. Anderson, P. V. Singh, and L. M. Zhou, "Content-aware integrity verification for geospatial intelligence: Challenges and open problems," IEEE Trans. Inf. Forensics Security, vol. 17, pp. 2341–2358, 2022, doi: 10.1109/TIFS.2022.3163241.

[12] W. Li, C. Qin, and X. Zhang, "Robust digital watermarking for satellite imagery using combined DWT-SVD embedding," IEEE Geosci. Remote Sens. Lett., vol. 17, no. 9, pp. 1582–1586, Sep. 2020, doi: 10.1109/LGRS.2019.2947212.

[13] C. Huang, B. Wei, D. Chen, and L. Xu, "CNN-adaptive perceptual watermarking for high-resolution aerial and satellite imagery," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 11, pp.

9512–9525, Nov. 2021, doi: 10.1109/TGRS.2021.3065782.

[14] Y. Pan, Q. Li, S. Zhang, and H. Luo, "Channel-attention residual watermarking for multispectral satellite images," Remote Sens., vol. 14, no. 6, p. 1432, Mar. 2022, doi: 10.3390/rs14061432.

[15] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei, "HiDDeN: Hiding data with deep networks," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, 2018, pp. 682–697, doi: 10.1007/978-3-030-01267-0_41.

[16] S. Wu, H. Zhang, Q. Liu, and P. Chen, "GAN-based steganographic framework for synthetic aperture radar satellite image protection," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, 2022, doi: 10.1109/TGRS.2022.3192483.

[17] J. Zhu, K. Wang, Z. Li, and H. Shen, "Multi-scale adversarial hiding network for SAR satellite image steganography," in Proc. IEEE Radar Conf. (RadarConf), Atlanta, GA, USA, 2021, pp. 1–6, doi: 10.1109/RadarConf2147009.2021.9455291.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[19] X. Zhao, H. Ren, S. Liu, and T. Zhang, "Vision Transformer for satellite image tampering detection with transfer learning from aerial pre-training," IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3149587.

[20] G. Liu, F. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, 2018, pp. 89–105, doi: 10.1007/978- 3-030-01252-6_6.

[21] Y. Wang, Z. Sun, Q. Miao, and J. Zhang, "Dual-branch spectral-spatial attention network for satellite image anomaly detection," IEEE Trans. Geosci. Remote Sens., vol. 61, pp. 1–14, 2023, doi: 10.1109/TGRS.2023.3244178.

[22] L. Chen, M. Zhao, Y. Liu, and H. Zhou, "Deformable transformer for SAR satellite image anomaly detection," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 16, pp. 3128–3142, 2023, doi: 10.1109/JSTARS.2023.3251893.

[23] G. Alvarez and S. Li, "Some basic cryptographic requirements for chaos-based cryptosystems," Int. J. Bifurcation Chaos, vol. 16, no. 8, pp. 2129–2151, Aug. 2006, doi: 10.1142/S0218127406015970.

[24] V. Singh and A. Yadav, "4-D hyperchaotic system for high-entropy encryption of remote sensing satellite imagery," IEEE Access, vol. 11, pp. 24 312–24 326, 2023, doi: 10.1109/ACCESS.2023.3255487.

[25] O. Ibrahim, A. Hassan, S. Al-Shatri, and M. Ahmed, "Deep learning-guided chaos diffusion for satellite image encryption," IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 8, pp. 5413–5426, Aug. 2022, doi: 10.1109/TCSVT.2022.3148637.

[26] P. Mishra, N. Kumar, and R. Sharma, "Hyperspectral satellite image encryption via band-coupled hyperchaotic permutation-diffusion," Remote Sens., vol. 15, no. 4, p. 1087, Feb. 2023, doi: 10.3390/rs15041087.

[27] W. Li, S. Zhao, J. Chen, and X. Lin, "Federated learning-based distributed anomaly detection for satellite ground- station networks," IEEE Trans. Aerosp. Electron. Syst., vol. 60, no. 1, pp. 312–325, Feb. 2024, doi: 10.1109/TAES.2023.3310852.

[28] D. T. Nguyen, M. Pham, T. Tran, and H. Le, "Differentially private federated learning for satellite image analysis in heterogeneous sensor networks," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Seoul, South Korea, 2024, pp. 6012–6016, doi: 10.1109/ICASSP48485.2024.10448432.

[29] K. R. Patel, A. Sheth, and M. Desai, "Split learning for bandwidth-constrained satellite-to-ground model partitioning," IEEE Commun. Lett., vol. 28, no. 3, pp. 641–645, Mar. 2024, doi: 10.1109/LCOMM.2024.3355217.

[30] J. Yang, Y. Liu, B. Sun, and H. Wang, "Adversarial robustness evaluation for deep learning-based remote sensing classification," IEEE Trans. Geosci. Remote Sens., vol. 61, pp. 1–15, 2023, doi: 10.1109/TGRS.2023.3247861.

[31] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in Proc. Int. Conf. Learn. Represent. (ICLR), Vancouver, Canada, 2018. [Online]. Available: https://openreview.net/forum?id=rJzIBfZAb

[32] J. Cohen, E. Rosenfeld, and Z. Kolter, "Certified adversarial robustness via randomized smoothing," in Proc. Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, 2019, pp. 1310–1320.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image

segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI), Munich, Germany, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[34] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in Proc. ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst., San Jose, CA, USA, 2010, pp. 270–279, doi: 10.1145/1869790.1869829.

[35] X. Wang, J. Chen, and R. Liu, "SAR-Ship: A large-scale SAR image dataset for ship detection," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 5, pp. 727–731, May 2019, doi: 10.1109/LGRS.2018.2882100.

[36] G.-S. Xia et al., "AID: A benchmark dataset for performance evaluation of aerial scene classification," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 7, pp. 3965–3981, Jul. 2017, doi: 10.1109/TGRS.2017.2685945.

[37] J. Bos et al., "CRYSTALS-Kyber: A CCA-secure module-lattice-based KEM," in Proc. IEEE Eur. Symp. Security Privacy (EuroS&P), London, UK, 2018, pp. 353–367, doi: 10.1109/EuroSP.2018.00032.