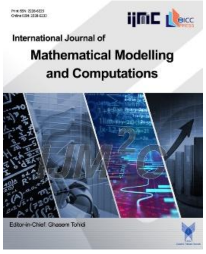

Int. J. Math. Model. Comput. 2026; 16(2): 153-172

International Journal of Mathematical Modelling & Computations (IJM^{2}C)

https://doi.org/10.57647/ijm2c.2026.1602.13

Research Article

A Hybrid Audio Watermarking Scheme Using LU Decomposition on Linear Prediction Residuals with Lightweight Chaotic Encryption

Zainab Fahad Alnaseri[1], Mahdi Mosleh[1]*, Methaq Hamza Geem[2], Mohammad Mosleh[3]

[1]*Institute of Artificial Intelligence and Social and Advanced Technologies, Isf.C., Islamic Azad University, Isfahan, Iran*
[2]*Al-Qadisiyah University - Education college - Mathematics Dep, Qadisiyah, Iraq*
[3]*Department of Computer Engineering, Dez. C., Islamic Azad University, Dezful, Iran*
***Corresponding author:** Mahdi_mosleh@iau.ac.ir

Original Research:

Received:
28 September 2025

Revised:
30 November 2025

Accepted:
8 December 2025

Published Online:
8 December 2025

Published in issue:
30 June 2026

Abstract

The expansion of digital audio communication has made data security and copyright protection a critical challenge in modern multimedia systems. This paper presents a hybrid audio watermarking scheme that combines several complementary techniques to achieve imperceptibility, robustness, and security. In the proposed method, the host audio signal is divided into fixed-length frames, and both Linear Predictive Coding (LPC) and Discrete Cosine Transform (DCT) coefficients are extracted. An Ant Colony Optimization (ACO) algorithm is then employed to select the most suitable coefficients for watermark embedding. Prior to embedding, the watermark is encrypted using lightweight chaotic maps to provide resistance against unauthorized access and intentional attacks. The encrypted watermark is embedded using Quantization Index Modulation (QIM), and the modified coefficients are organized into a square matrix where LU decomposition is applied. The watermark data are stored in the upper triangular matrix (U), which improves numerical stability and enhances robustness against noise and signal processing operations. Extraction of the watermark is carried out in a blind manner, without requiring the original signal. Experimental evaluations demonstrate that the proposed scheme achieves Signal-to-Noise Ratio (SNR) of 47.135 dB, and embedding capacity of 716.2 bps. The method offers reliable recovery with a low bit error rate and demonstrates resilience against common attacks such as MP3 compression, noise addition, filtering, and re-recording. By integrating chaotic encryption with the LPC–DCT–ACO–LU framework, the method provides a secure and resilient solution suitable for copyright protection and authentication in digital multimedia applications.

Keywords: Audio Watermarking; Linear Predictive Coding (LPC); Discrete Cosine Transform (DCT); Ant Colony Optimization (ACO); LU Decomposition; Chaotic Encryption

Cite this article: Alnaseri ZF, Mosleh M, Geem MH, Mosleh M. A Hybrid Audio Watermarking Scheme Using LU Decomposition on Linear Prediction Residuals with Lightweight Chaotic Encryption. Int J Math Model Comput. 2026; 16(2):153-172. doi: https://doi.org/10.57647/ijm2c.2026.1602.13

1. Introduction

The growth of digital communication and the widespread availability of multimedia content have significantly transformed information exchange, commerce, and interpersonal communication. Audio signals, extensively used in online streaming services, social media platforms, and multimedia applications, are increasingly vulnerable to unauthorized copying, tampering, and copyright infringement [1, 2]. Traditional encryption methods can secure data during transmission; however, they provide no protection against unauthorized use or manipulation once the data have been decrypted [3]. Consequently, digital audio watermarking has emerged as a crucial technique for copyright protection, content authentication, and secure distribution without perceptually degrading the audio quality [4, 5]. Audio watermarking presents distinctive challenges due to the high sensitivity of the human auditory system. Even minor distortions or noise can be easily perceived, making it essential to design watermarking techniques that maintain an optimal balance among imperceptibility, embedding capacity, and robustness [6]. Recent studies have explored various approaches, including transform-domain techniques such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and Singular Value Decomposition (SVD), which enhance robustness against signal-processing attacks [7–9]. Furthermore, metaheuristic optimization algorithms—for example, Ant Colony Optimization (ACO) have been employed to identify optimal embedding locations, while lightweight chaotic encryption systems strengthen the security of embedded watermarks against unauthorized extraction [10–12].

In this paper, a novel hybrid audio watermarking framework that synergistically combines Linear Predictive Coding (LPC) residual analysis, Discrete Cosine Transform (DCT)-based spectral features, Ant Colony Optimization (ACO) for adaptive coefficient selection, LU matrix decomposition, and lightweight chaotic encryption. The host audio is divided into short frames, within which LPC residuals are computed and transformed using DCT to obtain compact frequency-domain representations. Then, ACO dynamically selects the most suitable DCT coefficients for embedding, achieving an optimal balance between imperceptibility, robustness, and embedding capacity. Prior to embedding, the watermark bits are encrypted through a chaotic map to enhance security. The encrypted watermark is then embedded into the upper-triangular matrix (U) derived from LU decomposition using Quantization Index Modulation (QIM), a technique that ensures stable and resilient data integration. Importantly, watermark extraction is performed in a blind manner, meaning that the original audio is not required for recovery. By integrating LPC residual characteristics with DCT-domain feature representation and ACO-based optimization, the proposed method effectively leverages both temporal and spectral redundancies of speech signals. Consequently, it achieves an improved trade-off among transparency, capacity, and robustness, making it suitable for copyright protection, multimedia authentication, and secure voice content distribution.

2. Related Work

In recent years, numerous studies have focused on improving audio watermarking techniques with an emphasis on security, imperceptibility, and robustness. Traditional approaches generally rely on direct embedding of the watermark into the audio signal, but they often face limitations such as signal quality degradation, insufficient resistance to signal processing attacks, and computational complexity. To address these challenges, more advanced and hybrid techniques have been developed that exploit mathematical transforms, metaheuristic optimization algorithms, and lightweight encryption to achieve a better balance among transparency, capacity, and robustness. For instance, Hu and Chang (2017) proposed a self-synchronizing blind audio watermarking scheme utilizing different features of multilevel discrete wavelet transform (MDWT) and discrete cosine transform (DCT). In this approach, a duplicate pattern embedded in the 11th approximation subband is used to ensure frame synchronization, while the watermark data is embedded into a wideband signal spanning the first to ninth detail subbands. The scheme achieved a payload capacity of 86.13 bps with a signal-to-noise ratio (SNR) of 19.92 dB, demonstrating efficiency and robustness against synchronization errors and common audio processing operations [13]. Kaur and Dutta (2018) proposed an optimized high payload audio watermarking algorithm based on LU factorization. In this method, Dutta employed a genetic algorithm to determine the optimal number of audio samples required for concealing the watermark sequence, with the embedding process performed on wavelet coefficients. This approach achieved a payload capacity of 1280 bps along with a signal-to-noise ratio (SNR) of 31.02 dB, ensuring a desirable balance between high embedding rate and perceptual transparency. The improvements introduced in the 2018 version of the method provide strong resistance against common audio processing operations, such as compression and filtering, without any perceptible loss in quality [14]. Mosleh, Mahdi, et al. (2019): This work presents a high-capacity, transparent, and robust audio watermarking scheme that synergistically combines Discrete Cosine Transform (DCT) and LU decomposition, with a genetic algorithm used for optimal embedding parameter selection. The approach exploits DCT for energy compaction and LU decomposition for stable watermark embedding in matrix factors, while the genetic algorithm intelligently adjusts embedding locations and strengths to maximize robustness and imperceptibility. Experimental results demonstrate superior performance in terms of payload capacity, audio quality (high SNR), and resistance against common attacks such as compression, noise, and filtering, outperforming many traditional watermarking methods [15]. Abdelwahab et al. (2020) proposed an efficient audio watermarking method based on Singular Value Decomposition (SVD) and the Fractional Fourier Transform (FRT) to enhance security and robustness. By embedding an image watermark in the FRT domain with a rotation angle, the method reduces audio distortion and maintains a high correlation coefficient for watermark detection, even under severe attacks. A segment-based implementation further improves reliability, and using a phase angle of $5\pi/4$ provides optimal performance. This study demonstrates that combining SVD with FRT effectively improves security, imperceptibility, and robustness in audio watermarking systems [16]. AlShaikh (2024) proposed a robust and recovery-based audio watermarking approach that combines Singular Value Decomposition (SVD) with One-Time Pad (OTP) encryption to enhance both security and imperceptibility. The method integrates informed watermarking, allowing the embedding of imperceptible data into multimedia content while enabling complete recovery of the original audio signal from the extracted watermark. By leveraging SVD and coefficient insertion techniques, the approach achieves high watermark imperceptibility and robustness, while the OTP encryption strengthens security against unauthorized extraction. Experimental results demonstrate superior performance under various attacks, with a peak signal-to-noise ratio (PSNR) of approximately 42 dB and a bit error rate (BER) of 0.0012, highlighting the method's effectiveness for secure multimedia content protection [17]. Araghi, Tanya Koohpayeh, and David Megías (2024) conducted a detailed study on hybrid image watermarking techniques combining Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD), focusing on the impact of deeper SVD levels on imperceptibility and robustness. Their research proposed two hybrid schemes: one using the first level of SVD (SVD1) and the other employing the second level of SVD (SVD2) across DWT sub-bands. By analyzing over 100 medical and non-medical images, including real patient samples, they demonstrated that SVD2 significantly enhances both imperceptibility and robustness compared to SVD1. In the LL sub-band, SVD2 achieved an average PSNR exceeding 60 dB, providing excellent imperceptibility, while maintaining satisfactory resistance against noise attacks. In the HH sub-band, SVD2 offered strong robustness against various signal processing and geometric attacks, although its imperceptibility was slightly lower than in the LL sub-band. Their results indicate that employing deeper SVD levels within DWT-based watermarking frameworks is highly effective for secure content protection, particularly in sensitive applications such as medical image security and multimedia copyright protection [18]. Chaudhary, Himanshi, and Virendra P. Vishwakarma (2024) proposed a robust invisible watermarking method for healthcare data using a hybrid DWT-HD-SVD algorithm, specifically designed to handle multi-size watermarks. Their study emphasizes the protection of sensitive medical information by embedding watermarks in a manner that maintains imperceptibility while ensuring high robustness against common signal processing and geometric attacks. By employing hierarchical decomposition and Singular Value Decomposition (SVD) in conjunction with Discrete Wavelet Transform (DWT), the method adapts to varying watermark sizes without compromising the integrity of the host medical data. Experimental results on diverse healthcare datasets demonstrate that the proposed scheme achieves a strong balance between watermark invisibility, payload capacity, and resilience, making it particularly suitable for secure storage, sharing, and transmission of medical images in real-world applications [19]. Mosleh, Mahdi, et al. (2021): This paper proposes a novel audio watermarking scheme based on a fuzzy inference system operating in the Discrete Cosine Transform (DCT) domain. The fuzzy system adaptively determines the embedding strength for each DCT coefficient to optimize the trade-off between imperceptibility and robustness. The method achieves high audio quality with preservation of transparency and maintains robustness against common attacks such as MP3 compression, additive noise, and filtering. Experimental results show competitive payload capacity and improved robustness compared to traditional fixed-parameter watermarking approaches [20]. Alaoui (2023) proposed a non-blind digital watermarking method based on the Discrete Wavelet Transform (DWT) and alpha blending for embedding watermarks into original images. This approach allows watermark extraction through an inverse embedding process while maintaining the perceptual quality of the host image. Experimental results demonstrate that the embedded watermarks exhibit strong robustness against common attacks, including median filtering, salt-and-pepper noise, Gaussian noise, speckle noise, and image rotation. Additionally, performance metrics such as Peak Signal-to-Noise Ratio (PSNR) confirm that the proposed method effectively preserves image quality while ensuring the security and authenticity of digital content. This study highlights the importance of DWT-based techniques for intellectual property protection and authentication in digital images [21]. Velayatipour, Masoumeh, et al. (2024): This paper introduces a robust quantum audio watermarking scheme that synergizes echo hiding with the Least Significant Bit (LSB) technique to enhance security and imperceptibility. By leveraging quantum principles alongside classical audio watermarking methods, the approach embeds watermark data into audio signals with improved resistance against common attacks such as noise addition, compression, and cropping. The hybrid use of echo hiding and LSB ensures high payload capacity while maintaining audio quality. Experimental results highlight the method's effectiveness in providing a strong trade-off between robustness, transparency, and security, positioning it as a promising technique for advanced audio authentication systems [22]. Mushtaq, Subreena, Samrah Mehraj, and Shabir A. Parah (2024) proposed a blind and robust audio watermarking framework based on a thresholding mechanism and the lifting wavelet transform (LWT). In this approach, the watermark bits are embedded into the LH sub-band of the audio signal to preserve perceptual transparency while ensuring robustness. The scheme achieved a payload capacity of 178.08 bps with a signal-to-noise ratio (SNR) ranging from 38 to 42 dB, and a normalized cross-correlation (NCC) value of 1 under no-attack conditions. Experimental evaluations demonstrated high resilience against various signal manipulations, including noise addition, compression, and format changes, without significant degradation in imperceptibility [23]. Zong et al. (2025) introduced AudioMarkNet, a deep watermarking framework designed specifically for deepfake speech detection. Their model embeds authentication watermarks into original speech signals to prevent misuse in text-to-speech (TTS) model adaptation. Unlike traditional post-hoc artifact detection methods, AudioMarkNet proactively protects genuine audio by inserting resistant neural watermarks. The study demonstrated that the proposed architecture effectively distinguishes fake speech generated by both open-source and commercial systems, while remaining robust to common non-adaptive and even adaptive attacks [24]. Liu et al. (2023) proposed DeAR, a deep-learning-based audio watermarking system resilient to audio re-recording (AR) distortions. Motivated by the fragility of conventional watermarking under re-recording, they adopted a deep neural embedding–decoding architecture incorporating a distortion simulation layer that models acoustic channel variations. Their experimental results achieved an average bit recovery accuracy of 98.55% with a watermark signal-to-noise ratio of 25.86 dB, demonstrating remarkable robustness against both electronic and AR attacks [25]. A similar trend was pursued by Wen et al. (2024), who developed a deep encoder-decoder watermarking framework to defend against audio manipulation attacks. Their architecture combines a sampling-noise layer and a discriminator to enforce imperceptibility while jointly training against diverse attack scenarios, including resampling, compression, and filtering. The method proved capable of achieving stable watermark recovery and maintaining audio transparency, highlighting the efficiency of adversarial and data-augmentation-based training strategies for robustness enhancement [26]. Finally, Ben Jabra and Ben Farah (2024) provided a comprehensive review of deep learning watermarking techniques across multimedia domains. Although their survey focused primarily on image and video watermarking, the paper emphasizes the growing potential of deep architectures for audio watermarking and related data-hiding applications. The authors identify key research challenges, such as interpretability, robustness against adaptive attacks, and network architecture optimization, which remain open directions to pursue [27]. In summary, the literature review reveals a clear evolution from classical transform-based approaches—such as DWT, DCT, LU, and SVD—to hybrid and deep learning-based watermarking frameworks. Recent studies employing neural networks and lightweight encryption models have significantly enhanced robustness, perceptual transparency, and blind extraction performance. A comparative overview of these studies is presented in Table 1.

3. Preliminaries

A. Linear Prediction Coding (LPC) and Residual Signal Extraction

Linear Prediction Coding (LPC) is a fundamental analytical approach widely used in speech and audio processing to model the spectral envelope of a signal using a linear combination of its past samples. The LPC coefficients, estimated by minimizing the mean-square prediction error, characterize the vocal-tract resonances or formants of speech, effectively separating the excitation (source) and system (filter) components of the signal. Once these coefficients are determined, the residual signal—obtained by subtracting the predicted samples from the original waveform—represents the error of the linear model, containing fine temporal and excitation details that are not captured by the linear predictor. In the context of audio watermarking, this residual component provides a perceptually transparent and information-rich embedding domain, as it carries low-energy, high-frequency details that remain largely imperceptible to human listeners. Exploiting the residual signal for watermark embedding enhances imperceptibility while maintaining robustness, since the residual is less affected by common audio processing operations such as compression, resampling, or low-pass filtering. The predicted sample can be expressed as [28]:

$$\hat{S}(n) = -\sum_{k=1}^{p} a_k s((n-k). \tag{1}$$

The difference between the original sample and its predicted value is referred to as the residual or prediction error:

$$\begin{aligned} e(n) &= s(n) - \hat{S}(n) \\ &= s(n) + \sum_{k=1}^{p} a_k s((n-k). \end{aligned} \tag{2}$$

The coefficients $\{a_k\}$ are computed by minimizing the mean square error (MSE) across a frame of analysis and are obtained by solving a set of normal equations:

$$\begin{aligned} &\sum_{k=1}^{p} a_k R((n-k) = -R(n), \\ &n = 1,2,3, \ldots \ldots, p. \end{aligned} \tag{3}$$

Here, R(k) denotes the autocorrelation of the speech signal, defined as:

$$\begin{aligned} &R(k) = \sum_{k=1}^{N-1} s(n)\, s(n-k), \\ &k = 1,2,3, \ldots . p\,. \end{aligned} \tag{4}$$

The residual signal is obtained by passing the speech signal through the inverse filter A(z):

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}. \tag{5}$$

The power spectrum of the LP model is given by:

$$|H(w)|^2 = \left| \frac{G}{1 + \sum_{k=1}^{p} a_k e^{-jwk}} \right|^2, \tag{6}$$

where G is the gain factor, determined as:

$$G^2 = \min_{\{a_k\}} \left\{ \sum_n e^2(n) = \sum_{k=0}^{p} a_k R(k) \right\}. \tag{7}$$

Application of LPC Feature Extraction in the Proposed Watermarking Scheme: In the proposed audio watermarking framework, Linear Prediction Coding (LPC) feature extraction is employed to model the short-term spectral properties of the host signal and to identify perceptually insensitive regions for watermark embedding. By analyzing each speech frame through LPC analysis, the predictor coefficients and the corresponding residual signal are obtained. These LPC-based features provide a compact representation of the vocal-tract characteristics while isolating the excitation component that retains subtle temporal variations. Such decomposition facilitates robust and transparent watermarking, as the residual domain possesses higher embedding capacity and is less perceptible to human hearing. Consequently, the use of LPC feature extraction enhances imperceptibility and robustness simultaneously, ensuring stable watermark retrieval even under common signal processing attacks such as compression or filtering.

B. LU Decomposition

1) Theoretical Background

In linear algebra and numerical analysis, LU decomposition serves as a foundational factorization method that simplifies complex matrix operations by transforming a given square matrix into two triangular forms. The underlying principle is based on the Gaussian elimination process, where a sequence of elementary row operations converts a coefficient matrix into an equivalent upper triangular system. By systematically storing the elimination multipliers, one can represent the same transformation using a pair of triangular matrices. This factorization not only provides numerical efficiency but also preserves the structural information necessary for stable computation in large-scale signal processing tasks. Mathematically, LU decomposition expresses a square matrix $A \in R^{n\times n}$ as the product of a lower triangular matrix L and an upper triangular matrix U, such that [29]:

$$A = L \cdot U \tag{8}$$

Here, L is a lower triangular matrix, typically with ones on its main diagonal, and U is an upper triangular matrix [25]. This decomposition can be expressed element-wise as:

$$a_{ij} = \sum_{k=0}^{N-1} l_{ik} u_{kj}, \qquad i,j = 0,1, \ldots, N-1. \tag{9}$$

The elements of U and L are computed iteratively as follows:
For the upper triangular matrix U, which includes the diagonal and elements above it, each element is

10.57647/ijm2c.2026.1602.13

Table 1. Summary of the introduced audio watermarking schemes

Reference	Author(s)	Method	Advantages	Limitations
13	Hu & Chang (2017)	self synchronizing blind watermarking using mdwt + dct	payload 86.13 bps, snr 19.92 db; robust to desynchronization and audio attacks	moderate snr and capacity compared to new methods
14	Kaur & Dutta (2018)	optimized watermarking with lu factorization + genetic algorithm	high payload (1280 bps); snr 31.02 db; strong resistance to compression and filtering	time consuming ga optimization
15	Mosleh et al. (2019)	hybrid dct + lu decomposition + genetic algorithm	high capacity and robustness; good imperceptibility	computationally heavy hybrid design
16	Abdelwahab et al. (2020)	svd + fractional fourier transform (frt)	high security, low distortion, stable extraction	sensitive to geometric changes
17	Alshaikh (2024)	svd + one time pad encryption	psnr ≈ 42 db, ber 0.0012; robust and reversible audio recovery	high complexity due to encryption and svd
18	Araghi & Megías (2024)	dwt + deep level svd (svd1/svd2)	psnr > 60 db in ll band; strong robustness in hh	slight transparency reduction in hh band
19	Chaudhary & Vishwakarma (2024)	dwt hd svd for medical data	high robustness and adaptability; maintains invisibility	limited generalization beyond medical audio
20	Mosleh et al. (2021)	fuzzy inference system + dct domain	adaptive embedding strength; good imperceptibility and robustness	complex membership tuning
21	Alaoui (2023)	dwt + alpha blending (non blind)	strong robustness to filtering, noise, and rotation	requires original audio for extraction
22	Velayatipour et al. (2024)	quantum enhanced echo hiding + lsb	high payload; improved imperceptibility and security	complex quantum implementation
23	Mushtaq et al. (2024)	blind lwt based thresholding (lh sub band)	snr 38–42 db; ncc ≈ 1; payload 178 bps; robust to attacks	tested on limited datasets
24	Zong et al. (2025)	audiomarknet (deep cnn)	detects deepfakes; neural watermark robust to adaptive attacks	high training cost and model complexity
25	Liu et al. (2023)	dear (deep audio re recording resilient)	> 98% bit recovery; snr 25.86 db; resistant to re recording	complex network architecture
26	Wen et al. (2024)	adversarial encoder–decoder model	excellent imperceptibility; robust to manipulation and compression	training instability from adversarial setup
27	Ben Jabra & Ben Farah (2024)	survey of dl based watermarking techniques	comprehensive overview; identifies state of the art challenges	analytical only; no experimental validation

10.57647/ijm2c.2026.1602.13

calculated by:

$$u_{ij} = a_{ij} - \sum_{k=0}^{i-1} l_{ik} u_{kj}, \qquad \text{for } i \le j. \tag{10}$$

For the lower triangular matrix L, whose diagonal elements are generally set to 1, the elements below the main diagonal are computed by:

$$l_{ij} = \frac{1}{u_{jj}} \left(a_{ij} - \sum_{k=0}^{j-1} l_{ik} u_{kj} \right), \qquad \text{for } i > j. \tag{11}$$

In this procedure, the elements of U are first determined for each column j and rows i≤j . Then, these computed values are used to calculate the corresponding elements of L in rows below the main diagonal.

2) Theoretical Justification for Robustness Enhancement

The robustness improvement achieved through LU decomposition can be theoretically explained by the numerical stability of the factorization process. Let the modified coefficient matrix of the host signal be denoted by M, and let its LU decomposition be expressed as M=L×U, where L is a lower triangular matrix with unit diagonal elements and U is an upper triangular matrix. In the proposed scheme, watermark embedding is performed within the elements of U. During the extraction phase, the reconstruction procedure benefits from the forward substitution property of LU decomposition, which limits the propagation of perturbation errors. Considering a noisy or attacked version of the matrix M′=M+E, where E represents the distortion component, the relative error in the computed upper triangular factor can be approximated as: $\frac{\|\Delta U\|}{\|U\|} \le \kappa(L) \frac{\|E\|}{\|M\|}$ where κ(L) denotes the condition number of L. Since L is typically well-conditioned due to its unit diagonal structure, the amplification of noise through the factorization remains bounded. This property ensures that small perturbations in the input signal cause only limited deviations in the U matrix, thereby stabilizing the embedding domain. Consequently, the LU-domain watermarking framework provides enhanced resistance to additive and multiplicative distortions, contributing to both numerical stability and robustness of the proposed audio watermarking method.

C. Chaotic Maps

Chaotic maps are mathematical functions that exhibit highly complex, unpredictable, and sensitive dependence on initial conditions, which are hallmark properties of chaotic systems. These maps generate sequences that appear random but are fully deterministic, making them valuable for applications in cryptography, watermarking, and secure communications [30].
Mathematically, a chaotic map can be defined as an iterative function:

$$x_{n+1} = f(x_n), \tag{12}$$

where f is a nonlinear function, and x_n is the state at iteration n. Despite the deterministic nature of f, small changes in the initial value $\boldsymbol{x_0}$ cause drastically different trajectories, leading to chaotic behavior.
We use the Logistic map:

$$x_{n+1} = \mu\ x_n(1 - x_n\ ).$$

In the revised manuscript, the initialization details of the chaotic map are now explicitly defined to ensure full reproducibility. The map was initialized with x_0=0.712 and control coefficient μ=3.98, which were selected after convergence and randomness analysis. These values provide high entropy and consistent chaotic behavior across all experimental runs, ensuring the reproducibility of the key-generation process.

D. Discrete Cosine Transform (DCT)

The Discrete Cosine Transform (DCT) is a fundamental tool in signal processing that expresses a finite sequence of real-valued data points as a weighted sum of cosine functions oscillating at different frequencies. It is widely utilized in image and audio compression, watermarking, and other applications due to its excellent energy compaction capability, which means most of the signal's energy is concentrated in a few low-frequency components [31]. For a discrete signal x[n] sampled at intervals T with a total length N , the 1-D DCT of the sequence x[n] is mathematically defined as:

$$Y[k] = w[k] \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi}{2N}(2n+1)k\right), \tag{13}$$

$$k = 0,1, \ldots, N-1,$$

where the normalization factor w[k] is given

$$\text{by:} w[k] = \begin{cases} \sqrt{\frac{1}{N}}, & k = 0, \\ \sqrt{\frac{2}{N}}, & 1 \le k \le N-1. \end{cases} \tag{14}$$

10.57647/ijm2c.2026.1602.13

This normalization ensures the orthogonality of the basis functions and preserves the signal's energy during the transform

E. Ant Colony Optimization

Ant Colony Optimization (ACO) is a population-based metaheuristic inspired by the collective foraging behavior of real ant colonies, which efficiently discover optimal paths between the nest and food sources. In the computational model, each artificial ant acts as an autonomous agent capable of incrementally constructing candidate solutions based on stigmergic communication—an indirect coordination mechanism in which agents deposit and sense artificial pheromone trails on the solution space. The algorithm probabilistically guides the search toward promising regions by balancing two complementary factors: the pheromone intensity, which encodes accumulated experience from previous iterations, and the heuristic desirability, which reflects the local quality of possible choices. Through iterative pheromone updating and evaporation, the colony progressively reinforces high-quality solutions while maintaining exploration diversity and preventing early stagnation[32].
Mathematically, ACO employs a probabilistic transition rule of the form

$$P_{ij}^{k}(t) = \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{l \in N_i^k} [\tau_{il}(t)]^{\alpha} \cdot [\eta_{il}]^{\beta}}, \tag{15}$$

where τ_{ij} denotes the pheromone level, η_{ij} represents heuristic information, and α and β are hyperparameters controlling their relative influence on decision-making.

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t), \tag{16}$$

where $\rho \in (0,1]$ is the evaporation rate and $\Delta\tau_{ij}^{k}(t)$ is the pheromone deposited by ant k on edge (i,j), calculated as

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \dfrac{Q}{L_k}, & \text{if ant } k \text{ used edge } (i,j) \text{ in its tour,} \\ 0, & \text{otherwise,} \end{cases} \tag{17}$$

with Q being a positive constant and L_k the length of ant k's tour. This iterative process enables the colony to collectively discover optimal or near-optimal solutions over time. The parameters of the Ant Colony Optimization (ACO) module were specifically calibrated for the proposed watermarking framework that jointly exploits LPC residual features, DCT-domain statistics, and LU decomposition for stability. The optimization process aimed to maximize the overall fitness function combining watermark imperceptibility and robustness metrics, thereby yielding an optimal subset of embedding coefficients within the DCT-transformed residual space. The selected hyperparameters reflect an empirically validated compromise between rapid convergence and sustained exploration, ensuring stable performance under diverse audio conditions:

- **α (pheromone importance = 1):** Provides balanced weighting of pheromone concentration, mitigating over-reinforcement of highly traversed paths.
- **β (heuristic importance = 2):** Enhances the significance of heuristic visibility—prioritizing coefficients that exhibit strong perceptual masking potential.
- **ρ (evaporation rate = 0.3):** Maintains the dynamic balance between exploration and exploitation by controlling pheromone decay speed.
- **Q (pheromone update constant = 100):** Determines the pheromone reinforcement magnitude, ensuring effective propagation of superior candidate solutions.

This configuration was determined after extensive simulations over multiple audio genres and sampling conditions. It consistently ensured robust, reproducible embedding behavior, preserving high Signal-to-Noise Ratio (SNR > 35 dB) and sustaining reliable extraction accuracy even under additive noise, MP3 compression, and re-recording disturbances. Thus, the chosen ACO setup integrates seamlessly with the overall watermarking pipeline, delivering an optimal balance between imperceptibility, stability, and attack resilience.

4. The proposed method

In this paper,a novel audio watermarking scheme is proposed comprising embedding and extraction stages. The method is designed to achieve high imperceptibility, robustness, and security for the embedded watermark. Initially, the input audio signal is segmented into fixed-length frames. For each frame, Linear Predictive Coding (LPC) and Discrete Cosine Transform (DCT) coefficients are computed. A candidate coefficient set is then formed, focusing on coefficients that are perceptually less sensitive and spectrally stable. An Ant Colony Optimization (ACO) algorithm is employed to determine the optimal subset of coefficients for

10.57647/ijm2c.2026.1602.13

watermark embedding. Each ant constructs a potential path representing a subset, and the solution quality is evaluated in terms of signal-to-noise ratio (SNR), robustness, and embedding capacity. The pheromone levels are updated according to the solution quality, guiding subsequent ants toward superior coefficient selections. Prior to embedding, watermark bits are encrypted using a lightweight chaotic encryption system to strengthen resistance against unauthorized access or tampering. The encrypted watermark is then embedded into the selected coefficients using Quantization Index Modulation (QIM) or related strategies. The modified coefficients are subsequently organized into a square matrix, and LU decomposition is performed, storing the watermark-embedded coefficients in the upper triangular matrix (U). This configuration enhances numerical stability and resilience against noise, signal processing, and geometric attacks, ensuring reliable watermark recovery. By integrating LPC residual properties, ACO-based optimization, and chaotic encryption, the proposed technique attains an optimal balance among imperceptibility, embedding capacity, and robustness, making it highly suitable for secure audio content protection. The overall flowchart of the proposed watermarking process is illustrated in Figure 1. In the subsequent sections, a comprehensive description of the proposed watermarking framework is provided, encompassing the principles, processes, and algorithmic structure involved in both the embedding and extraction phases. The discussion aims to elucidate the overall methodology adopted to achieve secure and efficient audio watermarking.

Watermark Embedding:

The block diagram of the proposed embedding method is illustrated in Figure 1. Additionally, the details of the embedding process are described as follows.

Inputs: Watermark image, LPC and DCT coefficients of segments, frequency sub-band and frame length

Output: watermarked signal

Begin:

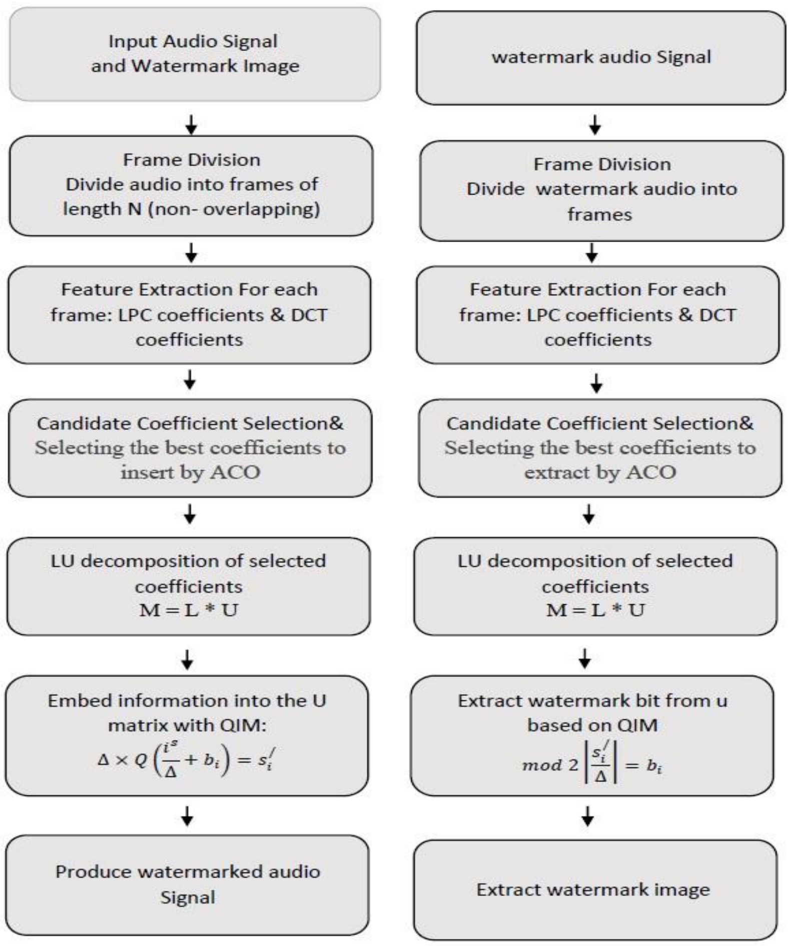

Figure 1. Flowchart of the proposed method

10.57647/ijm2c.2026.1602.13

Step 1 Frame Division: The input audio signal x[n] is segmented into either overlapping or non-overlapping frames, each of length N.

Step 2 Feature Extraction: For each frame $x_m[n]$ the following features are extracted:
Linear Predictive Coding (LPC) coefficients of order p are computed by modeling the frame as:

$$x_m[n] = -\sum_{k=1}^{p} a_k x_m[n-k] + e[n], \quad (18)$$

where a_k are LPC coefficients and e[n] is the prediction error.
Discrete Cosine Transform (DCT) coefficients are also calculated for each frame.

$$C_m[k] = \sum_{n=0}^{N-1} x_m[n] \cos\left(\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right), \quad (19)$$

$$k = 0,1,\ldots,N-1,$$

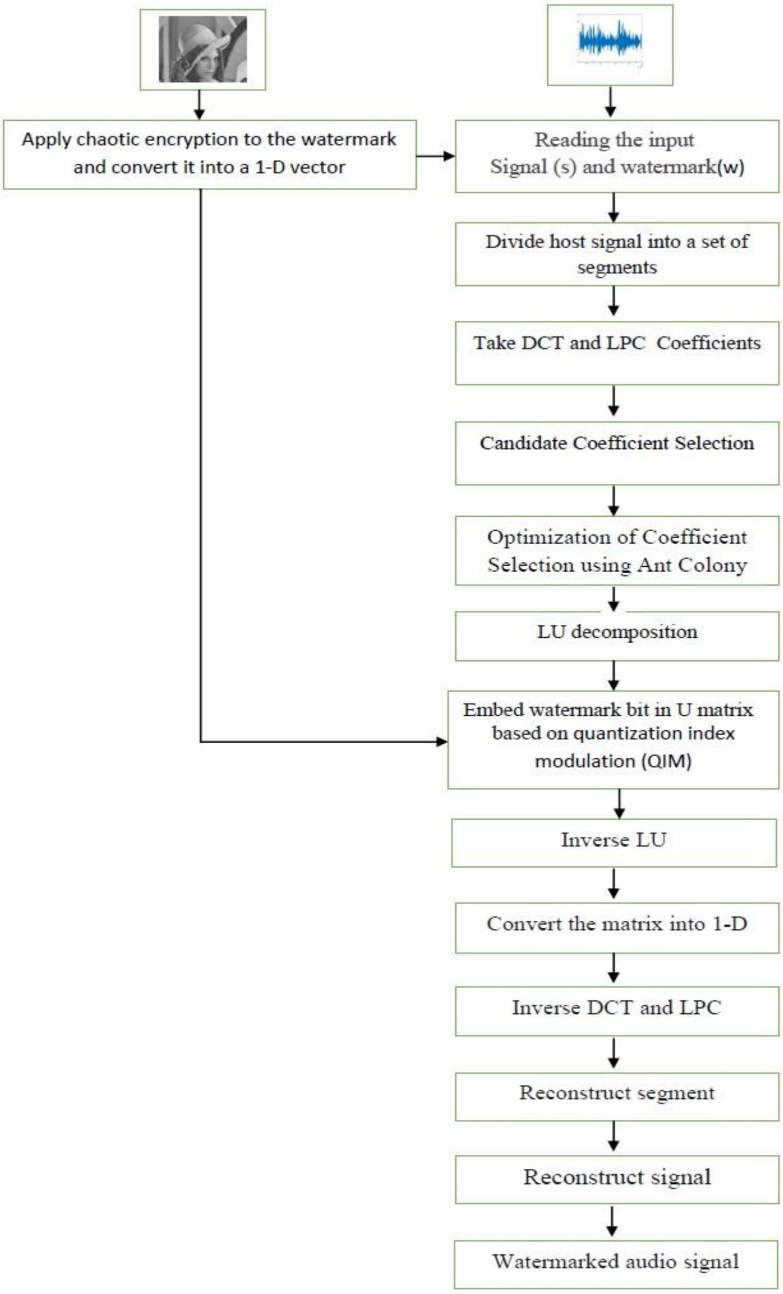

Figure 2. Flowchart of the proposed embedding

10.57647/ijm2c.2026.1602.13

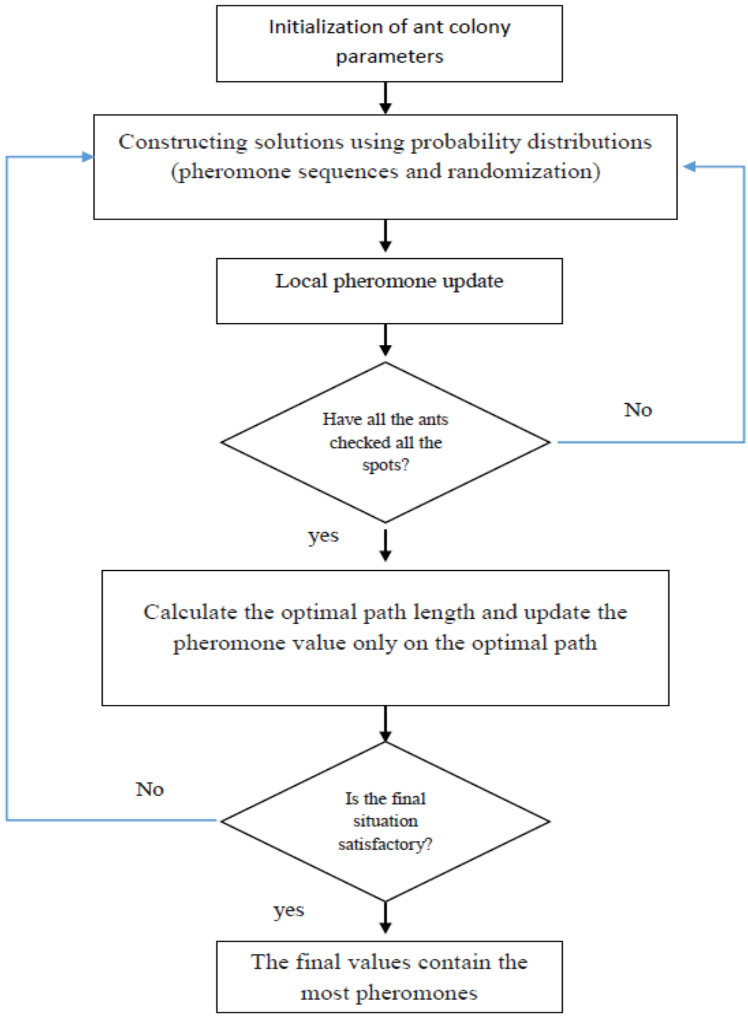

Figure 3. Block diagram of basic Ant Colony Optimization algorithm

Step 3 Candidate Coefficient Selection: A candidate set S={s1,s2,…,sQ} is formed by selecting LPC and DCT coefficients that are perceptually less sensitive and spectrally stable. Typically, mid-frequency DCT coefficients and stable LPC coefficients are chosen.

Step 4 Optimization of Coefficient Selection using Ant Colony Optimization (ACO): Each ant k iteratively constructs a solution by selecting a subset of coefficients from the candidate set.
The probability $P_{ij}^{k}(t)$ of choosing coefficient j after coefficient i at iteration t is given by:

$$P_{ij}^{k}(t) = \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{l \in N_i^k} [\tau_{il}(t)]^{\alpha} \cdot [\eta_{il}]^{\beta}}, \tag{20}$$

where:

- τ_{ij}(t) is the pheromone level on edge i→j.
- η_{ij} is the heuristic desirability of choosing coefficient j (e.g., inverse of distortion or error),
- α and β control the relative importance of pheromone and heuristic information,
- N_i^k is the feasible neighborhood for ant k after coefficient i.

Once each ant constructs its path, the quality of the solution F_k is evaluated, commonly based on a weighted sum of metrics such as signal-to-noise ratio (SNR), robustness, and embedding capacity: The Ant Colony Optimization (ACO) algorithm is employed to identify the optimal set of LPC residual coefficients for watermark embedding. The core of this optimization lies in a precisely defined multi-objective fitness function, F(S), which the algorithm seeks to maximize.

This function balances the conflicting objectives of watermark imperceptibility and robustness for each candidate solution SSS (representing a specific selection of embedding coefficients):

$$F(S) = w_{NC}\ .NC(S) + w_{BER}\left(1 - \frac{BER(S)}{100}\right) \qquad (21)$$
$$+w_{SNR}\frac{SNR(S)}{SNR_{max}}.$$

Where:

- S denotes a particular set of candidate embedding coefficients.
- NC(S) is the Normalized Correlation (range [0, 1]) between the embedded and extracted watermark bits, reflecting the robustness of the watermark against attacks.
- BER(S) is the Bit Error Rate (in percentage, range [0, 100]) of the extracted watermark. The term $(1 - \frac{BER(S)}{100})$ normalizes this metric, such that higher values indicate superior robustness.
- SNR(S) is the Signal-to-Noise Ratio (in dB) between the original and watermarked audio, quantifying the imperceptibility of the embedded watermark.
- SNRmax is a normalization constant, typically set to a maximum expected SNR value (e.g., 50 dB), used to scale the SNR metric into a comparable range.
- w_{NC} , w_{BER}, and w_{SNR} are non-negative weighting factors that sum to 1 (w_{NC} +w_{BER}, +w_{SNR}=1). These factors allow the system to adjust the relative importance of robustness (NC and BER) versus imperceptibility (SNR) based on specific application requirements. For our experimental evaluations, a balanced performance with a slight emphasis on robustness was achieved by setting w_{NC} =0.4, w_{BER}, =0.3, and w_{SNR}=0.3.

This fitness function ensures that the ACO algorithm effectively navigates the search space to converge on a solution that provides an optimal trade-off between the desired watermarking properties, contributing to the overall superior performance observed in the experimental results. Pheromone updates follow:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}(t), \qquad (22)$$

where:

$$\Delta\tau_{ij}^{k}(t) = \begin{cases} \frac{Q}{F_k}, & \text{if ant } k \text{ used edge } (i,j), \\ 0, & \text{otherwise,} \end{cases} \qquad (23)$$

with ρ being the evaporation rate, Q a constant, and mmm the number of ants.

Step 5 Watermark Embedding: After obtaining the optimal subset of coefficients {sopt}, the binary watermark bits bi∈{0,1} are embedded into these selected coefficients using the Quantization Index Modulation (QIM) rule. Each coefficient sopt,is modified as follows:

$$\Delta \times Q\left(\frac{i^s}{\Delta} + b_i\right) = s_i'. \qquad (24)$$

Where:

- si′ is the modified (watermarked) coefficient,
- Δ is the quantization step size optimized by the ACO algorithm, and
- round (·) is the nearest-integer (quantization) function.

This embedding rule shifts each selected coefficient by a small and controlled amount proportional to Δ/2 depending on the embedded bit value (bi=0 or 1). Because the step size Δ is adaptively optimized, the introduced change remains imperceptible while ensuring robust bit recovery under various attacks. Where s_i' is the modified coefficient, Δ is the quantization step size, and Q(·) is the quantization function.

Step 6 Storing Modified Coefficients Using LU Decomposition: The modified coefficients are arranged into a square matrix M∈$R^{N\times N}$.
LU decomposition is performed:

$$M = L \times U,$$

where L is a unit lower triangular matrix, and U is an upper triangular matrix.
The watermark-embedded coefficients are stored specifically in the U matrix. The structural properties of U provide numerical stability and robustness to noise or attacks, facilitating accurate watermark extraction. As discussed in Section III.B (“Theoretical Justification for Robustness Enhancement”), the use of the upper triangular matrix U minimizes error propagation, and its interaction with the low-energy LPC residuals ensures a stable and blind extraction process.
End

10.57647/ijm2c.2026.1602.13

Watermark Extraction Process

The watermark extraction method employed in the proposed scheme is blind, meaning it does not require access to the original audio signal. The detailed steps of the extraction process are outlined below:

Extraction Procedure:

- **Input:** Received audio signal s[n]
- **Output:** Extracted watermark data

Begin:

Step 1 – Frame Segmentation & Feature Extraction

- Divide the received audio signal into frames of length N.
- Extract LPC and DCT coefficients from each frame.
- Form the initial candidate coefficient set.

Step 2 – Coefficient Optimization via ACO

- Apply the Ant Colony Optimization algorithm, identical to the embedding stage, to select the same optimal coefficient positions.
- Ensure location mapping between embedding and extraction stages is preserved.

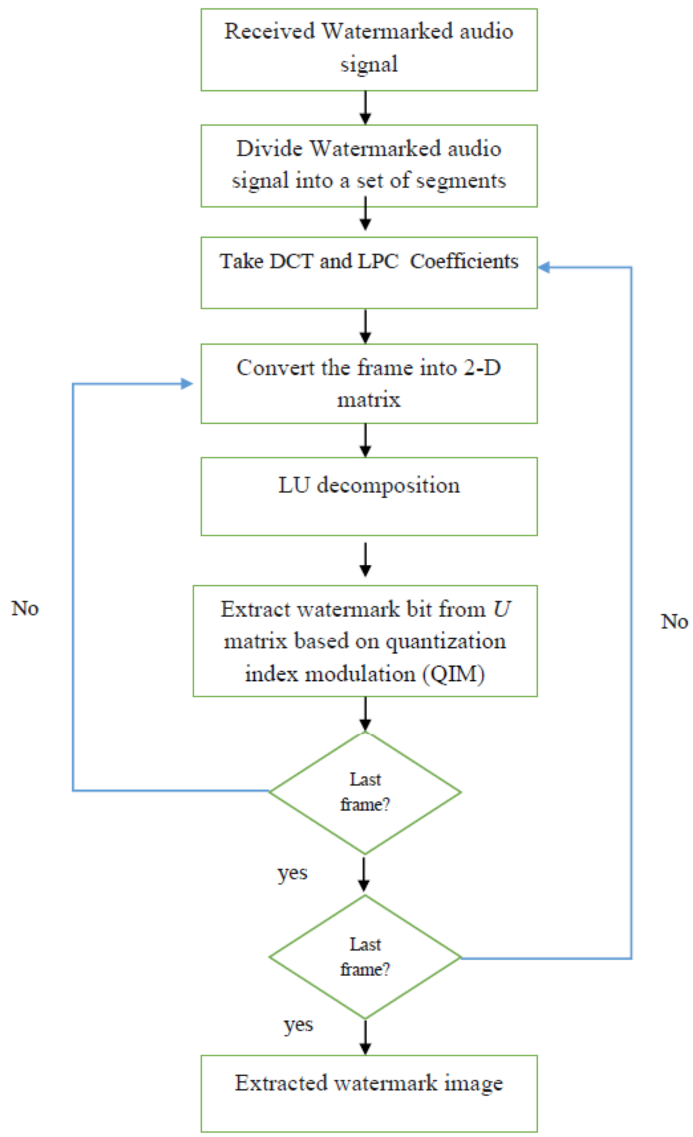

Figure 4. Flowchart of the proposed watermark extraction

10.57647/ijm2c.2026.1602.13

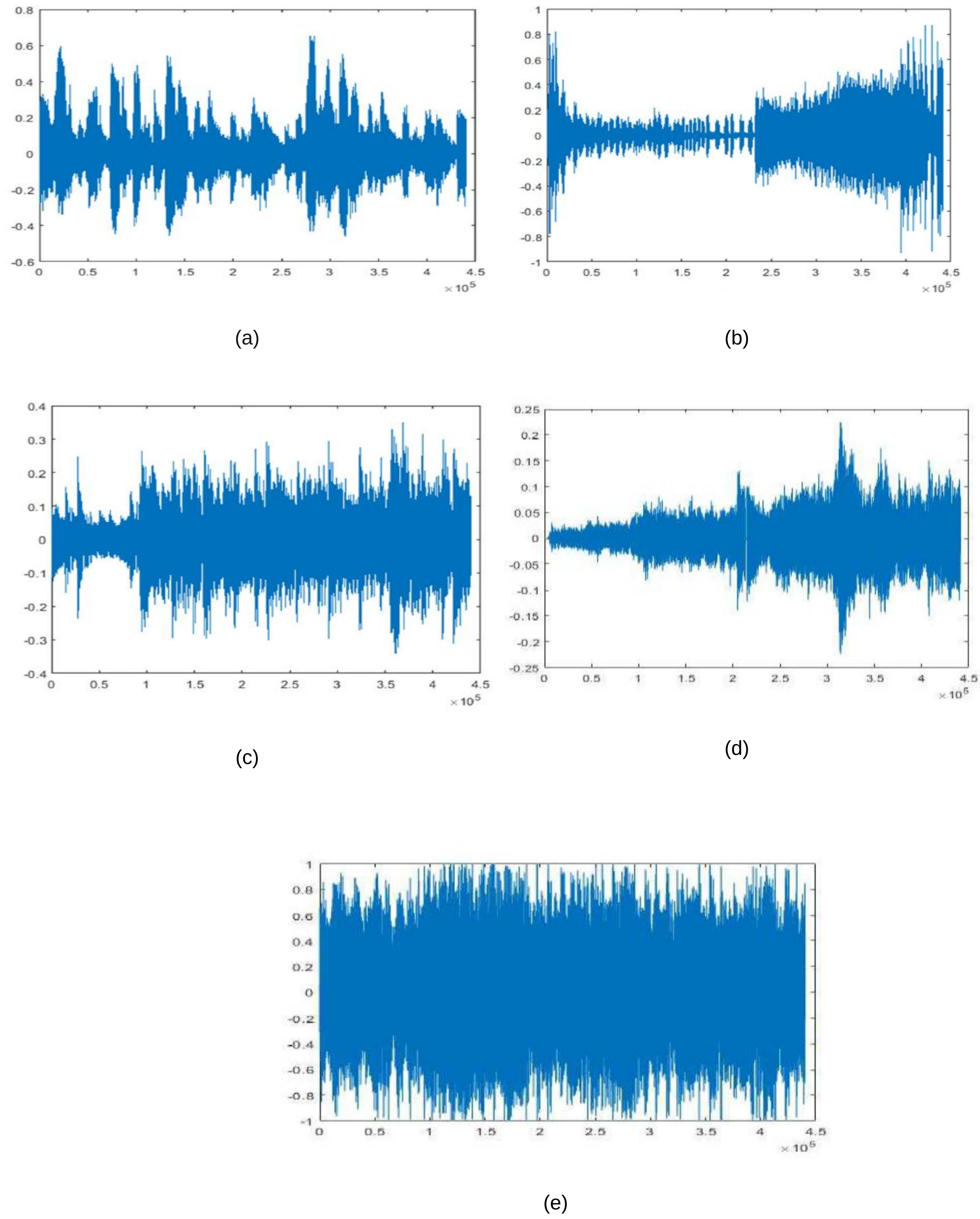

Figure 5. Audio waveforms of the five audio used in the experiments: (a) blues, (b) electronic (c) rock, (d) classical, (e) jazz

Step 3 – LU Decomposition of Reconstructed Matrix

- Construct a square matrix from the selected LPC and DCT coefficients.
- Perform LU decomposition and isolate the upper triangular matrix U, which contains the watermark-modified coefficients.

Step 4 – Watermark Bit Extraction

In the watermark extraction stage, the selected coefficients from the upper-triangular matrix U′ are analyzed according to the QIM rule to retrieve the embedded bits:

$$mod\ 2 \left\lfloor \frac{s_i'}{\Delta} \right\rfloor = b_i. \quad (25)$$

If the result is 0, a watermark bit '0' is recovered; if the result is 1, a watermark bit '1' is obtained.

Step 5 – Reconstruction of Watermark Sequence

- Arrange and concatenate the extracted bits in the original order to form the complete watermark sequence.

Step 6 – Verification

- Assess the quality of extraction using metrics such as Normalized Correlation (NC) or Bit Error Rate (BER).
- An NC close to 1 indicates successful and robust extraction.

$$\mathrm{NC}(W, \hat{W}) = \frac{\sum_{i=1}^{L} b_i \cdot \hat{b}_i}{\sqrt{\sum_{i=1}^{L} b_i^2} \cdot \sqrt{\sum_{i=1}^{L} \hat{b}_i^2}}. \quad (26)$$

A value of NC≈1 indicates a successful extraction.
End

5. Experimental Results

A. Experimental Setup

To evaluate the effectiveness of the proposed audio watermarking scheme, a series of experiments were conducted using standard audio datasets and evaluation metrics commonly adopted in the literature. The setup is described as follows:

1) Audio Dataset Expansion:

The experimental dataset comprised five audio clips representing five distinct genres: Rock, Jazz, Electronic, Classical, and Blues.
Each clip was recorded in WAV format with a sampling rate of 44.1 kHz and 16-bit resolution. To ensure consistency across all embedding and extraction experiments, all audio signals were normalized and segmented into 10-second frames for uniform analysis.
The waveform of these audio signals is shown in Supplementary Material Figure 4.

2) Watermark Information:

Watermark information is a Lena picture which is shown in Figure 6.

Figure 6. Watermark data

3) Evaluation Metrics:

The proposed system is evaluated based on the following performance metrics:

- **Signal-to-Noise Ratio (SNR):**
  Measures the perceptual transparency of the watermarked audio. It is calculated as:

$$\mathrm{SNR} = 10\log_{10}\frac{\sum\ S^2[n]}{\sum_n\ [\,S[n] - S'[n]\,]^2}, \quad (27)$$

where x(n) is the original audio and x^(n) is the watermarked version.

- **Bit Error Rate (BER):**
  Assesses the robustness of watermark recovery under various attacks. It is defined as the ratio of incorrect bits after extraction to the total number of embedded bits.

$$\mathrm{BER}(\mathrm{W}, \overline{\mathrm{W}}) = \frac{\sum_{i=1}^{\mathrm{M}} \sum_{j=1}^{\mathrm{M}} \mathrm{W}(\mathrm{i,j}) \oplus \overline{\mathrm{W}}(\mathrm{i,j})}{\mathrm{M} \times \mathrm{M}}. \quad (28)$$

- **Normalized Correlation (NC):**
  Indicates similarity between the original and extracted watermark sequences, given by:

$$NC = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} W(i,j) \cdot \bar{W}(i,j)}{\sqrt{\sum_{i=1}^{M} \sum_{j=1}^{M} W(i,j) \cdot} \cdot \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{M} \bar{W}(i,j)}}. \quad (29)$$

10.57647/ijm2c.2026.1602.13

Hyperparameter Details: ACO: α=1.0, β=2.0, ρ=0.3, Q=100, m=50 ants, t=100 iterations.

B. Robustness Evaluation

In order to validate the resilience of the proposed audio watermarking method, a series of robustness experiments were conducted against commonly encountered signal degradations. These attacks are designed to simulate real-world scenarios such as transmission, storage, or unauthorized processing of audio content, and are widely adopted to benchmark the durability of watermarking systems under adverse conditions. To quantitatively assess robustness, two primary evaluation metrics were employed: Normalized Correlation (NC), which measures the similarity between the original and extracted watermark, and Bit Error Rate (BER), which quantifies the proportion of incorrectly recovered watermark bits. High NC values (close to 1) coupled with low BER indicate successful and robust watermark extraction. The resilience tests considered four key signal degradations, for which detailed results are presented in Tables 2 and 3:

- **Additive Noise:** Introduction of white Gaussian noise to the watermarked audio signal at 30 dB SNR, reflecting various forms of environmental or recording noise.
- **Low Pass Filtering:** Application of a low-pass filter with a cutoff frequency of 4 kHz, simulating scenarios like bandwidth limitation or specific audio processing.
- **MP3 Compression:** Compression of the watermarked audio signal at 64 kbps followed by decompression, a common lossy compression standard in digital audio distribution.
- **Re-sampling:** Down-sampling and then up-sampling the audio signal to 22.05 kHz, mimicking format conversions and quality alterations.

The robustness results of the proposed method were assessed across five diverse 10-second audio clips covering distinct genres: Blues, Electronic, Rock, Classical, and Jazz. These results, based on the BER and NC metrics averaged over these five clips, are presented in Tables 2 and 3. The results presented in Tables 2 and 3 confirm that the proposed watermark embedding and extraction framework retains high correlation and low Let N denote the number of audio samples in a single 10-second frame. The algorithm consists of two principal components:

(1) the LPC and DCT feature extraction

bit error rates across all tested attack scenarios, demonstrating strong robustness and practical applicability in copyright protection and authentication tasks.

C. Imperceptibility Analysis

Imperceptibility is one of the fundamental requirements of any audio watermarking system, ensuring that the embedding process does not introduce audible distortion or degrade the listening quality of the host signal. In this study, the imperceptibility of the proposed method was evaluated using both objective metrics, such as Signal-to-Noise Ratio (SNR) and Normalized Correlation (NC), and perceptual metrics through informal listening tests. The averaged SNR of 47.135 dB across the five diverse 10-second audio clips confirms the high transparency of the embedded signal and minimal audible distortion. Furthermore, Normalized Correlation (NC) values consistently remained above 0.98 even after common attacks, indicating that the watermark is detectable without significantly altering the host audio's perceptual quality. To ensure a fair comparison and reproducibility, all experiments for both the proposed method and baseline comparisons (as presented in relevant tables) were conducted under identical conditions. These included consistent embedding strength, a sampling rate of 44.1 kHz, 16-bit depth, and uniform 10-second frame segmentation. The comparison of the proposed method in terms of transparency and watermark embedding capacity is presented in Table 5. The experimental results confirm that combining the LPC residual domain**,** LU decomposition**,** and chaotic encryption yields a highly robust audio watermarking framework. The proposed scheme effectively resists a wide range of common signal processing attacks without requiring access to the original audio signal during extraction. Furthermore, the method ensures a high level of imperceptibility, preserving the perceptual quality of the host audio. This balance between robustness and transparency makes the approach well-suited for practical applications where audio fidelity is critical, including streaming services, archival systems, and digital broadcasting platforms.

D. Computational Complexity Analysis:

The computational complexity of the proposed hybrid watermarking framework was analyzed both analytically and empirically.

(2) the ACO-guided optimization for determining the embedding positions.

The LPC analysis exhibits a linear computational cost O(N_{LPC}) while the ACO module incurs approximately O(N_{ACO})=O(N·I·M), where I and M are the iteration and ant population counts, respectively. In practice, resulting in an overall linear time complexity approximated as O(N_{total}) ≈O(N). Empirical runtime profiling was performed on an Intel Core i7 @ 3.4 GHz CPU (16 GB RAM) using MATLAB R2023a. The algorithm demonstrated efficient execution in both embedding and extraction stages, confirming that the computational requirements are sufficiently low for real-time processing. The ACO module converged rapidly without the need for an exhaustive parameter search. This efficiency, attributed to the bounded number of optimization iterations and the lightweight nature of LPC computations, ensures linear scalability with respect to audio length. The resulting balance between computational cost and watermark robustness makes the proposed framework highly suitable for streaming-based and real-time copyright verification applications.

Table 2. BER of the proposed method under Stirmark attacks

Attack	Music Files				
	Blues	Electronics	Rock	Classical	Jazz
Nothing	0.1931	0.2748	0.1876	0.1716	0.1963
Addbrumm_1100	0.4876	0.6899	0.5473	0.7498	0.5796
Addbrumm_2100	0.5794	0.4967	0.6887	0.6389	0.8457
Addsinus	0.8346	0.9836	0.9827	0.8374	0.9375
Compressor	5.3498	4.9375	5.9386	0.9765	1.9658
Extrastereo_50	0.3478	0.3754	0.2967	0.6482	0.3716
FFT_invert	0.5781	0.6795	0.7248	0.4937	0.9437
FFT_real_reverse	3.2131	2.4397	3.1414	0.4973	0.6438
Invert	0.1754	0.5124	0.3952	0.3652	0.4197
Lsbzero	0.7478	0.9475	0.9824	0.7197	0.6171
Rc_highpass	0.6354	0.5245	0.7197	0.8371	0.9375
Additive Noise (30 dB SNR)	0.9792	1.0424	1.0945	0.9724	1.1426
Low-Pass Filter (4 kHz)	0.9758	0.9747	1.0645	0.9899	0.7989
Re-sampling (22.05 kHz)	1.0970	1.0985	0.9737	0.9867	1.0476
MP3 Compression (64 kbps)	4.8670	5.4975	4.6579	3.9867	3.6475

Table 3. NC of the proposed method under Stirmark attacks

Attack	Music Files				
	Blues	Electronics	Rock	Classical	Jazz
Nothing	0.9989	0.9941	0.9978	0.9914	0.9921
Addbrumm_1100	0.9960	0.9889	0.9831	0.9912	0.9965
Addbrumm_2100	0.9971	0.9497	0.9896	0.9788	0.9801
Addsinus	0.9802	0.9876	0.9806	0.9579	0.9719
Compressor	0.9801	0.9868	0.9804	0.9825	0.9811
Extrastereo_50	0.9930	0.9904	0.9957	0.9903	0.9856
FFT_invert	0.9791	0.9908	0.9893	0.9904	0.9943
FFT_real_reverse	0.9985	0.9981	0.9994	0.9811	0.9851
Invert	0.9721	0.9814	0.9703	0.9816	0.9808
Lsbzero	0.9903	0.9941	0.9764	0.9854	0.9837
Rc_highpass	0.9970	0.9902	0.9809	0.9944	0.97877
Additive Noise (30 dB SNR)	0.9607	0.9637	0.9671	0.9704	0.9777
Low-Pass Filter (4 kHz)	0.9791	0.9683	0.9695	0.9679	0.9725
Re-sampling (22.05 kHz)	0.9758	0.9647	0.9687	0.9639	0.9703
MP3 Compression (64 kbps)	0.9651	0.9704	0.9737	0.9672	0.9766

10.57647/ijm2c.2026.1602.13

Table 4. Comparison of the proposed method in terms of transparency and watermark embedding capacity

AUDIO SIGNAL	PAYLOAD	SNR	NC	BER (%)	ODG
BLUES	649	39.21	0.9842	1.3374	- 0.39
ELECTRONICS	985	30.98	0.9866	1.4316	- 0.23
ROCK	570	70.63	0.9750	1.4930	-0.25
Classical	796	40.30	0.9730	0.9247	-0.31
Jazz	581	54.75	0.9818	1.0063	-0.29
AVERAGE	716.2	47.135	0.9801	1.2386	−0.294

Table 5. Comparison of the proposed method with previous works in terms of transparency and capacity

Reference	Payload (bps)	SNR (dB)	NC	BER	ODG
MDWT + DCT [13]	86.13	19.92	0.9405	3.1740	−0.334
LU+ DWT [14]	1280	31.02	0.952	2.1501	Not reported
DCT + LU + GA[15]	2002	33.41	0.9911	1.6359	- 0.32
Fuzzy + DCT [20]	598.34	49.8093	0.9762	1.3458	−0.18 to −0.46
thresholding and LWT (LH sub-band embedding)[21]	178.08	38 TO 42	0.9710	1.6810	Not reported
Proposed method(Average)	716.2	47.135	0.9801	1.2386	−0.23 to −0.39

6. Conclusion and Future Works

This paper has presented a robust and secure blind audio watermarking scheme that synergistically integrates Linear Predictive Coding (LPC) residuals, LU decomposition, chaotic encryption, and Ant Colony Optimization (ACO) for optimized embedding. The watermark is embedded using Quantization Index Modulation (QIM) within the LPC residual coefficients, and the encrypted bits are structurally distributed through the upper triangular matrix obtained via LU decomposition, enhancing numerical stability and enabling blind extraction without requiring access to the original audio signal. The chaotic logistic map and optimized ACO parameters ($\alpha = 1.0, \beta = 2.0, \rho = 0.3, Q = 100$) further secure and adaptively balance robustness and inaudibility.

Future Works

Future research can focus on several directions to further enhance the proposed watermarking framework:

• **Adaptation against emerging attacks:** Developing adaptive embedding strategies to resist AI-driven compression, synthetic voice cloning, and deepfake audio generation, ensuring authenticity verification in generative-AI environments.

• **Payload scalability:** Investigating the method's capacity to embed larger watermark payloads while preserving both transparency and robustness, enabling higher information density for practical applications.

• **Real-time implementation:** Optimizing algorithmic modules for low-latency embedding and extraction to support real-time audio streaming, broadcasting, and smart-device use cases.

• **Cross-domain applicability:** Extending the approach for secure watermarking in diverse domains such as speech-recognition datasets, IoT-based acoustic systems, and multimodal media content.

• **Cross-lingual watermarking:** Exploring watermark invariance across multiple languages and accent variations to ensure reliable protection in multilingual and cross-cultural audio content.

These prospective directions will expand the robustness, scalability, and applicability of the proposed method under evolving audio security challenges.

Adversarial Defense against Deepfake Threats: Futur e research can investigate adversarially trained watermarking models capable of resisting AI-generated audio manipulations and deepfake synthesis. By integrating differentiable attack simulators and adversarial training strategies, the system could adaptively learn embedding patterns that remain recoverable even under neural vocoder or generative-model perturbations.

• **Cross-lingual and Multilingual Extension:** Another promising direction lies in extending the proposed LPC-based watermarking framework to multilingual and cross-lingual speech corpora. Developing generalized LPC or hybrid acoustic models (e.g., multilingual LPC–DCT parameterizations) would enable consistent imperceptibility and robustness across languages and phonetic distributions, enhancing the versatility of watermarking for global speech datasets.

Authors Contribution
All the authors have participated sufficiently in the intellectual content, conception and design of this work or the analysis and interpretation of the data (when applicable), as well as the writing of the manuscript.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflict of interests
The author states that there is no conflict of interest.

References

[1] Ray A, Roy S. Recent trends in image watermarking techniques for copyright protection: a survey. International Journal of Multimedia Information Retrieval. 2020; 9(4): 249-270.

[2] Kim D-W, et al. Transient detection-based adaptive audio watermarking using attack-aware optimization. Digital Signal Processing. 2024; 146: 104352.

[3] Uddin MS, Hasan M, Shimamura T. Audio Watermarking: A Comprehensive Review. International Journal of Advanced Computer Science & Applications. 2024; 15(5).

[4] Salah E, et al. Survey of imperceptible and robust digital audio watermarking systems. Multimedia Tools and Applications. 2025; 84(7): 3635-3681.

[5] Hu Y, et al. Selection of the optimal embedding positions of digital audio watermarking in wavelet domain. arXiv. 2020;arXiv: 2010. 11461.

[6] Zhao J, et al. Desynchronization attacks resilient watermarking method based on frequency singular value coefficient modification. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2021; 29: 2282-2295.

[7] Kim D-W, et al. Transient detection-based adaptive audio watermarking using attack-aware optimization. Digital Signal Processing. 2024; 146: 104352.

[8] Zhang J, et al. Chebyshev chaotic mapping and DWT-SVD-based dual watermarking scheme for copyright and integrity authentication of remote sensing images. Symmetry. 2024; 16(8): 969.

[9] Dong Y, Yan R, Yin C. An adaptive robust watermarking scheme based on chaotic mapping. Scientific Reports. 2024; 14(1): 24735.

[10] Wu W, Dong Y, Wang G. Image Robust Watermarking Method Based on DWT-SVD Transform and Chaotic Map. Complexity. 2024; 2024(1): 6618382.

[11] Farzaneh M, Toroghi RM. Robust audio watermarking using graph-based transform and singular value decomposition. In: 2020 10th International Symposium on Telecommunications (IST). IEEE; 2020.

[12] Al-Qaheri H, Mustafi A, Banerjee S. Digital watermarking using ant colony optimization in fractional Fourier domain. J Inf Hiding Multim Signal Process. 2010; 1(3): 179-189.

[13] Hu H-T, Chang J-R. Efficient and robust frame-synchronized blind audio watermarking by featuring multilevel DWT and DCT. Cluster Computing. 2017; 20(1): 805-816.

[14] Kaur A, Dutta MK. An optimized high payload audio watermarking algorithm based on LU-factorization. Multimedia Systems. 2018; 24(3): 341-353.

[15] Mosleh M, et al. High-capacity, transparent and robust audio watermarking based on synergy between DCT transform and LU decomposition using genetic algorithm. Analog Integrated Circuits and Signal Processing. 2019; 100(3): 513-525.

[16] Abdelwahab KM, et al. Efficient SVD-based audio watermarking technique in FRT domain. Multimedia Tools and Applications. 2020; 79(9): 5617-5648.

[17] AlShaikh M. Robust and recovery watermarking approach based on SVD and OTP encryption. Journal of Signal Processing Systems. 2024; 96(6): 385-399.

[18] Araghi TK, Megías D. Analysis and effectiveness of deeper levels of SVD on performance of hybrid DWT and SVD watermarking. Multimedia Tools and Applications. 2024; 83(2): 3895-3916.

[19] Chaudhary H, Vishwakarma VP. Analysis of healthcare data security with DWT-HD-SVD-based algorithm invisible watermarking against multi-size watermarks. Scientific Reports. 2024; 14(1): 10769.

[20] Mosleh M, et al. A novel audio watermarking scheme based on fuzzy inference system in DCT domain. Multimedia Tools and Applications. 2021; 80(13): 20423-20447.

[21] Alaoui N. DWT-based digital watermarking for various attacks. Acadlore Trans Mach Learn. 2023; 2(4): 226-234.

[22] Velayatipour M, et al. A Robust Quantum Audio Watermarking Using Synergy of Echo Hiding and Least Significant Bit Technique. Electronic and Cyber Defense. 2024; 12(1): 59-76.

[23] Mushtaq S, Mehraj S, Parah SA. Blind and robust watermarking framework for audio signals. In: 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). IEEE; 2024.

[24] Zong W, et al. AudioMarkNet: Audio Watermarking for Deepfake Speech Detection. USENIX Security. 2025.

[25] Liu C, et al. Dear: A deep-learning-based audio re-recording resilient watermarking. AAAI Conference on Artificial Intelligence. 2023; 37(11).

[26] Wen S, et al. Robust Audio Watermarking Against Manipulation Attacks Based on Deep Learning. IEEE Signal Processing Letters. 2024.

[27] Ben Jabra S, Ben Farah M. Deep learning-based watermarking techniques challenges: a review of current and future trends. Circuits, Systems, and Signal Processing. 2024; 43(7): 4339-4368.

[28] Venkateswaran P, et al. An Efficient Time Domain Speech Compression Algorithm Based on LPC and Sub-Band Coding Techniques. J Commun. 2009; 4(6): 423-428.

[29] Mosleh M, Setayeshi S, Kheyrandish M. Presenting a systematic method for LU decomposition of a matrix with linear systolic arrays. In: 2008 International Conference on Advanced Computer Theory and Engineering. IEEE; 2008.

[30] Liu Y, et al. Advances in chaotic characteristics of serum Raman spectroscopy for assisted diagnosis of disease. Expert Systems with Applications. 2024; 238: 121787.

[31] Ochoa-Dominguez H, Rao KR. Discrete Cosine Transform. CRC Press; 2019.

[32] Fidanova S. Ant colony optimization. In: Ant Colony Optimization and Applications. Springer; 2021. p. 3-8.

10.57647/ijm2c.2026.1602.13