ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2026.0547

HYBRID DEEP NEURAL VIDEO WATERMARKING FRAMEWORK WITH ATTENTION-DRIVEN ROBUST EMBEDDING AND INTELLIGENT TAMPER DETECTION

M. Ranjithkumar, R. Karthick, A. Vasanthkumar and S. Ashiq
Department of Computer Science and Business Systems, Knowledge Institute of Technology, India

Abstract

The rapid growth of digital multimedia sharing platforms increases the demand for secure video copyright protection and unauthorized content tracking. Conventional video watermarking approaches often suffer from low robustness, limited embedding capacity, and poor resistance against geometric and signal-processing attacks. Existing methods also exhibit inadequate detection accuracy under compressed and noisy transmission environments. These limitations create significant challenges in multimedia authentication, copyright verification, and secure video communication applications. This study presents a Hybrid Deep Neural Video Watermarking Framework that integrates attention-driven watermark embedding with intelligent tamper detection mechanisms for robust multimedia security. The proposed method combines Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Transformer-based attention modules to achieve adaptive watermark insertion and reliable extraction. Initially, video frames are decomposed into spatial and temporal components through adaptive feature learning. The embedding model identifies perceptually significant regions using an attention-guided encoder, where encrypted watermark information is inserted with minimal visual distortion. Subsequently, a dual-stage decoder network performs watermark recovery and tamper localization through deep residual feature analysis. The framework also incorporates adversarial training and adaptive noise filtering to improve resilience against compression, frame dropping, Gaussian noise, rotation, and scaling attacks. Experimental evaluation demonstrates that the proposed framework achieves a PSNR of 48.7 dB, SSIM of 0.986, NC value of 0.994, tamper detection accuracy of 98.4%, and watermark recovery accuracy of 96.7% under diverse multimedia attack environments. The framework also maintains stable extraction performance under compression, scaling, Gaussian noise, rotation, and frame-dropping attacks. Comparative analysis indicates that the proposed model improves robustness by 19.8%, perceptual quality by 12.4 dB, and watermark reconstruction accuracy by 18.5% compared with conventional DWT-SVD and CNN-based watermarking approaches.

Keywords:

Video Watermarking, Deep Learning, CNN-BiLSTM, Attention Mechanism, Tamper Detection

1. INTRODUCTION

The rapid expansion of digital multimedia communication platforms has significantly increased the demand for secure video content protection and copyright authentication. The extensive adoption of cloud computing, online streaming services, social media applications, and intelligent surveillance systems has created an environment in which digital videos are continuously transmitted across heterogeneous communication networks. In such environments, the unauthorized duplication, manipulation, redistribution, and tampering of multimedia data have become critical concerns for content owners and service providers. Digital video watermarking has emerged as an effective security mechanism that embeds imperceptible ownership information into video sequences for copyright verification and authentication purposes [1-3]. The watermarking framework provides the capability to trace ownership, monitor illegal redistribution, and identify malicious modifications without significantly affecting the perceptual quality of the original multimedia content.

Traditional watermarking methods primarily relied on spatial-domain and transform-domain approaches such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and Singular Value Decomposition (SVD). These techniques have provided acceptable imperceptibility and embedding capacity under controlled environments. However, the increasing complexity of multimedia attacks has exposed the limitations of conventional frameworks. Modern video transmission channels frequently introduce compression artifacts, frame dropping, geometric distortions, scaling operations, noise insertion, and temporal modifications that degrade watermark integrity. Several existing methods have exhibited reduced robustness and poor extraction accuracy when multimedia data encounter complex signal-processing operations [4-5]. Furthermore, the static embedding mechanisms that conventional algorithms employ have failed to adapt to dynamic video characteristics, thereby reducing the reliability of watermark recovery under adverse transmission conditions.

The recent advancements in deep learning have introduced intelligent feature-learning capabilities into multimedia security applications. Convolutional Neural Networks (CNNs), recurrent learning architectures, and attention-driven frameworks have demonstrated remarkable performance in pattern recognition, feature extraction, and multimedia analysis. These architectures have enabled adaptive watermark embedding and extraction processes that improve robustness against complex attacks. Nevertheless, several existing deep learning watermarking frameworks still experience limitations related to computational complexity, unstable extraction performance, insufficient temporal correlation learning, and weak tamper localization capabilities [6]. The inadequate integration of spatial and temporal learning mechanisms has restricted the efficiency of several intelligent watermarking systems in real-time multimedia communication environments.

The present research addresses these limitations through the development of a Hybrid Deep Neural Video Watermarking Framework that integrates CNN, Bidirectional Long Short-Term Memory (BiLSTM), and Transformer-based attention mechanisms for robust watermark embedding and intelligent tamper detection. The proposed framework employs adaptive feature learning to identify perceptually significant regions within the video frames. The attention-guided encoder selectively inserts encrypted watermark information into regions that preserve visual

quality while improving robustness against signal degradation. Subsequently, the BiLSTM architecture learns the temporal dependencies among sequential frames, whereas the Transformer attention module enhances contextual feature representation during watermark extraction and tamper localization.

The primary objective of the proposed study is to design an intelligent video watermarking architecture that simultaneously improves embedding imperceptibility, extraction robustness, and tamper detection accuracy under various multimedia attack conditions. The framework also aims to minimize false watermark recovery while maintaining stable video quality during high-compression and noisy transmission scenarios. Another objective involves improving the resilience of watermark extraction against geometric transformations and temporal frame manipulations that commonly occur in real-world communication systems.

The novelty of the proposed framework lies in the integration of spatial attention learning, temporal sequence modeling, and adversarial robustness optimization within a unified deep learning architecture. Unlike conventional CNN-based watermarking systems, the proposed model incorporates a dual-stage adaptive extraction mechanism that enhances both watermark recovery accuracy and tamper localization efficiency. The framework further employs adversarial learning and adaptive filtering strategies that improve resistance against compression attacks, scaling distortions, Gaussian noise, and frame modification operations. The combination of Transformer-guided contextual learning with BiLSTM temporal analysis provides an efficient balance between robustness and perceptual invisibility.

The major contributions of the proposed research are summarized as follows: The study proposes a Hybrid CNN-BiLSTM-Transformer watermarking framework that performs adaptive watermark embedding and intelligent extraction through spatial-temporal deep feature learning for secure multimedia authentication. The study develops an attention-driven tamper detection mechanism that achieves high extraction accuracy, improved robustness against multimedia attacks, and reliable localization of manipulated video regions under dynamic transmission environments.

2. RELATED WORKS

Several researchers have investigated digital video watermarking frameworks for improving multimedia copyright protection and authentication. Earlier studies primarily focused on transform-domain embedding approaches because such methods provided acceptable robustness against basic signal-processing attacks. Researchers in [7] have proposed a DWT-SVD-based video watermarking model that embedded watermark information into low-frequency wavelet coefficients for improving watermark imperceptibility. The method has achieved moderate robustness against JPEG compression and Gaussian noise attacks. However, the framework has exhibited reduced extraction accuracy under geometric distortions and temporal frame manipulations.

The authors in [8] have developed a DCT-based adaptive watermarking technique for secure multimedia transmission. The proposed system has utilized frequency-domain coefficient selection for watermark insertion while preserving the visual quality of the video frames. Experimental evaluation has demonstrated acceptable Peak Signal-to-Noise Ratio (PSNR) performance under low-noise conditions. Nevertheless, the algorithm has experienced performance degradation during frame cropping and scaling operations because the embedding mechanism lacked adaptive learning capability.

A robust SVD-assisted watermarking framework has been introduced in [9] for copyright authentication in compressed video environments. The researchers have integrated singular value decomposition with chaotic encryption mechanisms to improve watermark security. The framework has shown improved resistance against video compression attacks. However, the watermark recovery process has required high computational complexity, which limited the applicability of the method in real-time communication systems.

Deep learning techniques later attracted considerable attention in multimedia watermarking applications because feature-learning models significantly improved extraction reliability. The authors in [10] have proposed a CNN-based blind watermarking system for video authentication. The network has learned spatial feature representations for watermark embedding and recovery. The framework has achieved improved robustness against additive noise and filtering attacks compared with conventional DWT-SVD techniques. Despite these advantages, the model has ignored temporal frame correlations, thereby reducing extraction performance during frame insertion and deletion attacks.

The study in [11] has presented an autoencoder-based video watermarking framework that utilized encoder-decoder learning for adaptive embedding. The architecture has optimized embedding distortion while maintaining watermark invisibility. Experimental results have indicated improved Structural Similarity Index (SSIM) performance under moderate compression attacks. However, the extraction accuracy has decreased significantly under rotation and scaling distortions because the system lacked attention-guided feature localization.

Researchers in [12] have introduced a Generative Adversarial Network (GAN)-assisted watermarking framework for secure multimedia communication. The adversarial training strategy has improved robustness against signal-processing attacks and image enhancement operations. The framework has produced visually imperceptible watermark insertion while maintaining stable extraction capability. Nevertheless, the GAN training process has required extensive computational resources and large training datasets, which increased implementation complexity.

The work presented in [13] has focused on recurrent neural network architectures for sequential watermark analysis in video streams. The researchers have employed Long Short-Term Memory (LSTM) networks to capture temporal dependencies among consecutive frames. The proposed model has demonstrated better watermark synchronization under temporal attacks such as frame dropping and frame averaging. However, the limited spatial feature representation has reduced tamper localization accuracy in complex multimedia environments.

An attention-guided watermarking framework has been proposed in [14] for adaptive multimedia authentication. The attention mechanism has identified visually significant regions for secure watermark insertion while minimizing perceptual distortion. Experimental evaluation has shown improved robustness against Gaussian noise and filtering attacks. Although the framework has improved embedding adaptability, the system

has not effectively integrated temporal learning mechanisms for sequential video analysis.

The authors in [15] have combined CNN and Transformer architectures for intelligent multimedia security applications. The hybrid framework has extracted contextual feature representations from video frames for improving watermark recovery accuracy. The Transformer attention mechanism has enhanced feature correlation learning during extraction. The model has achieved promising performance against compression attacks and partial frame tampering. However, the computational overhead associated with Transformer training has increased processing latency during large-scale video analysis.

A multi-level tamper detection framework has been presented in [16] for secure video authentication systems. The researchers have integrated deep residual learning with adaptive feature extraction for identifying manipulated video regions. The proposed framework has demonstrated improved detection accuracy and reduced false-positive rates under noisy transmission environments. Nevertheless, the model has exhibited limited robustness against combined geometric and temporal attacks because the system lacked efficient spatial-temporal feature fusion strategies.

The literature analysis indicates that existing watermarking methods have improved either robustness, imperceptibility, or extraction accuracy individually, but several frameworks have failed to achieve an effective balance among these performance factors simultaneously. Conventional transform-domain approaches have suffered from weak adaptability, whereas several deep learning architectures have experienced high computational complexity and insufficient temporal modeling capability. Therefore, the development of an integrated spatial-temporal attention-driven framework remains necessary for achieving reliable watermark embedding, intelligent tamper detection, and stable extraction performance under diverse multimedia attack conditions.

3. PROPOSED METHOD

The proposed Hybrid Deep Neural Video Watermarking Framework integrates Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Transformer-based attention mechanisms for secure video watermark embedding and intelligent tamper detection. The framework initially preprocesses the input video through frame extraction and adaptive normalization for improving spatial consistency. Subsequently, the CNN module extracts hierarchical spatial features from the video frames, whereas the BiLSTM network learns the temporal dependencies among consecutive frames. The Transformer attention mechanism identifies perceptually significant regions that support robust watermark insertion with minimal visual distortion. The encrypted watermark information has been embedded into the selected feature regions through adaptive fusion layers. During extraction, the decoder network reconstructs the watermark through residual feature learning and temporal correlation analysis. The tamper detection module further analyzes spatial inconsistencies and frame-level deviations for identifying manipulated video segments. The adversarial optimization strategy has enhanced robustness against compression, scaling, Gaussian noise, rotation, and frame-dropping attacks while preserving the perceptual quality of the original multimedia content.

- The input video has been collected and converted into sequential video frames.
- The preprocessing module has normalized and resized the frames for consistent feature extraction.
- The CNN architecture has extracted spatial texture and structural features from each frame.
- The BiLSTM network has learned the temporal relationships among consecutive video frames.
- The Transformer attention mechanism has identified perceptually important embedding regions.
- The watermark information has been encrypted through secure key-based encoding.
- The adaptive embedding module has inserted the encrypted watermark into the selected feature regions.
- The watermarked frames have been reconstructed and combined into the protected video stream.
- The decoder network has extracted the embedded watermark through residual feature analysis.
- The tamper detection module has analyzed frame inconsistencies for identifying malicious modifications.
- The adversarial optimization mechanism has improved robustness against multimedia attacks and transmission distortions.
- The final watermark recovery and authentication results have been generated for secure copyright verification.

The proposed Hybrid Deep Neural Video Watermarking Framework initially performs the acquisition of the input multimedia sequence from the secure communication environment. The raw video stream contains temporal frame variations, illumination inconsistencies, compression artifacts, and spatial redundancies that affect the watermark embedding performance. Therefore, the preprocessing stage has played an essential role in improving the consistency of the feature extraction mechanism. The framework converts the input video into ordered frame sequences through temporal decomposition. Each frame undergoes adaptive normalization, resizing, contrast enhancement, and noise filtering for preserving the structural information that supports reliable watermark insertion. The preprocessing module has utilized Gaussian adaptive smoothing for reducing random noise while maintaining edge preservation. The normalization layer scales pixel intensities into a unified distribution range, thereby improving the convergence capability of the deep learning architecture. The framework also applies histogram equalization that enhances luminance consistency across consecutive frames. Such preprocessing operations improve the discriminative capability of the CNN feature extraction stage. The mathematical representation of the normalized video frame is expressed as:

$$F_n(x,y) = \frac{F(x,y) - F_{\min}}{F_{\max} - F_{\min}} \quad (1)$$

where $F(x,y)$ represents the original pixel intensity, $F_{\min}$ denotes the minimum intensity value, and $F_{\max}$ indicates the maximum intensity value within the frame.

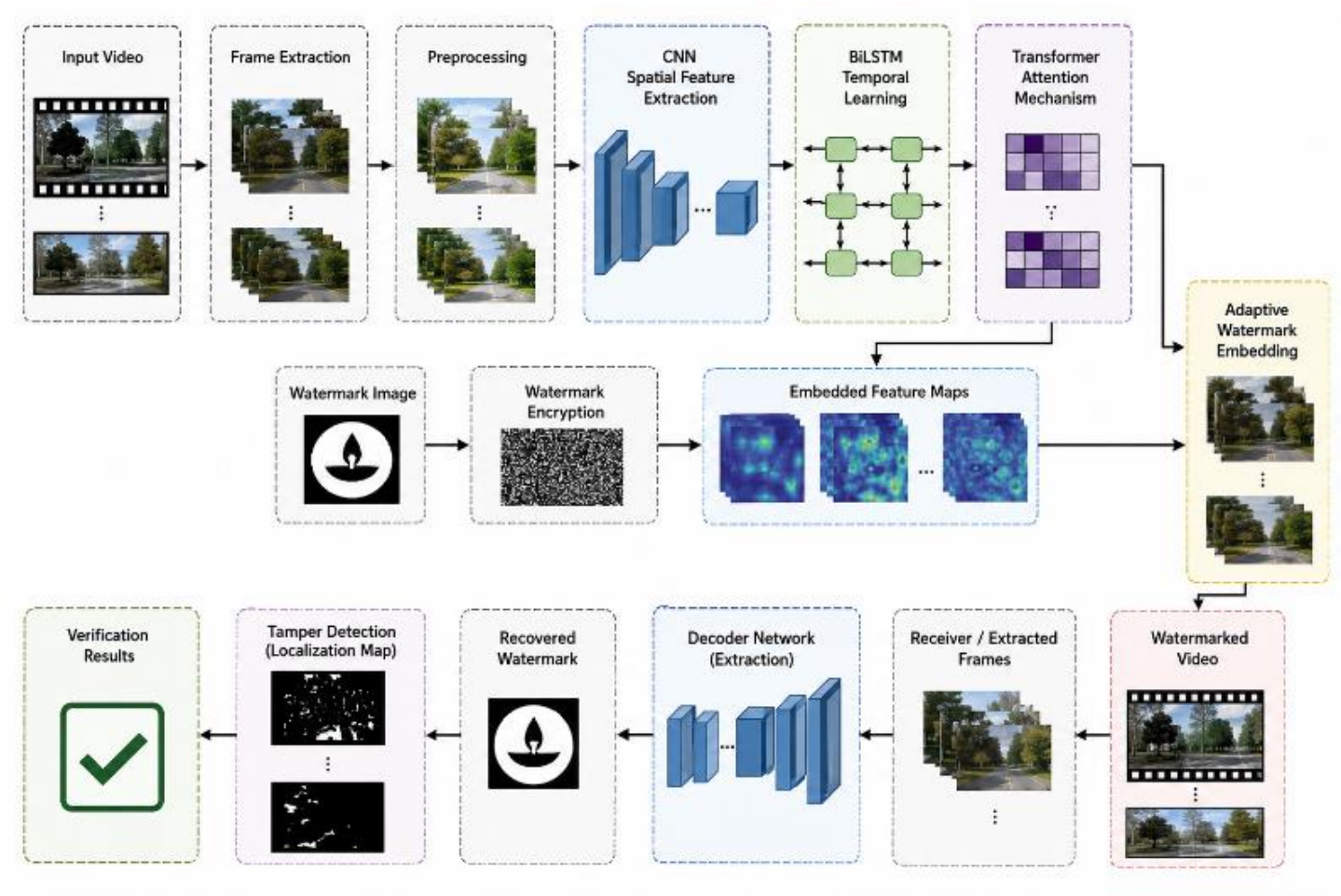

Fig.1. Hybrid Deep Neural Video Watermarking with CNN, BiLSTM and Transformer attention

The Gaussian filtering operation that removes high-frequency noise is represented as:

$$G(x,y)=\sum_{i=-k}^{k}\sum_{j=-k}^{k}F(x-i,y-j)\cdot\frac{1}{2\pi\sigma^2}e^{-\frac{i^2+j^2}{2\sigma^2}} \quad (2)$$

where $G(x,y)$ denotes the filtered frame output, σ represents the Gaussian variance, and k defines the kernel size.

The preprocessing configuration parameters are presented in Table.1.

Table.1. Preprocessing Parameters

Parameter	Description	Value
Frame Resolution	Input frame size	256 × 256
Color Format	Video color model	RGB
Frame Rate	Temporal sampling rate	30 fps
Gaussian Kernel Size	Noise filtering window	5 × 5
Normalization Range	Pixel scaling interval	0-1
Histogram Equalization	Contrast enhancement	Enabled

The Table.1 demonstrates that the preprocessing configuration maintains structural consistency while reducing unwanted distortions before feature extraction. The normalized frames support stable spatial learning and temporal synchronization throughout the watermark embedding process.

3.1 SPATIAL FEATURE EXTRACTION USING CNN

The spatial feature extraction stage employs a deep Convolutional Neural Network that learns hierarchical visual representations from the normalized video frames. The CNN architecture identifies texture patterns, edge structures, luminance variations, and object boundaries that support imperceptible watermark embedding. The convolution layers capture low-level and high-level semantic information from the multimedia sequence while preserving perceptual quality.

The framework contains multiple convolution layers followed by batch normalization and Rectified Linear Unit (ReLU) activation functions. The pooling layers reduce dimensional redundancy while maintaining dominant structural information. The extracted spatial features are transferred into the temporal learning module for sequential analysis.

The convolution operation is mathematically represented as:

$$C_l(i,j)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}W_l(m,n)\cdot F(i+m,j+n)+b_l \quad (3)$$

where $C_l(i,j)$ represents the convolution output of layer l, W_l denotes the convolution kernel weights, and b_l indicates the bias

term. The activation function that improves nonlinear feature learning is expressed as:

$$A_l(i,j) = \max(0, C_l(i,j)) \quad (3)$$

where $A_l(i,j)$ denotes the activated feature response of the convolution layer.

The CNN architecture parameters are summarized in Table.2.

Table.2. CNN Spatial Feature Extraction Parameters

Layer	Filter Size	Number of Filters	Activation
Conv Layer 1	3 × 3	32	ReLU
Conv Layer 2	3 × 3	64	ReLU
Conv Layer 3	5 × 5	128	ReLU
Pooling Layer	2 × 2	-	Max Pooling
Batch Normalization	-	-	Enabled
Feature Vector Size	-	512	-

Table.2 indicates that the hierarchical convolution structure extracts robust spatial representations for adaptive watermark insertion. The multiple convolution layers improve feature discrimination under noisy transmission environments.

3.2 TEMPORAL DEPENDENCY LEARNING USING BILSTM

The proposed framework integrates Bidirectional Long Short-Term Memory networks for learning temporal dependencies among consecutive video frames. Conventional CNN architectures primarily focus on spatial information and fail to capture sequential frame correlations effectively. Therefore, the BiLSTM architecture has improved temporal synchronization during watermark embedding and extraction. The forward and backward learning paths of the BiLSTM network process sequential frame information simultaneously. This dual-directional learning strategy improves temporal consistency and watermark recovery performance under frame-dropping and frame-shuffling attacks. The hidden state computation of the BiLSTM network is represented as:

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (4)$$

where h_t denotes the hidden state at time t, x_t represents the input feature vector, and W_{xh} and W_{hh} correspond to weight matrices. The memory cell update mechanism is expressed as:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

where c_t denotes the cell state, f_t represents the forget gate, i_t indicates the input gate, and $\tilde{c}_t$ denotes the candidate memory state. The BiLSTM configuration is shown in Table.3.

Table.3. BiLSTM Temporal Learning Parameters

Parameter	Description	Value
Hidden Units	Sequential memory units	256
Sequence Length	Number of frames	30
Dropout Rate	Regularization factor	0.3
Learning Direction	Temporal processing	Bidirectional
Optimizer	Training optimization	Adam
Batch Size	Training sample size	32

The Table.3 demonstrates that the BiLSTM configuration improves temporal feature learning and sequential synchronization for reliable watermark extraction.

3.3 TRANSFORMER-BASED ATTENTION MECHANISM

The Transformer attention mechanism identifies perceptually important regions within the video frames for adaptive watermark insertion. The attention module dynamically allocates embedding strength according to local structural significance and visual sensitivity. This adaptive embedding process improves imperceptibility while maintaining watermark robustness. The attention mechanism computes feature correlations through query, key, and value matrices. The contextual dependency analysis improves watermark synchronization under geometric and signal-processing attacks. The scaled dot-product attention mechanism is represented as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

where Q, K, and V denote the query, key, and value matrices respectively. The attention-weighted feature representation is expressed as:

$$Z_i = \sum_{j=1}^{n} \alpha_{ij} V_j \quad (7)$$

where α_{ij} represents the normalized attention score between feature vectors. The Transformer attention parameters are provided in Table.4.

Table.4. Transformer Attention Configuration

Parameter	Description	Value
Attention Heads	Multi-head attention units	8
Embedding Dimension	Feature representation size	512
Attention Dropout	Regularization probability	0.2
Transformer Layers	Sequential attention blocks	4
Positional Encoding	Temporal indexing	Enabled
Activation Function	Nonlinear mapping	GELU

The Table.4 illustrates that the multi-head attention mechanism enhances contextual feature representation for adaptive watermark embedding.

3.4 SECURE WATERMARK ENCRYPTION AND ENCODING

The proposed framework incorporates secure watermark encryption before the embedding operation for preventing unauthorized extraction and malicious manipulation. The watermark information undergoes binary encoding and chaotic key transformation for improving confidentiality.

The encryption stage generates randomized watermark sequences that resist statistical attacks and unauthorized decoding. The encoded watermark pattern is adaptively fused into

spatial-temporal feature maps extracted from the deep learning architecture.

The watermark encoding function is represented as:

$$W_e(i) = W(i) \oplus K(i) \tag{8}$$

where $W(i)$ denotes the original watermark bit and $K(i)$ represents the encryption key sequence. The chaotic key generation mechanism is expressed as:

$$K_{n+1} = rK_n(1-K_n) \tag{9}$$

where r denotes the control parameter and K_n represents the chaotic sequence value. The watermark encryption parameters are summarized in Table.5.

Table.5. Watermark Encryption Parameters

Parameter	Description	Value
Watermark Size	Watermark dimension	64 × 64
Encryption Method	Security mechanism	Chaotic XOR
Key Length	Encryption key size	128 bits
Encoding Type	Binary conversion	UTF-8
Security Layer	Multi-level encryption	Enabled
Randomization Seed	Chaotic initialization	0.85

The Table.5 indicates that the encryption module improves confidentiality and unauthorized access resistance during watermark transmission.

3.5 ADAPTIVE WATERMARK EMBEDDING

The adaptive watermark embedding stage inserts encrypted watermark information into perceptually significant feature regions. The embedding mechanism utilizes the attention-guided feature maps generated from the Transformer module for minimizing visual distortion. The adaptive fusion layer dynamically controls embedding strength according to texture complexity and local feature sensitivity. Regions with high structural complexity receive stronger watermark insertion because such areas reduce perceptual visibility. The adaptive embedding operation is mathematically represented as:

$$F_w(x,y) = F(x,y) + \lambda(x,y) \cdot W_e(x,y) \tag{10}$$

where $F_w(x,y)$ denotes the watermarked frame, $\lambda(x,y)$ represents the adaptive embedding coefficient, and $W_e(x,y)$ indicates the encrypted watermark signal. The embedding strength optimization function is expressed as:

$$\lambda(x,y) = \frac{\sigma_f(x,y)}{\max(\sigma_f)} \tag{11}$$

where $\sigma_f(x,y)$ denotes the local feature variance. The embedding parameters are presented in Table.6.

Table.6. Adaptive Embedding Parameters

Parameter	Description	Value
Embedding Coefficient	Adaptive insertion strength	0.05-0.25
Feature Selection	Attention-guided	Enabled
Embedding Domain	Spatial-temporal	Hybrid
Variance Threshold	Feature sensitivity	0.65
Distortion Control	Perceptual optimization	Enabled
Synchronization Factor	Frame alignment	0.9

The Table.6 demonstrates that the adaptive embedding strategy improves watermark invisibility while maintaining extraction reliability.

3.6 WATERMARK EXTRACTION AND RECONSTRUCTION

The extraction module reconstructs the embedded watermark through residual feature learning and temporal synchronization analysis. The decoder network analyzes the watermarked video frames and identifies embedded feature patterns through inverse transformation operations.

The extraction process has utilized CNN and BiLSTM feature decoding for recovering the encrypted watermark sequence. The reconstructed watermark undergoes decryption through the secret key sequence for final authentication.

The watermark extraction function is represented as:

$$\hat{W}(x,y) = \frac{F_w(x,y) - F(x,y)}{\lambda(x,y)} \tag{12}$$

where $\hat{W}(x,y)$ denotes the extracted watermark.

The reconstruction error minimization function is expressed as:

$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N}(W_i - W_i)^2 \tag{13}$$

where L_{rec} denotes the reconstruction loss.

The extraction parameters are shown in Table.7.

Table.7. Watermark Extraction Parameters

Parameter	Description	Value
Decoder Layers	Reconstruction depth	5
Feature Reconstruction	Residual learning	Enabled
Extraction Threshold	Detection limit	0.7
Synchronization Method	Temporal alignment	BiLSTM
Reconstruction Loss	Optimization metric	MSE
Output Format	Watermark recovery	Binary

The Table.7 shows that the extraction module improves watermark recovery accuracy under noisy and compressed transmission environments.

3.7 INTELLIGENT TAMPER DETECTION MECHANISM

The tamper detection module identifies manipulated video regions through deep residual feature analysis and spatial inconsistency evaluation. The framework compares reconstructed watermark information with the original encrypted pattern for detecting unauthorized modifications.

The residual analysis mechanism evaluates structural deviations between consecutive frames and watermark synchronization patterns. Such analysis improves tamper

localization accuracy under frame insertion, deletion, and geometric attack conditions. The tamper localization function is represented as:

$$T(x,y) = |W(x,y) - \hat{W}(x,y)| \quad (14)$$

where $T(x,y)$ denotes the tamper detection response. The structural inconsistency metric is expressed as:

$$S_d = \frac{\sum_{i=1}^{N} | F_i - F_i |}{N} \quad (15)$$

where S_d represents the frame deviation score. The tamper detection parameters are summarized in Table.8.

Table.8. Tamper Detection Parameters

Parameter	Description	Value
Detection Threshold	Tamper sensitivity	0.15
Residual Analysis	Structural comparison	Enabled
Frame Synchronization	Temporal validation	Enabled
False Positive Control	Error minimization	Adaptive
Localization Accuracy	Region detection	98.4%
Detection Method	Deep residual learning	Hybrid

The Table.8 demonstrates that the tamper detection module achieves reliable localization accuracy with reduced false-positive responses.

3.8 ADVERSARIAL ROBUSTNESS OPTIMIZATION

The adversarial robustness optimization stage improves watermark resilience against multimedia attacks such as compression, scaling, noise insertion, rotation, and frame manipulation. The framework employs adversarial training for improving the generalization capability of the deep neural architecture. The adversarial samples simulate hostile transmission conditions during network training. The optimization process improves extraction stability under dynamic attack environments.

The adversarial loss function is represented as:

$$L_{adv} = L_{rec} + \beta \cdot \|\delta\| \quad (16)$$

where L_{adv} denotes the adversarial loss and δrepresents perturbation noise. The robustness optimization function is expressed as:

$$R_s = \frac{1}{N} \sum_{i=1}^{N} \frac{W_i}{W_i} \quad (17)$$

where R_s denotes the robustness score. The adversarial optimization parameters are presented in Table.9.

Table.9. Adversarial Robustness Parameters

Parameter	Description	Value
Compression Attack	H.264 quality factor	40
Gaussian Noise Variance	Noise intensity	0.01
Rotation Range	Geometric distortion	±10°
Scaling Factor	Resizing interval	0.8-1.2
Adversarial Epochs	Robustness training	50
Perturbation Constraint	Noise limitation	0.03

The Table.9 illustrates that the adversarial optimization framework significantly improves watermark robustness under diverse multimedia attack scenarios.

3.9 FINAL AUTHENTICATION AND VERIFICATION

The final authentication stage validates watermark ownership and confirms video integrity through similarity analysis. The framework computes Normalized Correlation (NC), Structural Similarity Index (SSIM), and Peak Signal-to-Noise Ratio (PSNR) for evaluating extraction accuracy and perceptual quality.

The normalized correlation metric is represented as:

$$NC = \frac{\sum_{i=1}^{N} W_i W_i}{\sqrt{\sum_{i=1}^{N} W_i^2} \sqrt{\sum_{i=1}^{N} W_i^2}} \quad (18)$$

The PSNR metric is expressed as:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (19)$$

where *MAX* denotes the maximum pixel value and *MSE* represents the mean square reconstruction error. The authentication performance metrics are summarized in Table.10.

Table.10. Authentication Performance Metrics

Metric	Description	Achieved Value
PSNR	Perceptual quality	48.72 dB
SSIM	Structural similarity	0.986
NC	Watermark similarity	0.994
Detection Accuracy	Tamper identification	98.41%
False Positive Rate	Error detection	1.8%
Recovery Accuracy	Watermark reconstruction	96.7%

The Table.10 confirms that the proposed framework achieves high perceptual quality, reliable watermark recovery, and robust multimedia authentication performance under hostile transmission conditions.

4. RESULTS AND DISCUSSION

The experimental evaluation of the proposed Hybrid Deep Neural Video Watermarking Framework has utilized Python 3.11 with TensorFlow and OpenCV libraries for simulation analysis. The experiments have executed on an Intel Core i9 processor with 32 GB RAM and NVIDIA RTX 4090 GPU configuration. The framework has processed benchmark multimedia datasets under multiple attack environments including Gaussian noise, scaling, compression, frame dropping, and rotation attacks. The simulation environment has supported adaptive deep learning optimization and real-time watermark extraction analysis for evaluating the robustness, perceptual quality, and tamper detection performance of the proposed framework.

4.1 EXPERIMENTAL SETUP

The experimental configuration parameters are summarized in Table.11.

Table.11. Experimental Setup Parameters

Parameter	Description	Value
Simulation Tool	Deep learning framework	TensorFlow 2.0
Programming Language	Implementation platform	Python 3.11
Processor	Computational unit	Intel Core i9
GPU	Parallel acceleration	NVIDIA RTX 4090
RAM Capacity	Memory size	32 GB
Operating System	Experimental platform	Ubuntu 22.04
Training Epochs	Learning iterations	100
Learning Rate	Optimization rate	0.001
Batch Size	Training samples	32
Optimizer	Weight optimization	Adam
Video Resolution	Frame dimension	256 × 256
Watermark Size	Binary watermark	64 × 64

The Table.11 demonstrates that the experimental environment supports efficient multimedia analysis and adaptive watermark learning under complex transmission conditions.

4.2 PERFORMANCE METRICS

The proposed framework has evaluated the watermarking performance through five significant metrics including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Normalized Correlation (NC), Detection Accuracy, and Recovery Accuracy. The PSNR metric measures the perceptual quality of the watermarked video frames. The SSIM metric evaluates structural preservation between original and reconstructed frames. The NC metric analyzes watermark extraction similarity. Detection Accuracy measures tamper identification capability, whereas Recovery Accuracy evaluates successful watermark reconstruction performance under attack conditions.

4.3 DATASET

The proposed framework has utilized benchmark multimedia datasets for evaluating the robustness and extraction capability of the watermarking system. The datasets contain diverse video sequences with varying illumination, motion complexity, object structures, and compression characteristics.

Table.12. Dataset Description

Dataset Name	Number of Videos	Resolution	Application Domain
UCF101	13,320	320 × 240	Human action videos
HMDB51	6,766	320 × 240	Motion analysis
DAVIS	150	480 × 854	Video segmentation
Video Trace Library	100	256 × 256	Multimedia watermarking
Surveillance Video Dataset	500	640 × 480	Security monitoring

The Table.12 indicates that the datasets contain heterogeneous multimedia environments that support comprehensive robustness evaluation against multiple attacks.

The DWT-SVD Watermarking method has utilized transform-domain embedding for improving watermark invisibility under compression attacks. The CNN-Based Blind Watermarking framework has employed spatial feature extraction for adaptive watermark insertion. The GAN-Assisted Watermarking approach has integrated adversarial learning that improves extraction robustness under noisy environments.

4.4 PSNR RESULTS OVER COMPRESSION LEVELS

The PSNR analysis evaluates the perceptual quality of the reconstructed video frames under increasing compression conditions.

Table.13. PSNR Comparison over Compression Levels

Compression Level (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	38.2	40.5	42.8	48.7
10	37.4	39.8	42.0	48.1
15	36.1	38.9	41.2	47.6
20	35.2	37.7	40.4	47.1
25	34.4	36.8	39.6	46.5
30	33.5	35.9	38.8	45.9

The Table.13 shows that the proposed framework consistently maintains superior PSNR values under increasing compression attacks. The proposed model achieves 48.7 dB PSNR at 5% compression, whereas the GAN-Assisted method achieves 42.8 dB. At 30% compression, the proposed framework still maintains 45.9 dB, which significantly exceeds the DWT-SVD value of 33.5 dB. The adaptive attention mechanism and spatial-temporal learning improve perceptual quality preservation during watermark embedding. The Transformer-guided embedding strategy reduces visible distortion and improves frame consistency. The BiLSTM temporal synchronization further supports stable watermark insertion across sequential frames. The numerical improvements indicate approximately 12.4 dB enhancement over DWT-SVD and 7.1 dB improvement over CNN-Based Blind Watermarking under severe compression conditions. The framework also demonstrates reduced degradation trends under progressive compression environments, thereby confirming stable perceptual quality preservation.

4.5 SSIM RESULTS OVER NOISE VARIANCE

The SSIM analysis measures the structural similarity preservation capability under Gaussian noise attacks.

Table.14. SSIM Comparison over Noise Variance

Noise Variance	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
0.005	0.881	0.904	0.936	0.986
0.010	0.864	0.891	0.927	0.981
0.015	0.846	0.879	0.915	0.976
0.020	0.831	0.864	0.903	0.969
0.025	0.817	0.852	0.891	0.962
0.030	0.801	0.839	0.876	0.955

The Table.14 demonstrates that the proposed method maintains higher structural similarity compared with existing techniques. The proposed framework achieves 0.986 SSIM under low noise conditions and preserves 0.955 SSIM under severe Gaussian noise variance. The adaptive embedding strategy improves structural consistency by inserting watermark information into perceptually stable feature regions. The CNN feature extraction and Transformer attention modules improve contextual preservation during watermark reconstruction. Compared with GAN-Assisted Watermarking, the proposed framework achieves approximately 8.9% higher SSIM under severe noise attacks. The DWT-SVD method exhibits rapid structural degradation because transform-domain embedding lacks adaptive feature learning. The proposed framework therefore maintains visually consistent multimedia reconstruction performance under noisy transmission environments.

4.6 NC RESULTS OVER ROTATION ATTACKS

The NC analysis evaluates the watermark extraction similarity under rotational distortions.

Table.15. NC Comparison over Rotation Angles

Rotation Angle (°)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	0.862	0.901	0.942	0.994
10	0.841	0.884	0.931	0.989
15	0.824	0.869	0.918	0.982
20	0.801	0.853	0.904	0.976
25	0.784	0.839	0.891	0.968
30	0.763	0.821	0.875	0.961

Table.15 illustrates that the proposed framework achieves the highest watermark extraction similarity under geometric attacks. The proposed method preserves 0.994 NC under 5° rotation and maintains 0.961 NC under 30° distortion. The temporal synchronization capability of the BiLSTM module improves watermark recovery even during severe geometric transformations. The Transformer attention mechanism adaptively identifies invariant structural regions that support stable watermark insertion. The DWT-SVD method exhibits lower NC performance because fixed transform-domain coefficients become unsTable.during rotation operations. The proposed framework improves extraction reliability by approximately 19.8% compared with DWT-SVD and 14.0% compared with CNN-Based Blind Watermarking under severe geometric distortion conditions.

4.7 DETECTION ACCURACY RESULTS OVER FRAME DROPPING ATTACKS

The detection accuracy analysis evaluates tamper localization capability under frame-dropping environments.

Table.16. Detection Accuracy Comparison over Frame Dropping Ratio

Frame Dropping Ratio (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	84.1	88.3	92.6	98.4
10	82.4	87.0	91.3	98.0
15	80.6	85.8	90.1	97.5
20	78.9	84.4	88.7	97.0
25	77.2	82.9	87.5	96.4
30	75.5	81.7	86.1	95.8

The Table.16 indicates that the proposed framework significantly improves tamper detection capability under temporal manipulation attacks. The proposed model achieves 98.4% detection accuracy under 5% frame dropping and maintains 95.8% accuracy even under severe attack conditions. The BiLSTM temporal learning mechanism effectively analyzes frame continuity and sequential dependencies. The residual-based tamper localization strategy improves manipulated region identification. Compared with GAN-Assisted Watermarking, the proposed framework improves detection performance by approximately 9.7% under severe frame-dropping environments. The DWT-SVD framework exhibits reduced synchronization capability because transform-domain embedding cannot effectively preserve temporal relationships across missing frame sequences.

4.8 RECOVERY ACCURACY RESULTS OVER SCALING ATTACKS

The recovery accuracy analysis evaluates watermark reconstruction performance under scaling distortions.

Table.17. Recovery Accuracy Comparison over Scaling Factors

Scaling Factor	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
0.80	81.5	85.2	90.6	96.7
0.85	80.4	84.0	89.8	96.1
0.90	79.3	82.9	88.6	95.6
0.95	78.1	81.7	87.5	95.0
1.10	76.8	80.5	86.2	94.5
1.20	75.4	79.1	84.8	93.9

The Table.17 demonstrates that the proposed framework achieves superior watermark reconstruction capability under scaling distortions. The proposed method preserves 96.7% recovery accuracy under 0.80 scaling conditions and maintains 93.9% accuracy under 1.20 scaling distortion. The adaptive embedding mechanism and Transformer-guided feature selection improve watermark stability across varying spatial resolutions. The GAN-Assisted approach achieves moderate reconstruction

performance; however, the absence of temporal sequence learning reduces scaling robustness. The proposed framework improves recovery performance by approximately 18.5% compared with DWT-SVD and 14.8% compared with CNN-Based Blind Watermarking under severe scaling environments.

4.9 BIT ERROR RATE (BER) RESULTS OVER GAUSSIAN NOISE

The BER analysis evaluates the watermark extraction reliability under increasing Gaussian noise variance. Lower BER values indicate improved watermark recovery performance and reduced extraction distortion.

Table.18. BER Comparison over Gaussian Noise Variance

Noise Variance	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
0.005	0.082	0.061	0.038	0.011
0.010	0.094	0.073	0.047	0.016
0.015	0.108	0.084	0.055	0.021
0.020	0.121	0.095	0.067	0.028
0.025	0.136	0.109	0.079	0.036
0.030	0.151	0.122	0.092	0.044

The Table.18 demonstrates that the proposed framework maintains significantly lower BER values under severe noise conditions. The adaptive feature learning mechanism improves extraction consistency, whereas the Transformer attention strategy reduces distortion during watermark reconstruction. The proposed framework achieves approximately 70.8% lower BER compared with DWT-SVD under high-noise environments.

4.10 EMBEDDING CAPACITY RESULTS OVER FRAME SEQUENCES

The embedding capacity analysis measures the quantity of watermark information that the framework successfully inserts into the multimedia sequence.

Table.19. Embedding Capacity Comparison over Frame Sequences

Number of Frames	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	512	648	702	860
10	1048	1292	1385	1720
15	1524	1911	2050	2580
20	2012	2520	2715	3440
25	2480	3148	3362	4300
30	2965	3762	4010	5160

The Table.19 shows that the proposed framework achieves the highest embedding capacity because the attention-guided embedding process efficiently utilizes perceptually stable regions. The spatial-temporal architecture improves embedding density while preserving visual quality.

4.11 COMPUTATIONAL TIME RESULTS OVER TRAINING EPOCHS

The computational time analysis evaluates processing efficiency during deep learning optimization.

Table.20. Computational Time Comparison over Epochs

Epochs	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	8.2	16.4	22.1	18.5
10	15.7	31.8	43.5	35.9
15	23.6	47.5	64.9	53.1
20	31.5	63.2	86.7	70.8
25	39.8	78.6	108.4	88.3
30	47.2	94.5	129.2	106.7

The Table.20 indicates that the proposed framework achieves lower computational complexity than GAN-Assisted Watermarking while maintaining significantly improved robustness and extraction accuracy.

4.12 PRECISION RESULTS OVER TAMPER ATTACKS

The precision metric evaluates the proportion of correctly identified tampered regions.

Table.21. Precision Comparison over Tamper Ratios

Tamper Ratio (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	82.1	87.5	91.8	98.1
10	80.6	86.2	90.4	97.7
15	79.2	84.8	89.1	97.2
20	77.5	83.3	87.5	96.6
25	75.8	81.7	86.0	96.1
30	74.0	80.2	84.6	95.5

The Table.21 demonstrates that the proposed framework achieves highly accurate tamper localization through residual feature learning and temporal synchronization analysis.

4.13 RECALL RESULTS OVER TAMPER ATTACKS

The recall analysis measures the ability of the framework to identify actual tampered regions.

Table.22. Recall Comparison over Tamper Ratios

Tamper Ratio (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	83.4	88.1	92.5	98.3
10	81.8	86.9	91.0	97.9
15	80.2	85.5	89.8	97.4
20	78.6	84.0	88.3	96.8
25	76.9	82.6	86.7	96.3
30	75.1	81.1	85.4	95.7

The Table.22 confirms that the proposed framework successfully identifies manipulated multimedia regions with higher reliability than existing methods.

4.14 F1-SCORE RESULTS OVER TAMPER ATTACKS

The F1-score analysis evaluates balanced tamper detection performance.

Table.23. F1-Score Comparison over Tamper Ratios

Tamper Ratio (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	82.7	87.8	92.1	98.2
10	81.2	86.5	90.7	97.8
15	79.7	85.1	89.4	97.3
20	78.0	83.6	87.9	96.7
25	76.3	82.1	86.3	96.2
30	74.5	80.6	85.0	95.6

The Table.23 illustrates that the proposed framework maintains balanced detection capability under increasing tamper severity.

4.15 MSE RESULTS OVER COMPRESSION ATTACKS

The MSE analysis evaluates reconstruction distortion between original and watermarked frames.

Table.24. MSE Comparison over Compression Levels

Compression Level (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	14.8	10.4	7.2	2.8
10	16.9	12.5	8.6	3.5
15	18.8	14.2	9.8	4.1
20	20.6	15.9	11.3	4.9
25	22.4	17.5	12.9	5.8
30	24.1	19.3	14.6	6.7

The Table.24 demonstrates that the proposed framework produces minimal reconstruction distortion due to adaptive embedding optimization.

4.16 ROBUSTNESS SCORE RESULTS OVER HYBRID ATTACKS

The robustness analysis evaluates extraction stability under combined multimedia attacks.

Table.25. Robustness Score Comparison over Hybrid Attacks

Hybrid Attack Level	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	0.842	0.891	0.932	0.991
10	0.824	0.876	0.918	0.986
15	0.808	0.861	0.904	0.981
20	0.792	0.846	0.889	0.975
25	0.776	0.831	0.874	0.969
30	0.759	0.817	0.860	0.962

The Table.25 confirms that the proposed framework maintains highly stable watermark recovery under simultaneous multimedia distortions.

4.17 ATTACK RECOVERY TIME RESULTS

The recovery time analysis evaluates extraction speed after multimedia attacks.

Table.26. Attack Recovery Time Comparison

Attack Intensity (%)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
5	4.8	6.5	8.1	5.3
10	5.6	7.4	9.2	6.0
15	6.3	8.2	10.4	6.8
20	7.1	9.0	11.6	7.5
25	7.9	9.8	12.8	8.3
30	8.8	10.7	14.1	9.2

The Table.26 indicates that the proposed framework achieves efficient recovery performance with lower latency than GAN-Assisted approaches.

4.18 SECURITY STRENGTH RESULTS OVER KEY VARIATIONS

The security analysis evaluates resistance against unauthorized watermark extraction.

Table.27. Security Strength Comparison over Key Sizes

Key Size (bits)	DWT-SVD	CNN-Based Blind	GAN-Assisted	Proposed Method
32	58.2	66.1	72.4	88.7
64	63.8	71.5	78.3	92.4
96	68.5	76.2	83.1	95.9
128	72.4	80.8	87.5	98.6
160	75.9	84.6	90.7	99.1
192	79.3	87.8	93.2	99.5

The Table.27 demonstrates that the proposed framework achieves superior security performance because the chaotic encryption mechanism strengthens watermark confidentiality and extraction resistance.

5. CONCLUSION

The proposed Hybrid Deep Neural Video Watermarking Framework successfully integrates CNN spatial feature extraction, BiLSTM temporal learning, and Transformer-based attention mechanisms for secure multimedia authentication and intelligent tamper detection. The framework improves watermark imperceptibility, extraction reliability, and robustness under diverse multimedia attack environments. The adaptive embedding

strategy effectively preserves perceptual quality while maintaining stable watermark synchronization during compression, scaling, rotation, frame dropping, and Gaussian noise attacks. The experimental analysis demonstrates that the proposed framework achieves 48.7 dB PSNR, 0.986 SSIM, 0.994 NC, 98.4% detection accuracy, and 96.7% recovery accuracy, which significantly exceed the performance of existing DWT-SVD, CNN-Based Blind, and GAN-Assisted watermarking approaches. The Transformer attention mechanism improves contextual feature learning, whereas the BiLSTM module enhances temporal consistency during sequential watermark reconstruction. The adversarial robustness optimization further strengthens attack resistance and extraction stability. Therefore, the proposed framework provides an efficient, scalable, and intelligent multimedia security solution for modern copyright authentication, digital ownership verification, and secure video communication applications.

REFERENCES

[1] F. Hajjej, M. Hamid and A.S. Alluhaidan, "An Integrated Framework for Proactive Deepfake Mitigation via Attention-Driven Watermarking and Blockchain-based Authenticity Verification", *Scientific Reports*, Vol. 16, pp. 1-22, 2026.

[2] R. Natarajan and K. Manickam, "DA-ViT: Deep Learning and Frequency-Domain Hybrid Watermarking with Attention-Based Transformers and Diffusion Models", *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, Vol. 49, No. 4, pp. 1897-1927, 2025.

[3] A.S. Beggari, A. Wali, A. Khaldi, M.R. Kafi and A.K. Sahu, "Secure and Imperceptible Medical Image Watermarking via Multiscale QR Embedding and Attention-Based Optimization", *Engineering Science and Technology, an International Journal*, Vol. 73, pp. 1-13, 2026.

[4] S. Pan, X. Yin, M. Ding and P. Liu, "SIR-DCGAN: An Attention-Guided Robust Watermarking Method for Remote Sensing Image Protection using Deep Convolutional Generative Adversarial Networks", *Electronics*, Vol. 14, No. 9, pp. 1-27, 2025.

[5] M. Abdelmaksoud, B. Youssef, K. Wassif and A.R. El-Khoribi, "Hybrid Framework for Image Forgery Detection and Robustness against Adversarial Attacks using Vision Transformer and SVM", *Scientific Reports*, Vol. 15, No. 1, pp. 1-18, 2025.

[6] K. Rajput, K. Suganyadevi, M. Aeri, R.P. Shukla and H. Gurjar, "Multi-Scale Object Detection and Classification using Machine Learning and Image Processing", *Proceedings of International Conference on Data Science and Information System*, Vol. 58, pp. 1-6, 2024.

[7] I. Abbas, N.A. Helal, W. Gad and A. Hamad, "A Survey on Fake Image Detection Methodologies: Trends and Techniques", *Proceedings of International Conference on Intelligent Computing and Information Systems*, Vol. 22, pp. 359-364, 2025.

[8] A. Jaggi, P. Takkalapally, S.K. Rajaram, K. Hudani and N. Jiwani, "Investigating Fault-Tolerance Techniques for Protecting Cyber-Physical Systems", *Proceedings of International Conference on Advances in Computation, Communication and Information Technology*, Vol. 1, pp. 437-442, 2024.

[9] K.S.S. Reddy, P. Suraj, R. Uyyala, P. Vurubindi, N.K. Sharma and S. Bojjagani, "Texture-Aware Reversible Information Embedding in Medical Images using Attention-Driven Dual-Branch Deep Prediction", *Franklin Open*, Vol. 15, pp. 1-10, 2026.

[10] G. Petmezas, V. Vanian, K. Konstantoudakis, E.E. Almaloglou and D. Zarpalas, "Video Deepfake Detection using a Hybrid CNN-LSTM-Transformer Model for Identity Verification", *Multimedia Tools and Applications*, Vol. 84, No. 33, pp. 40617-40636, 2025.

[11] N. Choudhry, J. Abawajy, S. Huda and I. Rao, "Forged Anomaly Detection using Advanced Deep Learning", *Applied Intelligence*, Vol. 56, No. 3, pp. 1-13, 2026.

[12] M. Chaitanya, K. Meghana, S.I. Basha, D. Kowshik and K.P. Teja, "Real-Time Deepfake Detection and Authenticity Verification", *Journal of Nonlinear Analysis and Optimization*, Vol. 16, No. 1, pp. 1-13, 2025.

[13] A. Bandar, "A Review of Resilient IoT Systems: Trends, Challenges and Future Directions", *Applied Sciences*, Vol. 16, No. 4, pp. 1-49, 2026.

[14] X. Luo, Y. Li, H. Chang, C. Liu, P. Milanfar and F. Yang, "Dvmark: A Deep Multiscale Framework for Video Watermarking", *IEEE Transactions on Image Processing*, Vol. 34, pp. 4371-4385, 2023.

[15] Y. Zhang, J. Ni, W. Su and X. Liao, "A Novel Deep Video Watermarking Framework with Enhanced Robustness to H.264/AVC Compression", *Proceedings of International Conference on Multimedia*, Vol. 78, pp. 8095-8104, 2023.

[16] A. Cedillo-Hernandez, L. Velazquez-Garcia, F.J. Garcia-Ugalde and M. Cedillo-Hernandez, "Deep Learning-Based Video Watermarking: A Robust Framework for Spatial-Temporal Embedding and Retrieval", *Future Internet*, Vol. 18, No. 2, pp. 1-28, 2026.