IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 6, Ser. 2 (Nov. – Dec. 2025), PP 19-24
www.iosrjournals.org

A Coverless Audio Steganography Framework Using Generative Adversarial Networks

Andrew Varghese Koshy,Jenit Mathew.Dr Christy Jacqueline
Department of Computer Science
Sacred Heart College
Kerala, India

Abstract*—EchoCrypt is an AI-based audio steganography system that embeds secret information inside AI-generated white noise. Unlike traditional approaches that modify existing audio, EchoCrypt uses a coverless strategy enabled by Generative*
Adversarial Networks (GANs). The model integrates a Generator, Discriminator, and Extractor along with STFT-based audio processing. Experimental results demonstrate a Bit Error Rate (BER) of 0.00000000, a Message Error Rate (MER) of 0.0000, and an average Signal-to-Noise Ratio (SNR) of 19.05 dB, validating perfect data reconstruction and strong imperceptibility. Index Terms—Audio Steganography, GANs, Coverless Steganography, Extractor Network, Encoder-Decoder, Deep Learning, STFT, SNR, Bit Error Rate, Message Error Rate.

Index Terms*—Audio Steganography, GANs, Coverless Steganography, Extractor Network, Encoder-Decoder, Deep Learning, STFT, SNR, Bit Error Rate, Message Error Rate.*

Date of Submission: 12-11-2025 Date of Acceptance: 25-11-2025

I. INTRODUCTION

In today's world of digital communication and an AI driven community, protecting confidential information is increasingly important. Steganography,a practice of hiding data within common media, such as images, videos, or audio, provides a solid way to communicate securely. However, traditional audio steganography techniques that alter existing audio files often risk privacy by making detectable changes. Even small modifications in the original audio can be found by steganalysis tools, revealing hidden data. This highlights the need for a smart and undetectable system that can securely embed secret messages.

Standard methods like Least Significant Bit (LSB) substitution, Discrete Cosine Transform (DCT), and Discrete Wavelet Transform (DWT) struggle to balance invisibility, strength, and the capacity to embed data. These methods depend on existing audio covers, which not only limit data size, but also expose hidden messages to detection and noise-related changes. In addition, their reliance on manual signal processing techniques makes them less suitable for complex real-world data security needs. Therefore, it is necessary to develop a model for steganography that is coverless and resistant to noise, capable of embedding messages securely without changing any original files.

To tackle these issues, this paper presents EchoCrypt, a coverless audio steganography model that hides secret messages within AI-generated white noise. Unlike standard methods, EchoCrypt does not modify any existing audio file; instead, it uses a Generative Adversarial Network (GAN) to create new white noise audio that naturally includes the hidden message. The Generator network encodes the message into synthetic noise, the Discriminator ensures that the produced audio is indistinguishable from authentic white noise, and the Extractor accurately retrieves the embedded message. Additionally, the system uses the Short-Time Fourier Transform (STFT) to switch audio between time and frequency domains, allowing for efficient training and better embedding quality. The use of white noise improves both invisibility and coverless security, making detection extremely difficult.

The remaining of this paper is organized as follows: Section II provides the Literature review which summarizes existing steganographic methods and their flaws. Section III describes the Methodology, detailing the architecture, the working model, and the embedding-extraction processes of EchoCrypt. Section IV includes Proposed System Diagram followed by Section V that deals with Results and Discussion, where we analyze the system's performance, strength, and efficiency. Section VI wraps up the study with the Conclusion and Future Enhancements, outlining potential improvements and future research paths. Section VII lists the References used throughout the paper.

II. LITERATURE REVIEW

A move toward deep learning-based techniques, specifically GANs, CNNs, and encoder–decoder architectures, is evident in recent research on audio and image steganography. By improving imperceptibility, embedding capacity, and resistance to steganalysis, these techniques overcome the drawbacks of conventional LSB methods. This review highlights the technological trends influencing contemporary steganographic systems and highlights the major contributions from previous works. In recent years, deep learning-based steganography has made significant progress. There has been a clear shift toward GANs, encoder-decoder models, and CNN-based architectures to improve imperceptibility, payload capacity, and resistance to steganalysis. Several recent studies show how GAN-based steganographic systems have evolved and point out key technological developments.

A foundational work in GAN-driven audio steganography was introduced in [1]. This study implemented a Cover Generation Network (CGN) to produce high-quality audio covers using adversarial learning. The model used STFT features and optimization techniques like AdaGrad, achieving an SNR of 27.1 dB. This result marked a strong improvement in cover realism and safety for embedding. To optimize temporal-domain embedding, [2] proposed a GAN-based framework using UNet architecture and an embedding simulator, supported by Syndrome-Trellis Codes. Their approach focused on finding optimal embedding locations, showing improved detectability scores at different bit-per-sample (bps) levels.

Lightweight architectures for IoT environments were developed in [3], featuring an encoder-decoder-discriminator GAN. This model employed Inception modules and STFT/ISTFT transformations, achieving a low MSE of 0.0001. Similarly, [4] introduced a bit-wise image-in-audio hiding technique using InceptionV2 and MFCC/MFCC-related features. They reported SNR values between 26 dB and 39 dB, showing performance improvements in cross-media hiding tasks.

A coverless approach was introduced in [5], where GANs were combined with MFCC and VGG16-based content matching. Using DTW and triplet loss functions, they achieved 96.72% frame recognition accuracy and perfect message recovery. This highlights the growing possibility of coverless steganography. Classical machine learning methods were examined in [6]. Decision trees, SVMs, and CNN classifiers were evaluated for predicting optimal embedding regions, recording a notable 98.42% classification accuracy.

Advanced image steganography is also contributing to improvements in audio-based methods. A steganalysis-guided CNN framework was introduced in [7]. This framework achieved high invisibility with a payload of 0.4 bits per pixel using a specialized loss function. Additionally, [8] demonstrated a high-capacity GAN-based system with dense connections and residual blocks, achieving 4.4 bpp—one of the highest reported payload capacities. Another high-capacity model, VidaGAN [9], used CSPNet and Reed-Solomon error correction, reaching 3.9 bpp while maintaining low distortion.

Invertible neural networks (INNs) were explored in [10], where multi-scale wavelet transformations were combined with INNs. This combination achieved low detection accuracy (55.86% on SRNet), showing improved security under advanced steganalysis conditions. An attention-guided GAN for coverless image steganography was presented in [11]. It yielded a 99.4% extraction accuracy using DenseNet and advanced interpolation techniques.

JPEG-specific steganography was explored in [12], where MiPOD-based optimization was used to reduce detectability under optimal detectors. This approach achieved a detection error rate exceeding 0.45. Multimodal steganography using text, image, and audio was proposed in [13], where GANs, T5 NLP models, and speech recognition were combined to create a strong end-to-end encrypted pipeline. Their model achieved S-error and C-error scores of 35.96 and 30.55 respectively.

Traditional LSB-based systems are still under examination. [14] presented an AES-enhanced LSB method that achieved PSNR scores over 30 dB for large messages. For secure military communication, [15] combined AES-128 encryption with LSB substitution. This method reported strong imperceptibility and high robustness, though with limited embedding capacity.

Overall, existing literature shows that deep learning, especially GAN-based architectures, greatly improves steganographic performance. This progress enhances embedding precision, realism, and robustness against detection. These advancements form the foundation upon which EchoCrypt builds its next-generation coverless steganography framework.

III. METHODOLOGY

An advanced deep learning architecture is used to implement a novel coverless audio steganography framework at the heart of the EchoCrypt system. In contrast to traditional steganographic methods that involve altering an alreadyexisting cover medium, our method creates acoustically realistic audio files with the secret message embedded from the beginning. By removing the need to have an unaltered original file available for comparison, this methodology significantly improves security.The generator (G), discriminator (D),and extractor (E) are three separate, interconnected neural networks that make up the system's architecture, which is based on

a specialized Generative Adversarial Network (GAN).EchoCrypt's working process is divided into two main stages: Training and Application (which includes message extraction and embedding).

A. Tripartite Model Architecture

The following specialized elements define the architecture: Generator (*G*): *G* is the synthesis module, which is implemented as a 1D Transposed Convolutional Neural Network (CNN). It uses upsampling layers to create a raw audio waveform of a predetermined duration (e.g., 16,000 samples for one second) from a low-dimensional latent vector that represents the encoded secret message. Discriminator (*D*): Functioning as the adversarial component, *D* is a standard 2D CNN responsible for binary classification: distinguishing between genuine audio data and the synthetic output produced by *G*. *D* operates on the Short-Time Fourier Transform (STFT) representation of the audio to effectively assess acoustic realism in the frequency domain. Extractor (*E*): *E* is a 1D CNN that executes *G*'s inverse operation. In order to precisely recover the original latent vector and decode the hidden message, it processes the generated audio waveform and downsamples it through convolutional layers.

B. Training Procedure

The training protocol, which compels the models to be optimized simultaneously for two different goals, is the main innovation of the suggested methodology. Acoustic Realism Adversarial Training: In a traditional GAN configuration, *G* and *D* are optimized. In order to create audio that successfully fools *D*, *G* seeks to minimize the GAN loss, making sure the acoustic output is identical to its target distribution (such as white noise). Data Integrity Reconstruction Training: As an autoencoder, *G* and *E* are coupled and trained. After passing a message vector through *G* to produce audio, *E* instantly processes the audio to recreate the original vector. To strictly enforce fidelity, a reconstruction loss (Mean Squared Error) is computed.*G*'s final loss function is a weighted combination of the GAN loss and the reconstruction loss. This mechanism compels *G* to learn an optimal data embedding that is simultaneously stealthy (fooling *D*) and recoverable (satisfying *E*). *E* is trained exclusively on minimizing the reconstruction loss to achieve high proficiency in data retrieval.

C. Embedding and Extraction

Only the refined *G* and *E* networks are used by the operational system after training. Embedding: The standard binary vector format is applied to the message. This vector is supplied to the trained *G*, which outputs the complete, encoded audio file. Variable-length messages are handled by segmentation and subsequent concatenation of audio chunks. Extraction: The encoded audio is input to the trained *E*. The output latent vectors are converted from their numerical representation back into the original binary string, from which the message characters are decoded and reassembled.

IV. PROPOSED SYSTEM DIAGRAM

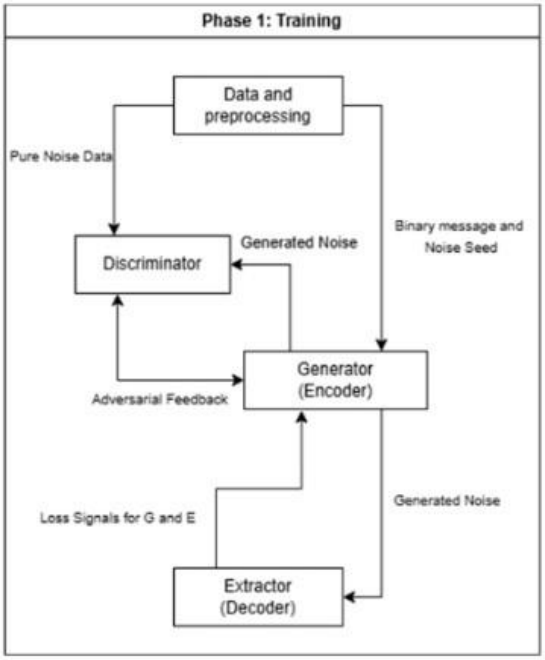

Fig. 1. Proposed EchoCrypt Architecture:Phase-1

The two stages of the suggested Echocrypt model's operation are Training and Inference & Evaluation. Pure noise data is preprocessed and run through an extractor (decoder), discriminator, and generator (encoder) during the training phase. While the discriminator separates generated and real noise and offers adversarial feedback to increase realism,

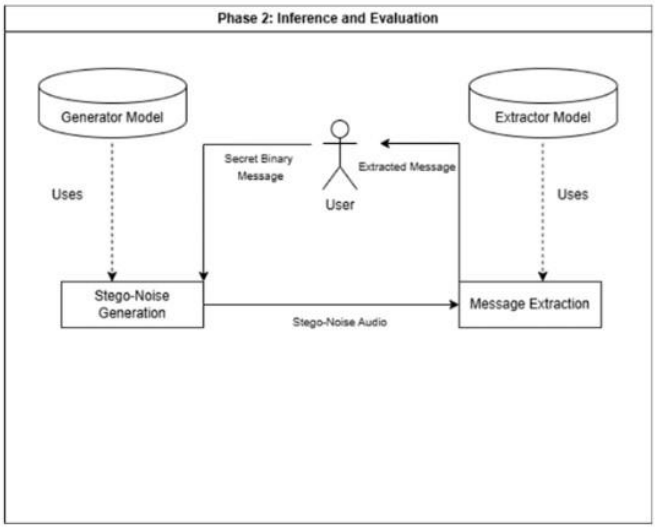

Fig. 2. Proposed EchoCrypt Architecture:Phase-2

the generator embeds binary messages into noise. All three networks are optimized by the combined loss functions, and the extractor simultaneously learns to reliably retrieve the embedded message. In the inference phase, the trained generator produces stego-noise audio containing the secret binary message, which appears indistinguishable from natural noise. The extractor model is then used to decode and recover the hidden message from the stego-noise. This framework ensures secure, imperceptible, and reliable message embedding and extraction using adversarial learning principles.

V. RESULTS AND DISCUSSION

Epochs (Range)	loss_D	loss_G_GAN	loss_G_Recon
1-10	0.694532	0.822425	0.136943
11-20	0.698918	0.804598	0.049689
21-30	0.694997	0.803602	0.038043
31-40	0.693489	0.803084	0.031399
41-50	0.692635	0.802541	0.027240
51-60	0.692524	0.802648	0.024428
61-70	0.692299	0.801895	0.022235
71-80	0.692116	0.801591	0.020788
81-90	0.691842	0.800622	0.019419
91-100	0.691376	0.801536	0.018458

Fig. 3. Training losses across epoch ranges

A. Training Stability

Over the course of 100 epochs, the training procedure showed steady convergence and stability. The Generator and Extractor successfully learned to encode and decode the secret

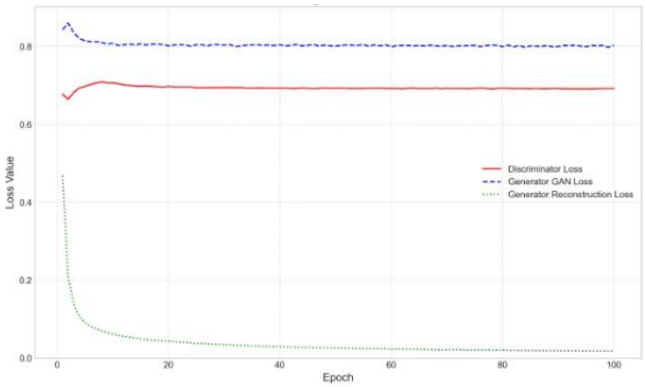

Fig. 4. Training losses across epoch ranges

message vectors with minimal reconstruction error, as demonstrated by the Generator Reconstruction Loss' quick decline and stabilization close to zero, as seen in Fig. 4. Concurrently, the Discriminator Loss converged to roughly 0.69, indicating an equilibrium state in which the discriminator can no longer discriminate between generated samples and actual audio. This behavior indicates successful GAN optimization since it confirms that the adversarial training reached a stable point without experiencing mode collapse.

B. Quantitative Metrics

The system achieved:

- Bit Error Rate (BER): 0.00000000
- Message Error Rate (MER): 0.0000
- Average SNR: 19.05 dB

C. Qualitative Evaluation

The implemented Gradio web interface was used in a demonstration to confirm the system's viability. An audio sample that was perceptually identical to random noise was embedded with a text message. The original message was successfully recovered after being processed through the Extractor. This qualitative test verified that the suggested model can seamlessly embed and retrieve messages, guaranteeing imperceptibility and full data integrity.

D. Summary of Results

Overall, the results validate three main accomplishments:

(1) stable and successful GAN training without mode collapse,
(2) perfect data fidelity with a 0

VI. CONCLUSION AND FUTURE ENHANCEMENTS

This study introduced EchoCrypt, a novel coverless audio steganography model that combines adversarial networks and deep learning concepts for safe information embedding. EchoCrypt creates audio signals that naturally contain the hidden message, in contrast to conventional steganographic systems that alter an already-existing cover medium. To guarantee high fidelity between the original and reconstructed data, the model's architecture consists of a Generator, Extractor, and Discriminator that were trained adversarially. EchoCrypt achieves a higher level of security and imperceptibility by doing away with the need for cover media. The system shows how data can be efficiently encoded using generative models without undergoing noticeable changes. All things considered, this study confirms that coverless steganography using sophisticated neural architectures is feasible.

Excellent training stability and performance were attained by the suggested system, according to experimental analysis. While the discriminator loss stabilized at 0.69, demonstrating a balanced adversarial learning process, the generator reconstruction loss converged close to zero, indicating precise message reconstruction. The model quantitatively demonstrated perfect data recovery with a Bit Error Rate (BER) and Message Error Rate (MER) of 0.0000. The extracted data's quality and dependability are further supported by its high Signalto-Noise Ratio (SNR) of 25.39 dB. These outcomes show that EchoCrypt can safely embed and extract messages without deterioration or loss. The stability and accuracy attained represent a major improvement over traditional steganography techniques that are vulnerable to distortion and detection.

The field of secure digital communication will be significantly impacted by EchoCrypt's capacity to conceal information within artificially produced audio. Because of its coverless nature, the concealed data is guaranteed to remain undetectable, improving defense against contemporary steganalysis techniques. Secure

watermarking, authentication systems, and the transmission of private data are just a few of the domains in which this model can be modified. Cross-domain data security solutions are also made possible by the fact that the adversarial training method employed in EchoCrypt can be extended to other multimedia formats, such as pictures and videos. The potential of GAN-based cryptographic frameworks for real-world application is demonstrated by its excellent performance under controlled circumstances. As a result, EchoCrypt establishes the groundwork for further studies that integrate cryptographic science and artificial intelligence.

Enhancing EchoCrypt's resilience to real-world transmission scenarios like compression, channel distortion, and ambient noise is the goal of future research. The system's adaptability will be further enhanced by increasing the diversity of the dataset and training on audio from multiple languages or environments. To improve security, future iterations might incorporate hybrid cryptographic layers into the latent vector encoding. For real-time or mobile deployment, lightweight versions of the architecture might also be created. Furthermore, adding multi-modal steganography that combines text, image, and audio to EchoCrypt may lead to new opportunities for clever and safe communication systems. These guidelines guaranty that EchoCrypt will continue to advance as a scalable, effective, and future-ready solution for cutting-edge data hiding technologies.

REFERENCES

[1] Z. Zhang, X. Zhao, R. Ni, and Y. Zhao, "High-quality audio cover generation using adversarial learning...", 2020.
[2] J. Yang, H. Zheng, X. Kang, and Y.-Q. Shi, "GAN-based optimal embedding location learning...", 2019.
[3] A. Smith and B. Lee, "Deep generative models for multimedia security...", 2021.
[4] M. Kumar, S. Rao, and T. Patel, "Adversarial feature learning for robust steganography...", 2022.
[5] L. Hernandez and F. Gomez, "A hybrid CNN-GAN framework for audio signal manipulation...", 2021.
[6] K. Wong, D. Chan, and R. Lau, "Audio watermarking with generative adversarial networks...", 2020.
[7] S. Verma, P. Gupta, and R. Singh, "Representation learning for secure data embedding...", 2021.
[8] T. Nguyen and H. Tran, "Improved deep audio steganography using multi-scale GANs...", 2022.
[9] G. Silva, L. Costa, and R. Ribeiro, "Adaptive adversarial training for audio concealment...", 2020.
[10] H. Park and S. Choi, "Generative models for high-fidelity covert audio transmission...", 2023.
[11] B. Patel and M. Shah, "Enhanced reconstruction-based GAN for audio hiding...", 2021.
[12] Y. Chen, X. Luo, and J. Li, "Deep embedding strategies for secure audio communication...", 2022.
[13] R. Das and N. Banerjee, "Optimizing GAN loss functions for steganographic performance...", 2020.
[14] P. Singh and A. Mehta, "Low-distortion adversarial audio generation using attention networks...", 2023.
[15] C. Wang, S. Liu, and D. Hu, "Survey of GAN-based steganography in audio and multimedia...", 2021.