Engineered Science
DOI: https://dx.doi.org/10.30919/es950

A Novel Speech Steganography Mechanism to Securing Data Through Shift Invariant Continuous Wavelet Transform with Speech Activity and Message Detection

Surendra Kumar,[1,*] Arun Singh Yadav,[2] Sanjay Yadav,[3] Niroj Kumar Pani,[4] Juan Carlos Cotrina-Aliaga,[5] Manuel Antonio Cardoza Sernaqué,[6] Christian Paolo Martel Carranza[7] and Dhiraj Kapila[8]

Abstract

Steganography is a technique of concealing secret information within an innocuous-looking carrier signal. Speech steganography has become increasingly important in a variety of applications, including military, satellite, and mobile communications. However, conventional methods have failed to meet the maximum-security requirements, resulting in reduced robustness and imperceptibility performance. In this paper, a novel speech steganography mechanism is proposed to address these issues by using shift invariant continuous wavelet transform (SI-CWT) for data hiding. The message data is hidden in the low-level speech bands using pseudo noise sequences, resulting in a steganography output speech signal. To recover the original speech and message signal, a zero-crossing rate (ZCR)-based speech activity and message detection (SAMD) mechanism is applied on the receiver side. The proposed steganography system is evaluated using various performance metrics, including robustness, imperceptibility, and security. Simulation results demonstrate that the proposed system outperforms conventional methods in terms of these metrics, indicating that it is a promising solution for secure speech communication. Additionally, potential limitations and weaknesses of the proposed system are discussed, highlighting the need for further research in this area. Overall, this novel speech steganography mechanism has significant potential for enhancing the security and reliability of speech communication in various applications.

Keywords: Speech steganography; Shift invariant continuous wavelet transform; Zero-crossing rate; Speech Activity and Message Detection.
Received: 25 May 2023; Revised: 14 July 2023; Accepted: 15 July 2023.
Article type: Research article.

1. Introduction

The globe has become a global village where information, entertainment, attitudes, and other things all move about at breakneck speed thanks to the exponentially developing internet technologies, mobile telephone, speech-video communications, *etc*.[1] There is a continuing need for effective data storage, transport, and retrieval techniques since the amount of data transferred across these networks keeps growing in exponential proportions on a daily basis. One of the technologies that enables this is data steganography,[2] which uses a variety of approaches to steganography huge amounts of data, allowing us to save money, time, space, and bandwidth.

In a wide range of applications,[3] including satellite, mobile, and military communications, speech steganography plays a significant role. The need for highly secure voice and message transfers in these communications necessitates the development of steganography systems with enhanced security. Our use of technology for communication is undergoing revolutionary changes on a constant basis.[4] Depending on how well the reconstructed signal must be, it is classified in this way.

The fundamental steganography technology utilized in real-time speech communications is shown in Fig. 1. Steganography is employed in this case to embed hidden

[1] *Department of Computer Engineering and Applications, GLA University, Mathura 281406, India.*
[2] *Department of Computer Science, University of Lucknow, Lucknow 226007, India.*
[3] *Department of Computer Science, Koneru Lakshmaaiah education foundation vaddeswaram 522302, Guntur A.P., India.*
[4] *Department of Computer Science Engineering and Applications, Indira Gandhi Institute of Technology, Sarang 759146, India.*
[5] *Faculty of Human Medicine, Universidad Privada San Juan Bautista, Chincha 11702, Perú.*

information into each subblock of the cover; the quantity of data samples used for the hidden information should range from 2 to 4, since these increases embedding capacity and minimizes loss.[5] In the original speech transmission, the concealed speeches are divided into intensity blocks, with the embedding process in each block being comparable to the genuine scale one. This procedure continues for another two intensity blocks.[6]

Figure 1 shows the steganography principle in the mobile communication system. Technical steganography and network steganography are two separate ways that have traditionally been used to classify speech steganography techniques.[7] To deliver the digital file flawlessly, technical steganography transmits the voice signals produced by the steganography as well as the digital media data used as the cover through transport protocols like TCP or IP. On the other hand, network steganography techniques inject the hidden information data using Speech over IP (SoIP) fields and use the protocol fields as the cover speech.[8] In the aforementioned approaches, the application areas and specifications differ. The prerequisite for technological steganography, for instance, is to create the system to withstand attacks like having an un-detectable ability. However, the more significant characteristics of network technology in real-time applications are jitter, which is a change in the received signal's latency, and packet loss.[9] Audio steganography is the art of protecting data by hiding it in an audio file; the hidden data may be a picture, a text, a video, or even another audio; this method is commonly used to secure copyright for audio recordings in order to retain the rights of the file's owners. The greatest difficulty that all steganographic methods face comes from the human auditory system's (HAS) extreme sensitivity.[10,11] Audio steganography can be accomplished using a variety of methods, including LSB, phase coding, echo hiding, and spread spectrum coding. The most well-known and convenient technique is called LSB, in which the LSB of the cover value is used to substitute the bits of hidden data.[12] Data of any sizes may be concealed using the LSB steganography approach. Sadly, the channel noise these data produce makes them perceptible,[13] whereas transform domain strategies made use of the auditory system's ability to conceal by placing the low frequencies next to the high frequencies that are heard. One of the drawbacks of the LSB approach is that it is not a secure method. This is due to the fact that it is very simple to hack the secret message and extract it by any anyone who is not permitted to do so.[14] Creating a mechanism to secure information and assure its security during transmission is the main challenge. In audio steganography, an audio serves as the cover, and an informative file, a picture, or a noise might serve as the secret information.

In Ref. [15] authors described the Pixel Value Difference (PVD) based speech steganography method, in which the cover speech was separated into non-overlapping blocks of 2 × 2 size, and the difference value between two samples in each block was determined. A pixel's amount of alteration has been made based on the difference value and the input data sequence. An updated hidden Markov model was used by the authors[16] in to reduce hidden time, reduce computing costs, and improve security while increasing data retrieval. They created an authentication technique that combines data concealing, chaotic encryption,[17] and semantic segmentation to increase security, bandwidth efficiency, and resistance to steganalysis assaults. In Ref. [18] writers concentrated on dismantling YASS, another steganography technique. The authors noted that the embedding site choices made by the YASS method are not random and are thus simple to identify.[19] The assault was aided by the characteristics being extracted using the steganalysis observation domain (SO-domain). The SO-domain only partially accesses the concealed data, and the placement of the embedding determines how well the system can identify it. Authors have suggested a safe encryption-based solution in Ref. [20] for obfuscating sensitive information in cover speech files employing Parity and XOR in addition to encryption to increase degrees of security. This technique increases the amount of LSB bits required for embedding while keeping the perceptual quality of the output voice signal from steganography at a high level. The primary goal of steganography-based approaches is to maximize the volume of hidden data that may be conveyed to the precise receiver. In order to have a big cover file relative to the secret information size and demand a considerable amount of bandwidth, the transform domain approach[21] and the temporal domain method are required. As a result, several voice steganography methods are used.

The secret information is categorized using vector quantization,[22] and the concealed information in the receiver is then reconstructed using the same codebook. The method[23] uses the threshold to cut down on the amount of time needed for fractal steganography coding. The drawback of this plan is that both ends of the code books must be present. To lower the temporal complexity compared to the conventional approach, authors in Ref. [24] proposed using projected fractal coding with affine mapping voice steganography technique and block index descriptor that selects the best matching block. High embedding capacity without descriptors and adequate voice steganography quality are provided by another method.[25] The widely used simplistic technique for message concealing is called LSB-based embedding,[26] and it involves hiding the concealed message at the least important sections of cover speech samples. It appears to be an original speech because the only change in the cover sample is in the LSB, which would not significantly alter the stego object. The intensity

[6] *Universidad Tecnológica del, Lima 15046, Perú.*
[7] *Universidad de Huánuco, Huánuco 10001, Peru.*
[8] *Department of Computer Science & Engineering, Lovely Professional University, Phagwara 144001, Punjab, India.*
[#] *These authors contributed to this work equally.*
**Email: kumar.surendra1989@gmail.com* (S. Kumar)

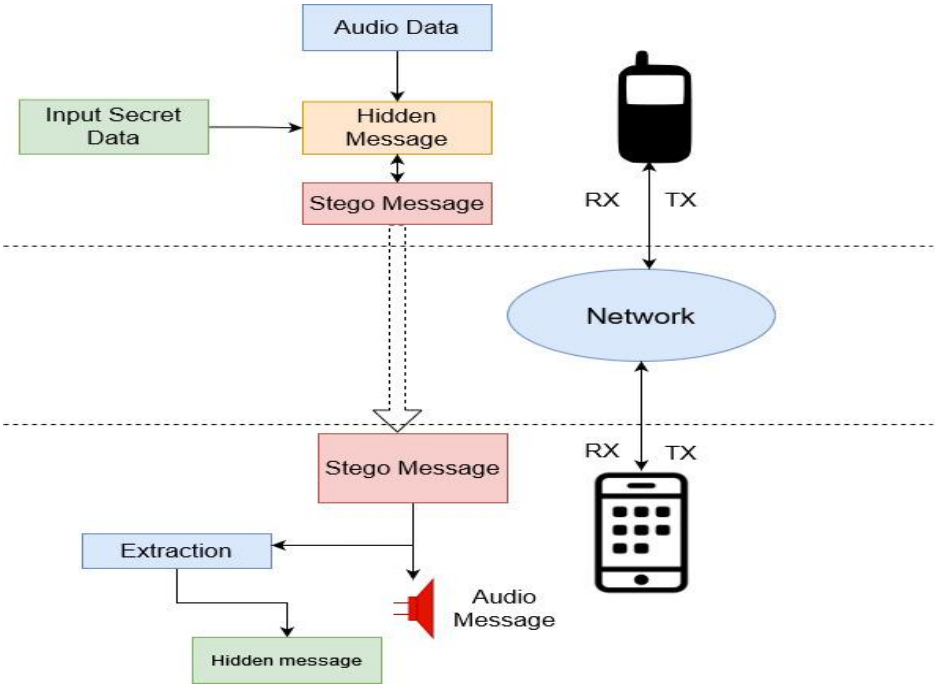

Fig. 1 Speech steganography system in Network communication.

components of the original cover object were split up, and each intensity speech frame underwent an embedding procedure. The stego object is created by combining the intensity speech frames after that. The PVD-based approaches[27] took advantage of the features of the human visual system, which is more sensitive to tiny changes in smooth regions than in edges. After comparing the values of a sample with those of its neighbors to determine the number of embedding bits in that sample, the method decides whether to improve the embedding capacity by using extra bits if the difference is significant.

The embedding samples in the sample mapping technique[28] are chosen using a mathematical function that is dependent on the sample intensity value of the seed sample and its neighbor samples. By applying an embedding process to both the seed sample and its neighbor, as well as a threshold value, the algorithm determines the boundary condition for the samples and their neighbors. The Fibonacci algorithm or other random generation techniques were used in the random sample approach[29] to choose the random sample sites where embedding happened. Texture-based embedding is a common spatial domain steganography[29] method. This method divides the cover speech and concealed speech into a number of subblocks of a given size. The ECG steganography methodology, which uses the tunable Q-factor wavelet transformation (TQWT) and SVD algorithms, provides a more secure and private way to encrypt patient data.[30,31] The fundamental difficulty in developing a steganographic system is maintaining a just balance between resilience, security, imperceptibility, and increased bit embedding rate.[32] A state-of-the-art coverless steganography that makes the most of the audio and frame picture features of the video. The first three characteristics that are extracted to produce hash bit sequences are the DWT (discrete wavelet transform) elements and instantaneous energy levels of audio, along with the SIFT (scale-invariant feature transformation) feature that characterizes frame pictures. Create a retrieval database based on the connections between the three characteristics of the related films and the produced bit sequences.[33] Using machine learning techniques, testing procedures can be automated, made more efficient, and the number of test cases needed can be cut in half. New metrics can also be introduced to measure the effectiveness of testing.[34] As a result of its increased redundancy and faster data transfer rates than analogue audio signals, digital audio signals are frequently employed for steganography. The LPC10, CELP, and MELP audio standards are used for audio and speech processing. They are strong, high-quality speech coding techniques that offer very precise estimations of audio characteristics and are frequently used in communications.[35] Hamiltonian cycles on triangular meshes are used to produce beautiful space loops that fill up 3D surfaces, whereas travelling salesman art is used to make artistic surface loops by mapping out picture shapes.[36] Most of the approaches have remain vulnerable, and we are working on developing a new hybrid steganography technique to address the security vulnerabilities[37] that have been brought to our attention. The CWT method utilizes a multi-scale analysis approach by first examining lower frequencies with a larger window size and then studying higher frequencies with a smaller window size. Because of this, the CWT will adaptively select a higher frequency resolution and a lower time resolution when it is processing lower frequencies, but it will select a frequency accuracy and a higher time resolution when it is processing higher frequencies.

The major contribution of the research are as follows:

- Consequently, the topic of this study is sophisticated speech steganography using SI-CWT. It is a technique for concealing data using pseudo noise patterns in low level speech bands that is also referred to as steganography. After that, the receiver reconstructs the original voice and message signals using a SAMD method based on ZCR.
- An extractor is a crucial component of a steganography technique that is used to recreate the secret audio from the stego-audio. This may significantly lessen feature loss and is made up of five resolution blocks that are constructed with a residual network structure to learn the acoustic characteristics of various frequency bands on the audio spectrum. The outcomes of the experiment indicate that this extractor may ensure the full transmission of a secret audio in both auditory and semantic aspects.
- The proposed method may practically withstand detection by steganalysis tools because to the similar distributions of the sample values between the stego-audio and actual audio. On the already-created or existing audio cover, it also doesn't perform any embedding actions.

Section 1 of the study, together with related work and introduction, is contributed. The suggested approach is covered in Section 2. The results and a discussion of the suggested method are covered in Section 3. The conclusion and foreseeable scope are covered in Section 4.

2. Methodology

2.1 Proposed methodology

The domains of the steganographic algorithm are spatial and transform. To lessen the distortion, the texture patterns from the cover speech are mapped with the secret speech using similarity measurement. In embedding techniques based on histograms, pairings of zero points, peak points, or points with a comparable intensity level are found in the cover speech's histogram, and fusion is then performed at the corresponding spots. Each instance of the input message was a speech signal that was approximately two minutes long, had a payload of more than two thousand bits, and was input. In SSIP, the encoder and decoder both made use of a similar pseudo-noise sequence generator and a key that was utilized to pinpoint the sample's location throughout the embedding process. Less mathematical complexity, a lower likelihood that cover speech will deteriorate, and a higher likelihood that more concealed information will be embedded are all benefits of spatial domain approaches. In contrast, the manipulation of the cover, the destruction of the hidden information by the eavesdropper, and the fact that the majority of the algorithms depend on speech cause loss in the hidden information in the spatial domain steganography. Thus, the implementation of transform domain-based voice steganography was taken into account in this study.

2.2 Steganography embedded system

The secret message is embedded in the cover object, which serves as a carrier, and once the embedding procedure is complete, the output object is known as the stego speech. The technique used to choose where alteration occurs inside the cover, the sorts of embedding operations, and the number of bits that must be updated in each sample of the cover have all had an impact on the steganographic security.

The block diagram of the transform domain steganography that has been suggested for embedding various kinds of confidential information is shown in Fig. 2. The cover speech was translated into frequency domain using the SI-CWT transform in the transform domain encoder, and any type of secret information was then inserted into the cover speech. Following embedding, the embedded signal, known as stego speech, is subjected to the inverse SI-CWT transform before being sent across the public channel. The stego speech has been subjected to the SAMD method in the decoder to obtain the secret data.

The protection of data from tampering or attacks by hackers or eavesdroppers on transmitting channels is referred to as robustness. Additionally, the terms imperceptibility and capacity refer to the quantity of data that may be concealed in a cover speech and the inability to tell the difference between the original cover speech and the hidden information.

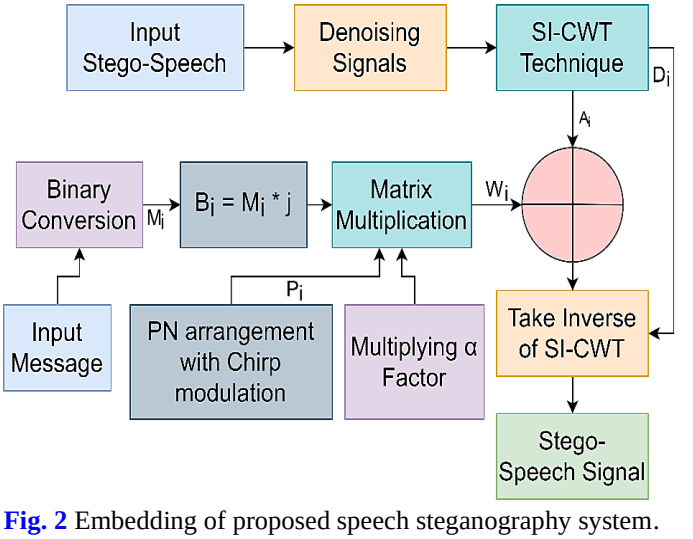

Fig. 2 Embedding of proposed speech steganography system.

2.3 Steganography extraction

A speech frame's speech content or the presence of background message is determined by the SAMD. The DTX/SCR codes the speech frame using either the ACELP encoder or the comfort messaging system depending on the decision made by the SAMD algorithm, which then sends the decision to the SAMD algorithm. 20 millisecond speech frames sampled at 12.8 kHz are used as the input to the SAMD. The signal is split into 12 subbands for each input speech frame, and the intensity of each subband is determined. For the identification of strongly periodic signals, like speech, a tone detection function based on the open-loop pitch gains obtained by voice encoder is used. The estimated speech and message levels for each speech frame are computed. While the message is estimated independently in each sub-band for speech, just

one guess is made for all bands. Next, an input SNR is determined using the estimated background message. An adaptive threshold that was based on message and speech estimations was used to evaluate the SNR in order to reach an intermediate conclusion. The final SAMD choice was made by adding a hangover to the intermediate decision so as not to code low pitch endings of speech segments as message.

Pitch computation and ZCR-based SAMD are the two components of this method. In classifying the message and speech frames in SAMD, the ZCR plays a crucial role. The ZCR uses the zero-crossing count (ZCC) as a conditional parameter to distinguish between speech and message signals. In order to produce ZCC, it was customary to keep track of how loudly each frequency sample in the speech spectrum was being presented. In contrast to message signals, which have high ZCC values due to the absence of pitch, speech signals have low ZCC values due to pitch presence. To recognize voice and message signals, Fig. 3 shows a detailed block layout of the SAMD of the ZCR.

Step 1: The voice signal is first applied to the end-to-end detection, identifying the upper and lower pitch levels as well as the frequency restrictions.

Step 2: To ensure accurate analysis, the speech signal is split up into several speech frames. In this case, the message data is handled as the minor speech frame, and the voice data is largely treated as the major speech frame.

Step 3: These speech frames are used to estimate the short time average zero crossing rate, effectively calculating the ZCC values.

Step 4: The ZCC-based ZCR values were then contrasted with the accepted threshold values; if the speech falls inside the threshold region, it is handled as a message signal; otherwise, it is treated as a speech signal. As indicated in Fig. 4, if a comparison is not made, it is seen as a scenario that is uncertain. The ZCC Distribution for messages and speech at various frequencies is explained.

Step 5: If there is a doubtful circumstance, the speech signal is once again separated into numerous sub-speech frames for in-depth examination.

Step 6: With the help of the hamming window, the sub speech frames are once again used to the calculation of the short time average zero crossing rate. In this case, the hamming window performed an accurate analysis of the data and calculated the message samples. Speech and message data are extracted through a series of iterations of the process.

The mathematical analysis of ZCR show in Fig. 4 and explained in Equation (1) as follow:

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]| w(n-m) \quad (1)$$

where

$$sgn[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x < 0 \end{cases} \quad (2)$$

Here, $x(m)$ is the present value of speech sample and $x(m-1)$ is the previous value of speech sample. Here, $w(n-m)$ is the hamming window coefficients shows in Equation (2) and Z_n is the ZCC count.

$$w(n) = \begin{cases} -\frac{1}{2N} for, 0 \leq n \leq N-1 \\ 0 \quad for, \ otherwise \end{cases} \quad (3)$$

Finally, $w(n)$ represents the extracted message data from using Equation (3).

3. Result and discussion

Examining the simulation results in great detail is the focus of this paper. To compare the effectiveness of the suggested strategy to other artificial intelligence methodologies now in use, a number of qualitative metrics are used.

3.1 Dataset

100 speakers divided across 5 training speech samples (100 speakers total) and 1 testing speech sample (100 speakers total). The dataset derives its information from the following sources: https://www.kaggle.com/datasets/piyushagni5/berlin-database-of-emotional-speech-emodb/code; and https://www.kaggle.com/datasets/kongaevans/speaker-recognition-dataset. A total of 600 speech snippets in different audio formats, including MPEG, MP4, and MP3, are included. Following pre-processing, the samples were added to the.wav

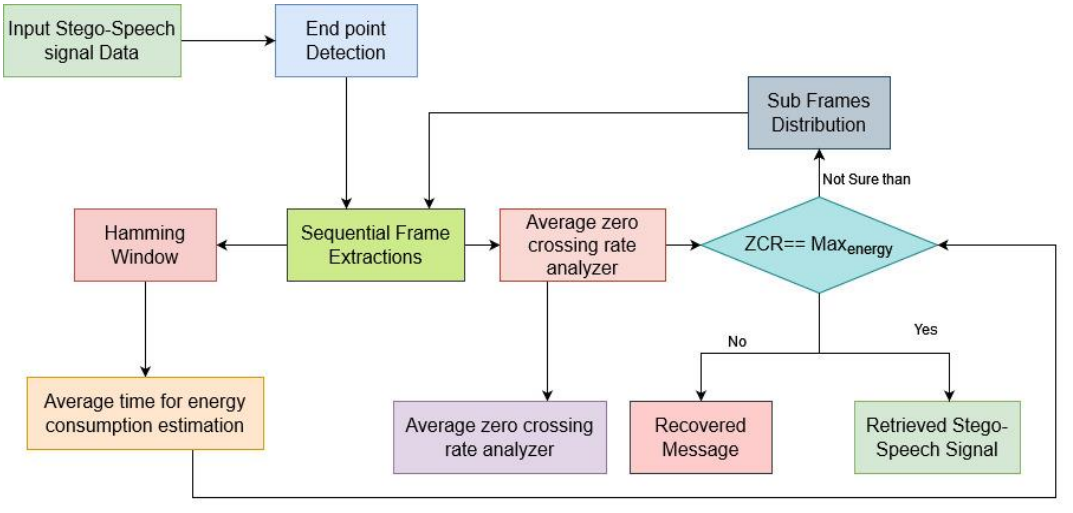

Fig. 3 SAMD based speech and message extraction system.

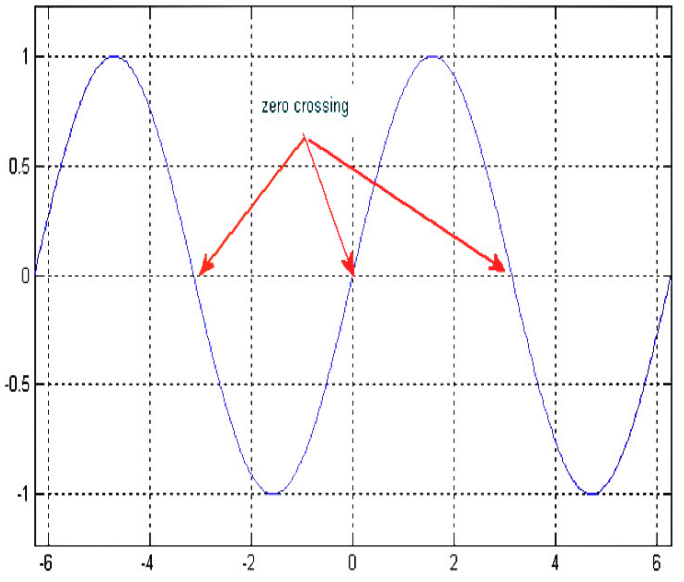

Fig. 4 Definition of zero-crossings rate.

file and made ready for use. It is vital to fine-tune the settings before utilizing them because each speech sample has a duration of 5 to 10 seconds. The total dataset has a length of 1 hour and 40 minutes and a size of 600 MB.

3.2 Subjective performance

On two separate speech signals, operations for message extraction and embedding are shown in Figs. 5 and 6. The message "anushasan" is stored in speech sample 1 and the message "MATHURA" is stored in voice sample 2. The final result is that both messages are flawlessly extracted and implanted, demonstrating increased imperceptibility performance. Additionally, there is no difference between the raw speech and the speech that was rebuilt, demonstrating increased robustness performance. Table 1 shows the Correlates of correlation and bit error rate (BER) values in comparison to noise attack.

Table 1. Correlates of correlation and BER values in comparison to noise attack.

Method	BER		CC
	Exclude noise	With noise	
FFT-based Approach[38]	0.000512	4.23	0.917324
DWT-based technique [20]	0.000001	0.00156	0.97243
SI-CWT with SAMD	0.00	0.00053	0.99998

Two sections make up the performance review. In the first part of the study of the confusion matrix, the rate of false negatives (RFN), the rate of false positives (RFP), the rate of true positives (RTP), as well as the rate of true negative (RTN) are all investigated. The confusion matrix was utilized in the training and testing of the datasets, as well as the measurement of the detection accuracy. Standard datasets were utilized in the experiment designed to discover anomalies. The confusion matrix is examined using the calculated values of RTP from Equation (4), RFN from Equation (7), RTN from Equation (5) and RFP from Equation (6). The false positive rate (RFP) is calculated by dividing the entire number of negative cases by the total quantity of false positives (FP) that were not correctly categorized as negative instances. The ratio of positive instances that were incorrectly labelled as false negatives (FN) to all positive cases is used to determine the false negative rate, or RFN. The number of negative instances that were mistakenly stated as true positives (TP) is divided by the total number of negative cases to arrive at the RTP. The ratio of positive instances that were mistakenly classified as true negatives (TN) to all positive cases is used to compute the true negative rate, or RTN and comparative results are shown is Table 2 and Fig. 8 presented an accuracy performance for confusion matrix. the time-domain waveforms of the original secret audio files, while the right column displays the reconstructed secret audio files presented in Fig. 7.

$$\text{Rate of True Positives} = \frac{\text{True Positive}}{\text{Ture Positive+False Negative}} \quad (4)$$

$$\text{Rate of True Negative} = \frac{True\ Negative}{False\ Positive+True\ Negative} \quad (5)$$

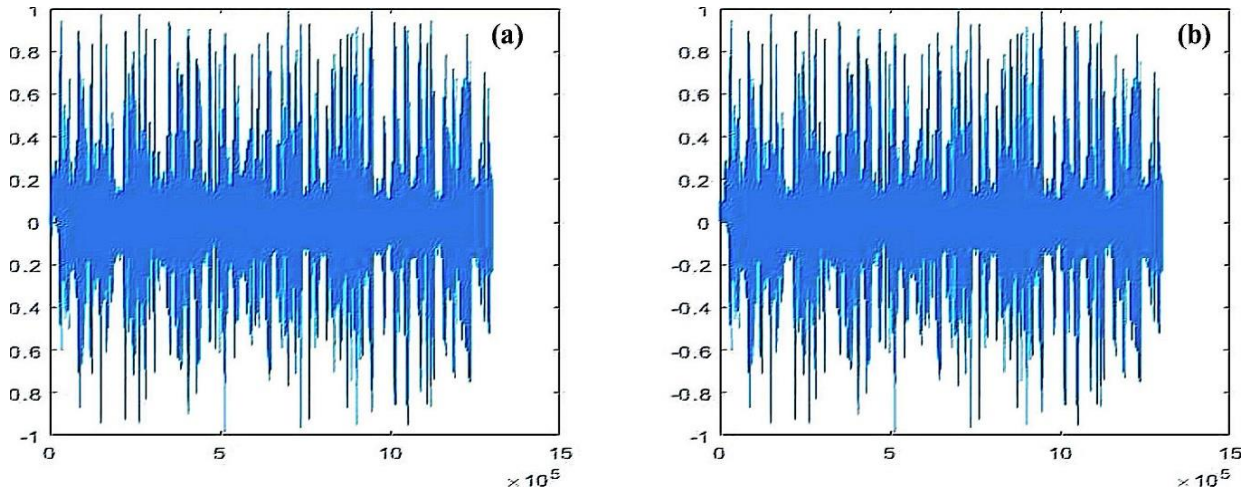

Fig. 5 (a) shows original audio and (b) shows the stego file with secret message.

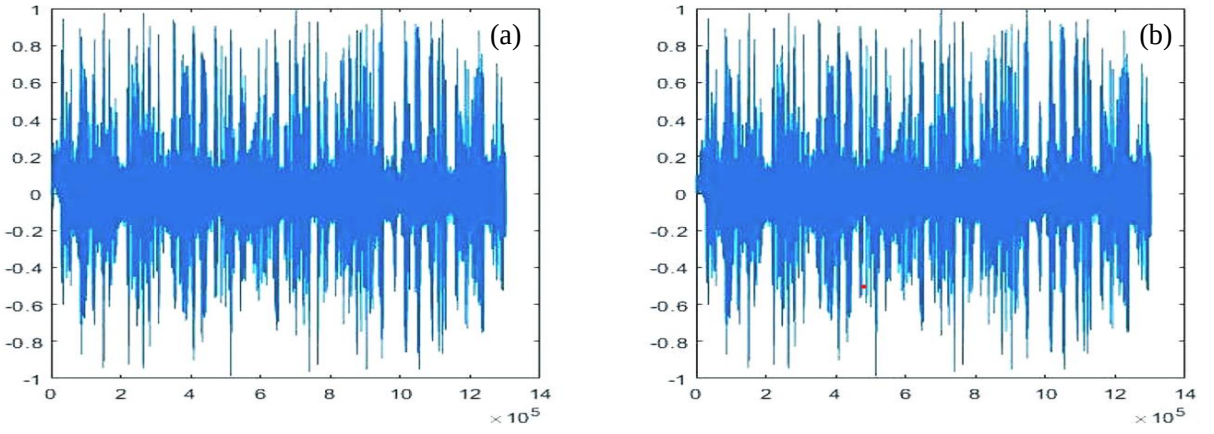

Fig. 6 (a) shows original audio and (b) shows the stego file with secret message.

$$\text{Rate of False Positives} = \frac{False\ Positive}{False\ Positive + True\ Negative} \quad (6)$$

$$\text{Rate of False Negatives} = \frac{\text{False Negative}}{\text{False Negative+True Positive}} \quad (7)$$

Table 2. Comparison of the various accuracy parameters.

ER%	TP	TN	FN	FP	TPR	TNR	TAR	FNR
10%	54	72	21	3	71.62	98.11	86.75	29.02
20%	59	72	15	3	82.56	98.11	89.99	18.74
50%	66	70	6	3	92.06	98.11	96.14	7.04
80%	68	70	4	3	96.83	98.11	97.53	4.77
100%	70	70	3	3	98.11	98.11	98.11	3.12

3.3 Objective performance

The effectiveness of the suggested method is contrasted in Table 3 with that of more established techniques like DWT,[16,39] SWT,[18] and RDWT.[20] Based on the simulations, it can be seen that the suggested speech steganography method outperformed all other methods for all measures presented in Fig. 9.

Table 3. Performance comparison.

	DWT[16,39]	SWT[18]	RDWT[20]	Proposed
Entropy	9.8260	9.9756	11.314	12.034
MI	1.7546	1.8687	2.39837	2.5142
PSNR	31.140	35.067	36.3268	58.231
SSIM	1.0730	1.1087	1.13664	1.2643
STD	0.0581	0.5141	0.04428	0.0258
RMSE	0.02009	0.2608	0.01918	0.0004

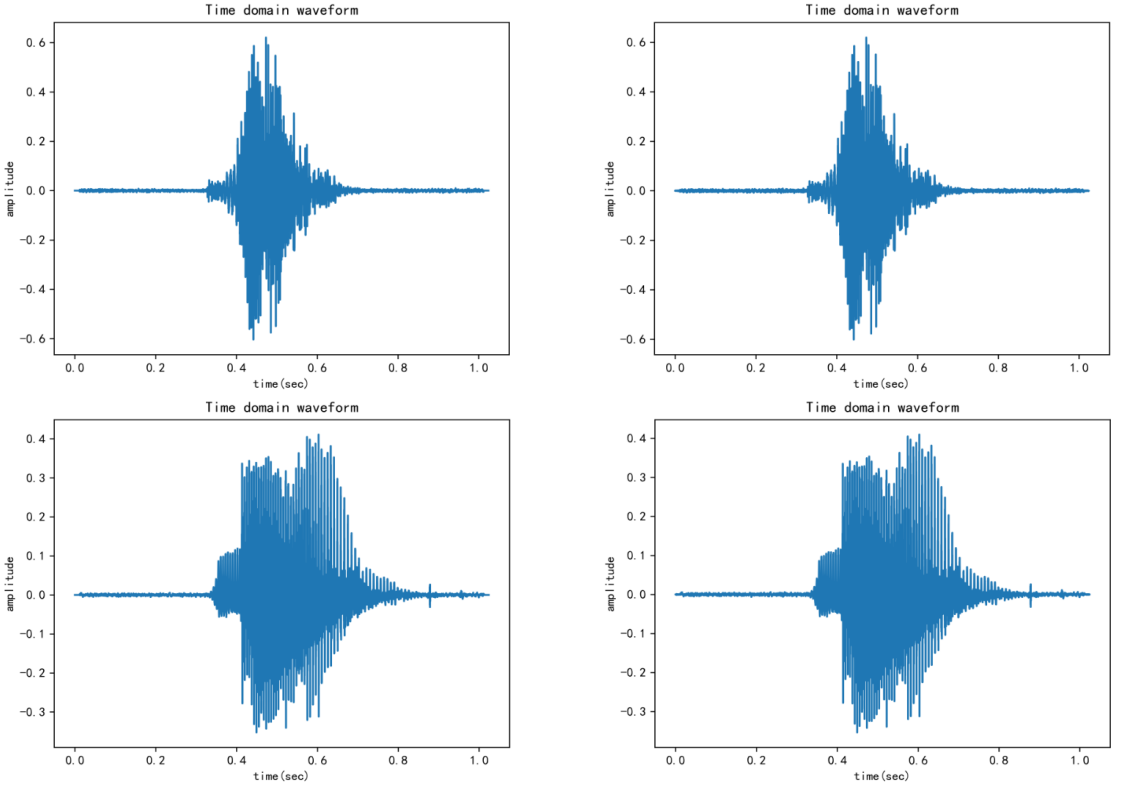

Fig. 7 Secret audio files' time-domain waveforms are displayed in the left column, while those of the reconstructed secret audio files are displayed in the right column.

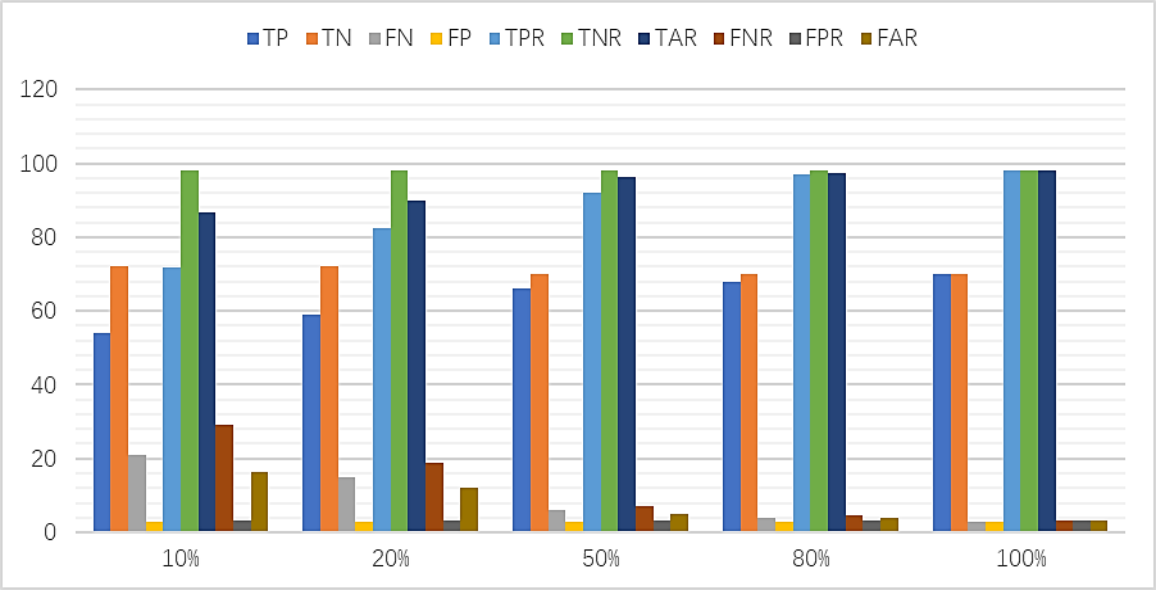

Fig. 8 Accuracy performance comparison.

The ambient sound impact is the main issue, and it needs to be decreased and optimized in order to increase the effectiveness of the speech steganography system. As a result, the effectiveness of the proposed audio steganography system is assessed in terms of the hidden message that can be gleaned from the stego speech through the inclusion of random impulse noise. Although the recovered secret messages utilizing the DWT-based approach[40] have a higher mistake rate, the ZCR-based voice steganography-based communications do contain noise attacks. Comparatively, a proposed SI-CWT with SAMD technique recovered encrypted data with zero error rates. In addition, bit error rate and correlation coefficient for both existing and proposed hybrid voice steganography systems are computed and displayed in Table 1 for comparison. The main limitation of the proposed method is to assess the algorithm's accuracy, various data kinds can be hidden from view in each frame, along with various attacks or noises.

The designed technique includes both a case for speech steganography efficiency alongside and without pause removal. CPU running time is also measured to evaluate the performance of the proposed voice steganography system based on pause elimination. Additionally, the suggested SI-CWT with SAMD methodology was tested for resistance to noise attack and outperformed other voice steganography methods on the basis of both BER and CC parameters. Last but not least, thorough simulation research showed that the suggested voice steganography delivers the best results when compared to the traditional methodologies. The efficiency of the designed SI-CWT, in which the cover speech, stego speech, and reconstructed speech are all of same size and look very similar to one another, which results in higher imperceptibility

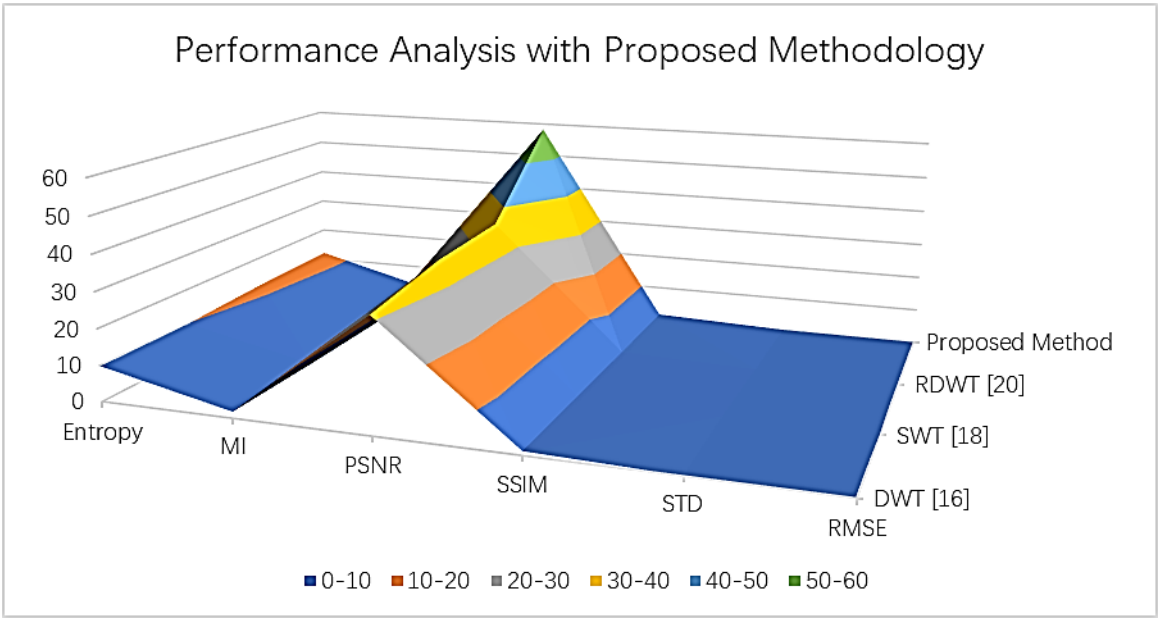

Fig. 9 Proposed Speech steganography mechanism performance comparison.

and robustness features as compared to existing speech steganography systems.

4. Conclusions

This article is focused on the construction of an advanced speech steganography system that makes use of the SI-CWT. The message data is then concealed into low-level speech bands with the use of pseudo noise sequences, which results in the generation of the steganography output speech signal as a result. Once this is accomplished, a ZCR based SAMD method is implemented on the receiver side in order to recover both the original speech and the message signal. Exceptional reliability as well as safety were demonstrated by the recommended method's covert communication. Experimental and theoretical analyses show that the offered method not only guarantees the full semantic communication of secret audio even with the face of distortion, but also offers exceptional security and undetectability. Future work on this topic could be expanded to include the implementation of high secured vocoders for military applications can be implemented. Offering security-based voice conversion is a vocoder's fundamental feature. In order to effectively incorporate the proposed method in speech steganography while protecting against additional signal processing attacks like MP3 compression, amplitude scaling, cropping, synchronization attacks, *etc*., it is effective for removing pauses from speech. This approach is based on deep learning convolutional neural networks.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] R. C. Rao, P. V. Y. Jayasree, S. S. Rao, G. S. Yeshwanth, K. R. S. Megana, K. Shreya, & Suprasen, Basic Framework of Different Steganography Techniques for Security Applications, *International Conference on Soft Computing and Signal Processing*, Springer, 2021, doi: 10.1007/978-981-16-7088-6_52.

[2] H. Dutta, R. K. Das, S. Nandi, S. R. M. Prasanna, An overview of digital audio steganography, *IETE Technical Review*, 2020, **37**, 632-650, doi: 10.1080/02564602.2019.1699454.

[3] A. Ahmed, AlSabhany, Digital audio steganography: systematic review, classification, and analysis of the current state of the art, *Computer Science Review*, 2020, **38**, 100316, doi: 10.1016/j.cosrev.2020.100316.

[4] A. Kanhe, G. Aghila, A DCT-SVD-based speech steganography in voiced frames, *Circuits, Systems, and Signal Processing*, 2018, **37**, 5049-5068, doi: 10.1007/s00034-018-0805-9.

[5] F. Kreuk, Y. Adi, B. Raj, R. Singh, J. Keshet, Hide and speak: towards deep neural networks for speech steganography, *proceedings of Interspeech*, 2020, 4656-4660. Doi: 10.21437/Interspeech.2020-2380

[6] Y. Qiu, H. Tian, H. Li, C.-C. Chang, A. V. Vasilakos, Separable convolution network with dual-stream pyramid enhanced strategy for speech steganalysis, *IEEE Transactions on Information Forensics and Security*, 2023, **18**, 2737-2750, doi: 10.1109/TIFS.2023.3269640.

[7] J. Wen, H. Zeng, Y. Wang, S. Liu, Y. Xue, An SVD-based adaptive robust speech steganography using MDCT coefficient, *Multimedia Tools and Applications*, 2021, **80**, 2517-2536, doi: 10.1007/s11042-020-09725-5.

[8] H. Kheddar, D. Megías, High-capacity speech steganography for the G723.1 coder based on quantised line spectral pairs interpolation and CNN auto-encoding, *Applied Intelligence*, 2022, **52**, 9441-9459, doi: 10.1007/s10489-021-02938-7.

[9] N. Amiri, I. Naderi, DWT-GBT-SVD-based robust speech steganography, 2020, https://arxiv.org/abs/2004.12569".

[10] C. Teck Jian, C. C. Wen, N. H. Binti Ab Rahman, I. R. Binti A Hamid, Audio steganography with embedded text, *IOP Conference Series: Materials Science and Engineering*, 2017, **226**, 012084, doi: 10.1088/1757-899x/226/1/012084.

[11] K. Thangadurai, G. Sudha Devi, An analysis of LSB based image steganography techniques, *2014 International Conference on Computer Communication and Informatics*. Coimbatore, India, IEEE, 2014, 1-4, doi: 10.1109/ICCCI.2014.6921751.

[12] M. Pooyan, A. Delforouzi, LSB-based audio steganography method based on lifting wavelet transform, *2007 IEEE International Symposium on Signal Processing and Information Technology*, Giza, Egypt. IEEE, 2008, 600-603, doi: 10.1109/ISSPIT.2007.4458198.

[13] Shunzhi, Jiang, SmartSteganogaphy: Light-weight generative audio steganography model for smart embedding application, *Journal of Network and Computer Applications*, 2020, **165**, 102689, doi: 10.1016/j.jnca.2020.102689.

[14] J. Chaharlang, M. Mosleh, S. Rasouli Heikalabad, A novel quantum audio steganography-steganalysis approach using LSFQ-based embedding and QKNN-based classifier, *Circuits, Systems, and Signal Processing*, 2020, **39**, 3925-3957, doi: 10.1007/s00034-020-01345-6.

[15] M. Baziyad, I. Shahin, T. Rabie, A. Bou Nassif, Maximizing embedding capacity for speech steganography: a segment-growing approach, *Multimedia Tools and Applications*, 2021, **80**, 24469-24490, doi: 10.1007/s11042-020-10228-6.

[16] S. Agarwal, S. Venkatraman, Deep residual neural networks for image in speech steganography, 2020: arXiv: 2003.13217. https://arxiv.org/abs/2003.13217.

[17] W. Yang, S. Tang, M. Li, B. Zhou, Y. Jiang, Markov bidirectional transfer matrix for detecting LSB speech steganography with low embedding rates, *Multimedia Tools and Applications*, 2018, **77**, 17937-17952, doi: 10.1007/s11042-017-5505-0.

[18] R. C. Rao, P. V. Y. Jayasree, S. S. Rao, Hybrid Speech Steganography System using SS-RDWT with IPDP-MLE approach, *Soft Computing*, 2023, **27**, 1117-1129, doi: 10.1007/s00500-021-05970-4.

[19] Y. Xue, K. Mu, Y. Wang, Y. Chen, P. Zhong, J. Wen, Robust

speech steganography using differential SVD, *IEEE Access*, 2019, **7**, 153724-153733, doi: 10.1109/ACCESS.2019.2948946.
[20] P. M. Kumar, K. Srinivas, Real time implementation of speech steganography, *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India. IEEE, 2020, 365-369, doi: 10.1109/ICSSIT46314.2019.8987785.
[21] R. Shrivastava, M. Singh, K.S.S.R. Teja, A Real-time Implementation for the Speech Steganography using Short-Time Fourier Transform for Secured Mobile Communication. *Journal of Physics: Conference Series IOP Publishing*, 2021, **2089**, 012066, doi:10.1088/1742-6596/2089/1/012066/meta.
[22] C. Gong, X. Yi, X. Zhao, Pitch delay based adaptive steganography for AMR speech stream. *Digital Forensics and Watermarking. Cham: Springer International Publishing*, 2019, 275-289, doi: 10.1007/978-3-030-11389-6_21.
[23] R. C. Rao, P. Jayasree, S. S. Rao, P. M. Kumar, Spread spectrum based speech steganography using RDWT, *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India. IEEE, 2021, 927-933, doi: 10.1109/ICESC51422.2021.9532839.
[24] Yuting, Hu, Detection of heterogeneous parallel steganography for low bit-rate VoIP speech streams, *Neurocomputing*, 2021, **419**, 70-79, doi: 10.1016/j.neucom.2020.08.002.
[25] J. He, J. Chen, S. Xiao, X. Huang, S. Tang, A novel AMR-WB speech steganography based on diameter-neighbor codebook partition, *Security and Communication Networks*, 2018, **2018**, 7080673, doi: 10.1155/2018/7080673.
[26] B. Gupta Banik, S. K. Bandyopadhyay, Novel text steganography using natural language processing and part-of-speech tagging, *IETE Journal of Research*, 2020, **66**, 384-395, doi: 10.1080/03772063.2018.1491807.
[27] A. H. Khaleel, I. Q. Abduljaleel, Secure image hiding in speech signal by steganography-mining and encryption, *Indonesian Journal of Electrical Engineering and Computer Science*, 2021, **21**, 1692, doi: 10.11591/ijeecs.v21.i3.pp1692-1703.
[28] A. M. K. Reddy, Optimized multirate wideband speech steganography for improving embedding capacity compared with neighbor-index-division codebook division algorithm, *Revista Gestão Inovação e Tecnologias*, 2021, **11**, 1362-1376, doi: 10.47059/revistageintec.v11i2.1763.
[29] A. S. Hameed, A high secure speech transmission using audio steganography and duffing oscillator, *Wireless Personal Communications*, 2021, **120**, 499-513, doi: 10.1007/s11277-021-08470-8.
[30] P. Mathivanan, A. Balaji Ganesh, ECG steganography based on tunable Q-factor wavelet transform and singular value decomposition, *International Journal of Imaging Systems and Technology*, 2021, **31**, 270-287, doi: 10.1002/ima.22477.
[31] A. Kumar, A novel privacy preserving HMAC algorithm based on homomorphic encryption and auditing for cloud, *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC*), Palladam, India, IEEE, 2020, 198-202, doi: 10.1109/I-SMAC49090.2020.9243340.
[32] I. Jawad, Kadhim, Comprehensive survey of image steganography: techniques, Evaluations, and trends in future research, *Neurocomputing*, 2019, **335**, 299-326, doi: 10.1016/j.neucom.2018.06.075.
[33] C. Zhang, Y. Tan, J. Qin, X. Xiang, Coverless video steganography based on audio and frame features, *Security and Communication Networks*, 2022, **2022**, 1-14, doi: 10.1155/2022/1154098.
[34] S. Kumar, Reviewing software testing models and optimization techniques: an analysis of efficiency and advancement needs, *Journal of Computers, Mechanical and Management*, 2023, **2**, 32-46, doi: 10.57159/gadl.jcmm.2.1.23041.
[35] S. Talati, P. EtezadiFar, M. R. H. Ahangar, M. Molazade, Investigation of Steganography Methods in Audio Standard Coders: LPC, CELP, MELP. Majlesi, *Journal of Telecommunication Devices*, 2023, **12**, 7-15, doi: 10.30486/mjtd.2022.695928.
[36] C. J. Tralie, Artistic curve steganography carried by musical audio. *Artificial Intelligence in Music, Sound, Art and Design. Cham: Springer Nature* Switzerland, 2023, 276-291, doi: 10.1007/978-3-031-29956-8_18.
[37] A. Kumar, V. Jain, A. Yadav, A new approach for security in cloud data storage for IOT applications using hybrid cryptography technique. *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)*, Mathura, India. IEEE, 2020, 514-517, doi: 10.1109/PARC49193.2020.236666.
[38] A. S. Patil, G. Sundari, An embedding of secret message in audio signal. 2018 *3rd International Conference for Convergence in Technology (I2CT)*. April 6-8, 2018, Pune, India. IEEE, 2018, 1-3, doi: 10.1109/I2CT.2018.8529549.
[39] S. Chakraborty, A. S. Jalal, A novel local binary pattern based blind feature image steganography, *Multimedia Tools and Applications*, 2020, **79**, 19561-19574, doi: 10.1007/s11042-020-08828-3.
[40] N. Bansal, V. K. Deolia, A. Bansal, P. Pathak, Comparative analysis of LSB, DCT and DWT for Digital Watermarking. *2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, 2015, 40-45, doi: 10.1109/EESCO.2015.7253657.

