Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.034270
Article

Tech Science Press

Comparative Analysis of Execution of CNN-Based Sanguine Data Transmission with LSB-SS and PVD-SS

Alaknanda S. Patil[1,*], G. Sundari[1] and Arun Kumar Sivaraman[2]

[1]Department of Electronics and Communication, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, 600119, India
[2]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600127, India
*Corresponding Author: Alaknanda S. Patil. Email: patilalaknanda@yahoo.com
Received: 12 July 2022; Accepted: 03 November 2022

Abstract: The intact data transmission to the authentic user is becoming crucial at every moment in the current era. Steganography; is a technique for concealing the hidden message in any cover media such as image, video; and audio to increase the protection of data. The resilience and imperceptibility are improved by choosing an appropriate embedding position. This paper gives a novel system to immerse the secret information in different videos with different methods. An audio and video steganography with novel amalgamations are implemented to immerse the confidential auditory information and the authentic user's face image. A hidden message is first included in the audio from the multimedia file; using LSB Technique. The Stego-video is created in the second stage by merging the authorized user's face into the frame of the video; by using PVD technology. Stego-audio is linked again with the stego-video in the third stage. The incorporated perspective techniques (LSB-SS and PVD-SS algorithms) with more significant data immersing capacity, good robustness and imperceptibility are proposed in this research work. The spread spectrum approach is used to increase the complexity of secret data recognition. Two different video files are tested with different voice files with the results such as PSNR, SSIM, RMSE and MSE as 52.3, 0.9963, 0.0024 and 0.0000059, respectively.

Keywords: Audio steganography; data hiding; information security; pixel value differencing

1 Introduction

In any application, the data being safeguarded is on top priority. So the research work on data preservation is majorly prompted. In the current era, the abundant usage of the internet and visual data transmission requires proper authentication to preserve & communicate confidential data by immersing it in the cover media, named steganography. The effective medium of hidden communication is achieved by steganography. An interest is renewed because industries want to safeguard trademarked digital music, image, video, and text. Steganography avoids the detection of counterfeiters and unauthorized individuals by embedding of a key, writer ID or biometric parameter. Steganography gives secret communication by

using the carrier of any media and payload of any media as per the requirement of the data to be preserved [1–5]. Imperceptibility, immersing capacity and robustness are the essential and crucial characteristics of a data-concealing system [6–14]. Depending upon the cover media, steganography can be elaborated in various methods, such as image steganography, video steganography, audio steganography and text steganography. Image steganography is segregated into two methods such as methods based on the spatial domains and frequency domains. In the First Approach, the confidential data is immersed directly into the pixel's intensity, in contrast, in the second technique, the pictures are first converted into the frequency domain, and then the confidential data is immersed into the coefficients. Video steganography receives lots of attention due to the vast usage of the internet & multimedia, since it has great spatial and temporal redundancy to immerse the confidential data. Multimedia-files can include much more information, than our Human Visual System (HVS) can see; thus, they can express more information than words [15–18].

Audio Steganography is developed chiefly for immersing confidential data regarding copyright and content integrity assurance. Data immersion can be done using a various methodologies, such as, Least Significant Bit (LSB), phase coding, parity coding, echo hiding, spread spectrum, and tone insertion [19]. A optimized novel method is proposed to get better results with robustness, imperceptibility and better immersing capacity [20].

The pixel-value differencing approach has the advantage of sending a large amount of data and retaining the image consistency after data embedding. The pixel-value differencing (PVD) scheme determines the number of secret bits to be inserted by comparing the values of two successive pixels in a block [21–24]. Neural networks reflect the behavior of the human brain, which allows computer programs to recognize patterns and solve common problems in the fields of AI, machine learning, and deep learning [25–29].

Different types of Neural Networks in Deep Learning

- Artificial Neural Networks (ANN)
- Convolution Neural Networks (CNN)
- Recurrent Neural Networks (RNN)

The varieties of CNN Architectures in Fig. 1 are LeNet, AlexNet, VGG, Google Net, ResNet and more. AlexNet is the first famous convolutional neural network (CNN) [30]. The structure of this document is as follows: Section 2 gives the content of the existing development, Section 3 provides the prospective system, Section 4 gives results, and Section 5 focuses on conclusions.

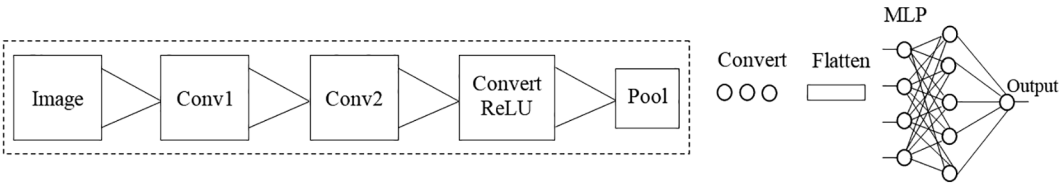

Figure 1: CNN structure

2 Existing Development

It is a crucial task to define the confidential message and immerse it in the cover media with good perceptibility. The varieties of video steganography with effective implementation are LSB replacement, Pixel Value Differencing and bit inversion [31–34]. Some studies related to the future research work are elaborated here.

The block-wise technique extends the pixel-value differencing approach. It increases the complexity because of four-pixel processing at a time [35]. Researchers proposed the optimum pixel adjustment process (OPAP) as a new data concealing approach that uses diamond encoding (DE) and PVD to adaptively embed data into the pixel pairs adaptively. But it does give lower immerging efficiency [36]. Researchers have recently worked with DCT and DWT coefficients to hide data. The distortions in the videos are reduced by a great amount, and the video quality levels are maintained. But the process adds complexity [37]. Recent research gives the Multi Curve and Elliptic Curve Cryptography technology for encrypting the confidential information. It is possible to select pixels from the frames using an optimization algorithm called Artificial Bee Colony while integrating the encrypted private details into the video; otherwise, stego-video distortion rises [38]. It's a difficult task to attain a high level of security in video steganography with avoiding intra-frame distortion drift [39–43].

3 Prospective Systems

The prospective system incorporates in mainly two modules–Face identification system & Trans-receiver system. Both modules are elaborated here.

3.1 Face Identification System

Fig. 2 shows an extensive diagram with the interconnections between steps of the Face identification system.

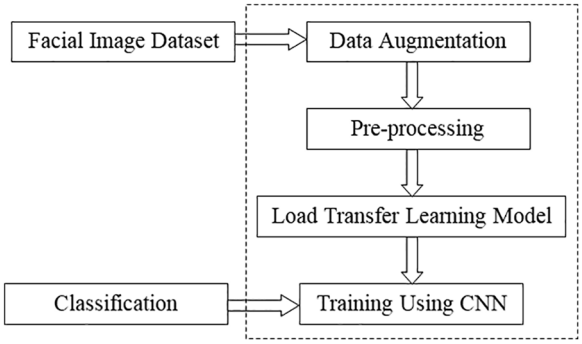

Figure 2: Face identification system

The main steps are the acquisition of images, database construction, face identification, pre-processing, augmentation of data, feature extraction, CNN training, and classification. The two primary phases are the training phase and the testing (identification) phase.

3.2 Capturing of Image

Real-time data is collected in the database below. The collection contains five distinct perspectives of the face pictures of five peoples in various light and brightness situations. The images are in the size of 227 × 227 pixels and are in RGB format. Table 1 shows the distribution of training and testing data in the system's database.

3.3 Preprocessing

Noise reduction, face cropping, scaling, thresholding, equalization of a histogram for removing illumination variation, converting to a binary or grayscale image, and so on are the things that could be done with recorded facial images. The RGB color standard is used for the input picture. The RGB picture

is then transformed into a grayscale image for further processing. This might be accomplished by averaging the three-channel signals; however, this approach fails since red has the most wavelengths of the three hues, whereas green has a shorter wavelength than red and provides a calming effect to the eyes. As a result, we see the need to lower the effect of the color red, enhance the impact of color green and place the influence of blue color somewhere between the two. The formula for converting RGB to grayscale is as follows:

$$0.30 * R + 0.69 * G + 0.11 * B \quad (1)$$

where, R, G, and B represent the values of red, green, and blue pixel intensity, respectively.

Table 1: Database distribution system

Data	Total facial images	Training facial images	Testing facial images
1	973	779	194
2	830	664	166
3	924	740	184
4	842	674	168
5	1453	1089	364

3.4 Augmentation of Data

Augmentation of data is the process of the artificial production of new training data from previous existing training data. This is accomplished using domain-specific techniques to build a new and dissimilar training example from instances in the training data. Image data augmentation is one of the most used types of data augmentation, and it entails creating refurbished versions of photos within the training dataset for the same picture class as the source image. Zooms, flips, shifts, etc., are the other picture modification operations that are included in the transformation.

The goal is to make the training dataset bigger by adding additional instances. As a result, the models will likely notice changes in the training set photos.

3.5 CNN Training and Testing

CNN's have demonstrated that they are capable of recognizing pictures as well as their categorization. CNN's are a type of multi-layer feed-forward neural network. CNN is made up of neurons, kernels and filters, all of which contain the weight, specification and bias. Every filter conducts convolution on a set of inputs. Convolutional, Rectified Linear Units (ReLU), pooling, and Fully Connected Layers are all part of CNN's structure (FCL).

- Layer of Convolutional

The layer of convolutional is the core building part of CNN, and it is responsible for the majority of the computational effort. The convolution layer's primary function is to extract specifications from the incoming data image. A collection of learnable neurons is used to convolve the input picture. In the output picture, this creates a feature map which is then supplied to the next convolutional layer, as input data.

- Pooling Layer

This layer minimizes the complexity of every activation map, while retaining the predominant data. The input photos are separated into a group of rectangles, which do not overlay. With a non-linear performance like average or maximum, each region is down-sampled. This layer generally placed within the convolutional

layers, achieves faster convergence and higher generalization with is less susceptible to relocation and modification.

- Layer of ReLU

This operation is non-linear, that incorporates rectifier-based units. It's applied per pixel and reset the feature map's values from negative to zero. To recognize how the ReLU works, let's suppose the input is 'x', and that the rectifier is f(x) = max (0, x) in the neural networks' literature. The FCL aims to utilize these characteristics to categorize the input picture in different groups, depending on the training dataset. The final pooling layer, FCL is responsible for putting the information into a classifier that employs the activation function of Softmax. The total of all Connected Layer's output possibilities is 1. This is supported by the use of Softmax in the form of an activation function.

3.6 CNN Training and Testing

AlexNet is a major deep network, which is utilized in a variety of computer vision applications. For face recognition, this method uses the trained CNN model's transfer learning called AlexNet. Fig. 3 depicts the AlexNet model architecture.

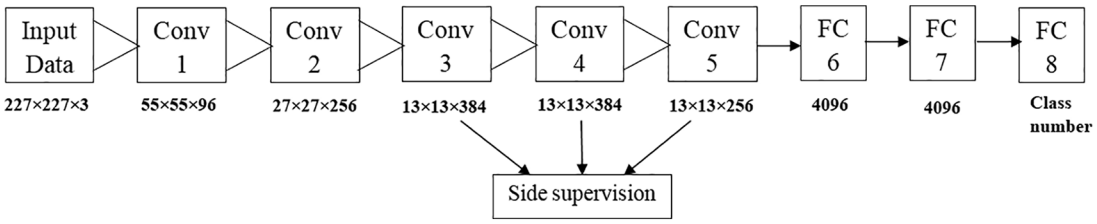

Figure 3: Architecture of AlexNet

AlexNet has three layers of fully connected and five layers of convolution. The image's key characteristics are extracted by these convolutional layers. The linear convolution filters in each convolutional layer are followed by normalization, ReLu activation and max pooling. The first input layer accepts pictures with dimensions of 227-by-227-by-3. Ninety-six filters are included in the first convolution layer, each measuring 11 × 11 × 3 with a speed of 4 and no padding. The ReLu layer receives outputs of the first convolutional layer. Then it goes to the max-pooling layer. The ReLu activation function is used to prevent the propagating of any number which is not positive across the network. The purpose of the Pooling Layer is computation and Control Over fitting Reduction. The second convolutional layer consists of 256 filters of 5 × 5 with pace one and padding two. 3 × 3 convolution with pace one and padding one is performed by the third to fifth convolution layers. Max-pooling is only present in convolutional layers 1, 2, and 5. Three fully linked layers follow the convolutional layers and down-sampling. The classification task is carried out by the last fully linked layer, which incorporates characteristics learned from the previous layer. A softmax layer follows this one, which will normalize the output.

3.7 LSTM

Organizations gain a competitive edge by early predicting the outcome of ongoing or completed operations. On this type of classification task, the performance of standard machine learning and more recently, the deep learning approaches like Long Short-Term Memory (LSTM), this has been thoroughly examined. Much recent research has been on using Convolutional Neural Networks (CNN) to solve time series challenges such as classification, but not on outcome prediction. The purpose of this paper is to close this gap and compare CNNs to LSTMs. Attention is another strategy used in time series

classification in conjunction with LSTMs and is included in this study. Our findings reveal that, given the necessary big datasets, all of the neural networks attain an adequate high prediction power. CNN's perform on par with LSTMs; thus, the Attention mechanism adds no value to the latter. CNN's are preferred because; they are one order of magnitude faster than both types of LSTM. All models are hyperparameter robust, and they acquire their maximum predictive power early in these cases, as well as only after a few events, which makes them ideal for runtime forecasts of the speed, early predictive capacity and resilience of CNNs. While it should pave the way for their use in process outcome prediction.

3.8 Trans-Receiver System

The multimedia file (.avi) is the input for this system. Firstly, the audio of the video multimedia file is retrieved. Pseudorandom Noise (PN) sequence is combined with personal audio, referred to as Spread Spectrum (SS) technique. Then the spreaded confidential signal is immersed with the original extracted voice file using the LSB method, named stego-audio file [40]. The chip rate of this spreaded code is more excellent and results in a wideband time-continuous scrambled signal.

The video (group of image frames) is also extracted from the original video, and then the authorized user face image is immersed in generating the stego-video. The immersing is done by two different algorithms in the proposed work, which is elaborated in detail below.

3.8.1 CASE 1-LSB Innovation

The LSB method is less complicated and more commonly utilized. It comes up with the optimal place in the retrieved video frames for hiding the face picture with a minimal distortion in the stego-video file. As a result, there is less distinction between stego-video and the original video.

C(i, j) will stay unaltered if the cover video bit C(i, j) is identical to m. The message bit of the face picture needs to be inserted; otherwise, set C(i, j) to m. In Eq. (2), the message immersion processing is as complex as it gets.

$$\begin{aligned} &S(i,\ j) = C(i,\ j) - \ 1, \quad \text{for LSB } (C(i,\ j)) = 1 \text{ and } m = 0 \\ &S(i,\ j) = C(i,\ j), \quad \text{for LSB } (C(i,\ j)) = m \\ &S(i,\ j) = C(i,\ j) + 1, \quad \text{for LSB } (C(i,\ j)) = 0 \text{ and } m = 1 \end{aligned} \tag{2}$$

The stego-audio and video are re-encased in a video file and sent through a secure communication channel in Fig. 4.

3.8.2 CASE II-PVD Innovation

Using the PVD technique, the facial photo of the authorized user is integrated with the retrieved multimedia frames. The procedure for insertion is outlined below.

Compute the variation between P(i, x) and P(i, y) for each consecutive pixel P(i, x) in the cover picture as di. Depending on the di value, get the lower value (Ij) and the upper value (uj) from the table of range (Rj).

Compute $w_j = u_j + I_j + 1$

Compute the value of $t_i = \log(w_j)$ using the log base 2.

The t_i value gives the maximum number of bits that can be entered.

i_θ is the decimal figure of the t_i message

Compute the figure $\delta_i = \theta_i + I_j$

Compute the value of $m = abs(\delta_i - d_i)$

Calculate p_((i, x))^' and p_((i, y))^' by using Eq. (3)

$$p'_{(i,x)},\ p'_{(i,y)} = \begin{cases} \left(P_{(i,x)} + \left\lceil \frac{m}{2} \right\rceil,\ P_{(i,y)} - \left\lceil \frac{m}{2} \right\rceil\right), \\ P_{(i,x)} \geq P_{(i,y)} \text{ and } d_i' > d_i; \\ \left(P_{(i,x)} - \left\lceil \frac{m}{2} \right\rceil,\ P_{(i,y)} + \left\lceil \frac{m}{2} \right\rceil\right) \\ P_{(i,x)} < P_{(i,y)} \text{ and } d_i' > d_i; \\ \left(P_{(i,x)} - \left\lceil \frac{m}{2} \right\rceil,\ P_{(i,y)} + \left\lceil \frac{m}{2} \right\rceil\right) \\ P_{(i,x)} \geq P_{(i,y)} \text{ and } d_i' \leq d_i; \\ \left(P_{(i,x)} + \left\lceil \frac{m}{2} \right\rceil,\ P_{(i,y)} - \left\lceil \frac{m}{2} \right\rceil\right) \\ P_{(i,x)} < P_{(i,y)} \text{ and } d_i' \leq d_i; \end{cases} \tag{3}$$

The audio and video steganography results are then incorporated and sent via additive white noise through the communication channel.

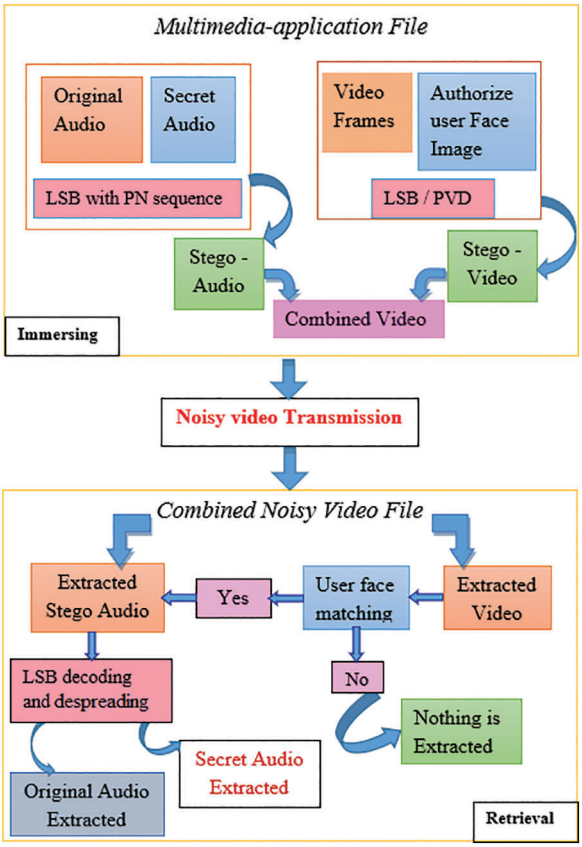

Figure 4: Block diagram of immersing and retrieval of prospective system

The stego-multimedia file's white noise is eliminated by the receiver side. Separating audio steganography and video steganography results, from the noiseless video file is the first step in the

retrieval procedure. The video file is retrieved from the stego-audio-video file in the decryption section, and the user picture from the specified frame is recovered using the reverse PVD technique.

In the stego-picture, for each consecutive pixel i.e., p_((i, x))^' and p_((i, y))^', compute the difference between p_((i, x))^' and p_((i, y))^' as d_i^'. Determine the lower bound (I_j) as well as the upper bound (u_j) from the table collection (R_j) based on the worth of d_i^'.

Actuate w_j = u_j − I_j + 1

Calculate the value of t_i = log(w_j) with 〖log〗_2.

t_i value specifies the maximum number of bits that may be entered.

Calculated d_i^" = d_i^' − l_j, transmit d_i^" into binary values using the length of t_i

The transformation of d_i^" converting to binary with a length of t_i is the immersed message.

The part of the differential value algorithm for two pixels are: Finding a distinct value for two neighboring pixels with a width difference of two pixels; using t length transforming d to binary; message power inserted in t; converting immersed messages to decimal; denoting two new pixels after immersing messages.

Then the input face picture is taken through the webcam and compared with the immersed image for authentication. Only if both pictures match, then only the user can retrieve the secret hidden message behind the audio; otherwise, the procedure will be delayed until the correct recipient appears in front of the webcam. The crucial process of this algorithm is to retrieve the original secret audio from stego-audio. It is executed by dispreading the SS signal followed by the reverse LSB method.

4 Results and Discussion

This section elaborates on the execution of the proposed face recognition and LSB-SS-based audio steganography method. MATLAB 2019aX64 bit version is used to implement the suggested system.

4.1 Face Recognition Analysis

The face recognition training parameters using the CNN algorithm are mentioned in Table 2.

Table 2: The suggested facial recognition system's CNN algorithm's training parameter

Sr. No.	Training parameters	Values
1	Training algorithm	'sgdm'
2	Momentum	0.9000
3	Batch size	10
4	Rate of initial learning	3e − 4
5	Drop period	10
6	Drop factor	0.1
7	Method of gradient threshold	'l2norm'

The Qualitative analysis is the irrational judgment of the performance of the system. The input from the testing dataset of four unauthorized people and one approved person is shown, as well as the output of the provided facial recognition system utilizing CNN.

The face recognition results, demonstrate that this CNN-based face recognition system is capable of accurately recognizing the face with greater accuracy.

4.2 Qualitative Analysis

4.2.1 Analysis of LSB-Based Steganography

The complete visual analysis or qualitative analysis of the prospective system for two different videos is amalgamated in the form of GUI independently.

The two different original audio-video multimedia files (VIDEO_STEGANOGRAPHY.avi and mylect.avi) are the input of this system. The audio file is retrieved from the multimedia file, and the waveform of the original audio file is presented under the host Audio. The secret audio is spread with the help of the PN sequence, and then, the listener in the original audio file is immersed by the LSB method. A stego-audio file is the output of this execution. The authentic user face and its binary representative image are used in the dataset.

These images are used to match the input image on the receiver side. The final stage-audio file after embedding the secret audio is shown under embedded multimedia. The waveform of the final immersed stage audio-video file will be sent over the channel.

The stego-audio-video file is obtained during the extraction stage, and the binary image recovered is shown in GUI. The authorized user face image captured by the webcam is toned with the binary image. Further, the process of extraction of the stego-hidden audio is performed and extracted secret audio.

From the qualitative analysis, it is noticed that the recovered binary image is used before to identify the authenticated user by comparing the image taken from the receiver's webcam with the transmitted image. Once the image matches, the secret audio file is achieved by final extraction.

In this approach of audio-video steganography, sustainable embedding is introduced that neither alters nor overwrites the bits. Since the cover file and the stego file have the same appearance, no one can predict the presence of stego in the communication channel.

4.3 Quantitative Analysis

The Peak Signal Noise Ratio (PSNR), Root Mean Square Error (RMSE) and Structural Similarity Index Matrix (SSIM) are used to evaluate the systems' results. The following is a more thorough description of this parameter:

4.3.1 PSNR

Peak Signal to Noise Ratio (PSNR) is an audio file characteristic for Peak Signal to Noise Ratio. PSNR and MSE are oppositely symmetrical, and PSNR can be calculated using the equation below.

$$\mathrm{PSNR} = 10\log_{10}\left[\frac{\mathrm{I}^2}{\mathrm{MSE}}\right] \tag{4}$$

where 'I' is the audio's highest possible value.

4.3.2 MSE

PSNR and MSE both are oppositely symmetrical to each other.

4.3.3 RMSE

The square root of MSE is used to calculate RMSE, which stands for Root Means Square Error.

$$RMSE = \sqrt{\frac{1}{[N \times M]^2} \sum_{i=1}^{N} \sum_{j=1}^{M} (X_{ij} - Y_{ij})^2} \tag{5}$$

4.3.4 SSIM

SSIM is a metric for the quality loss due to data change and loss during transmission. The SSIM is computed between the original audio and the extracted audio in this method.

$$SSIM(x,\ y) = \left(2\mu_x\mu_y + C_1\right) \tag{6}$$

where μ_x, μ_y are the local mean, σ_x, σ_y are the standard deviation, and σ_xy is the cross-variance for data x, y.

The mean, standard deviation, and cross-variance are calculated as follows:

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{7}$$

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)^2\right)^{\frac{1}{2}} \tag{8}$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)\left(y_i - \mu_y\right) \tag{9}$$

Two video embedding algorithms are employed, and the assessment factors are elaborated for various audio data in Table 3.

Table 3: Tabular data for quantitative analysis

Methodology proposed	Audio sample (.wav)	MSE	RMSE	PSNR	SSIM
LSB	LSB-duck	0.00000311	0.0018	55.0783	0.9964
LSB	LSB-as	0.00000556	0.0024	52.5505	0.9967
LSB	LSB-bird	0.00003400	0.00585	24.6598	0.9495
LSB	LSB-Dog	0.00001743	0.0042	47.5863	0.9963
LSB	LSB-Lion	0.00009547	0.0309	30.2015	0.9866
PVD	PVD-Duck	0.00000590	0.0024	52.2901	0.9963
PVD	PVD-as	0.00000521	0.0023	52.8355	0.9967
PVD	PVD-bird	0.00000453	0.0021	23.4337	0.9347
PVD	PVD-Dog	0.00000596	0.0024	42.2511	0.9962
PVD	PVD-Lion	0.00009247	0.0209	30.2015	0.9866

From the analysis of Figs. 5–8, it is understood that the PSNR and SSIM value of the samples is maximum while RMSE is lower. It shows the precision of the system. Comparative analysis of the prospective LSB & PVD embedding algorithm database distribution system is shown in Table 4.

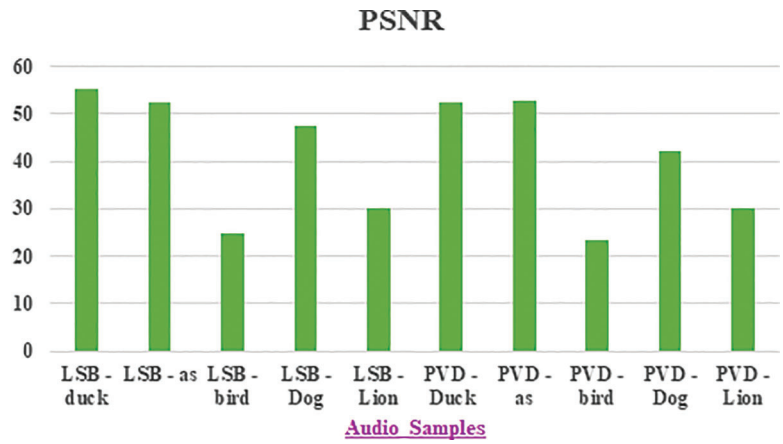

Figure 5: Graphical evaluation of PSNR for different audio samples and methods

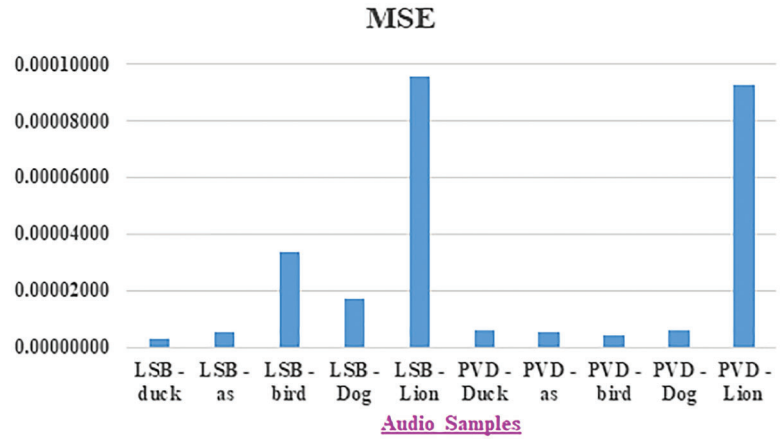

Figure 6: Graphical evaluation of MSE for different audio samples and methods

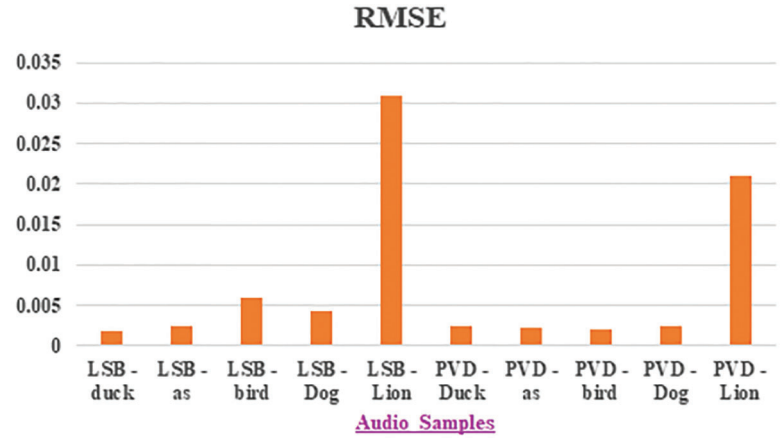

Figure 7: Graphical evaluation of RMSE for different audio samples and methods

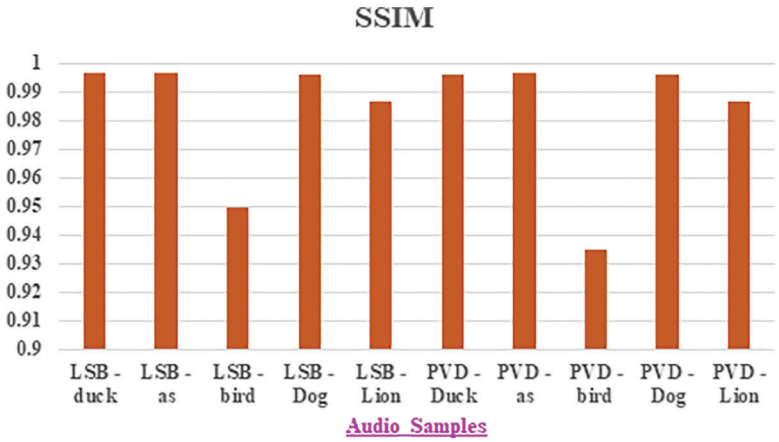

Figure 8: Graphical evaluation of SSIM for different audio samples and methods

Table 4: Comparative analysis of prospective LSB & PVD embedding algorithm database distribution system

Sr. No.	Comparative parameters	LSB immersing algorithm with video 1	LSB immersing algorithm with video 2	PVD immersing algorithm with video 1	PVD immersing algorithm with video 2
1	Video size before immersing	6.01 MB	2.59 MB	6.01 MB	2.59 MB
2	Video size after immersing	1 GB	976 MB	1.79 GB	1.54 GB
3	Video length before immersing	15 s	8 s	15 s	8 s
4	Video length after immersing	15 s	8 s	15 s	8 s
5	Secret audio size	23.8 KB	156 KB	23.8 KB	156 KB
6	Secret audio length	2 s	2 s	2 s	2 s
7	Embedding Capacity	@40%	@40%	@67.68%	@ 68%
8	PSNR	55.0783	52.5505	52.2901	52.8355
9	MSE	0.0000031058	0.0000055584	0.0000059	0.0000052053
10	RMSE	0.0018	0.0024	0.0024	0.0023
11	SSIM	0.9964	0.9967	0.9963	0.9967
12	Complexity	Less complex with less visual change	More complex with less visual change	More complex to encode and gives good visual quality with pre-processing	More complex to encode and gives good visual quality with pre-processing

The robustness of the steganography systems mostly depends on the scaling parameter for the immersing process. Most of the states of art methods use the static values of scaling parameters for the algorithm's simplicity. Even the statistic error values like MSE and RMSE have been found in the excellent embedding range, and statistically, there is no variance in the cover and stego-file. Hence, it can conclude that the system is sustainable for different embedding ranges.

It is observed from a comparative execution of MSE, RMSE, PSNR, and SSIM, the proposed methodology gives better results compared to the cutting-edge techniques.

5 Conclusion

This section of the paper elaborates on the sanguine data transmission done by three layers of security. This article examines the four main components of the picture steganography technique: the confidential auditory data propagating before embedding, the approved user facial embeds in the frame of the video, the embedding process and the extraction process. The converted sanguine spread signal immersed in the original audio of the video gives stego-audio. The authorized user face image immersed with independent algorithms as LSB and PVD in the frame of the original video gives the stego-video. The transmission of amalgamated two stego-signals combined with disturbance represents a high level of robustness. Only genuine receivers can retrieve confidential information with double security levels as the SS code of hidden data and authorized user's face image. This prospective work has also been done with application video, which results in ensuring the reliable video transmission is done, if accurate sanguine data is retrieved. The precision parameters like MSE, RMSE, PSNR and SSIM represent the excellent quality of noisy data transmission. The better immersing capacity is proved in comparative analysis.

Future improvements to the system could include adding different audio formats and video formats, as well as adjusting the SS code length. The SS code length variation will make the system more robust and genuine.

Acknowledgement: The authors wish to express their thanks to one and all, who supported them during this work.

Funding Statement: No funding was received to assist with the preparation of this manuscript.

Conflicts of Interest: The authors declare they have no conflicts of interest to report regarding the present study.

References

[1] S. Kumar, A. Singh and M. Kumar, "Information hiding with adaptive steganography based on novel fuzzy edge identification," *Defence Technology*, vol. 15, no. 3, pp. 162–169, 2019.

[2] E. Satir and H. Isik, "A huffman compression based text steganography method," *Multimedia Tools Application*, vol. 19, no. 2, pp. 283–298, 2012.

[3] A. Cheddad, J. Condell, K. Curran and P. M. Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, vol. 90, no. 15, pp. 727–752, 2010.

[4] A. Ioannidou, S. T. Halkidis and G. Stephanides, "A novel technique for image steganography based on a high payload method and edge detection," *Expert Systems with Applications*, vol. 39, no. 3, pp. 11517–11524, 2012.

[5] M. M. Sadek, A. S. Khalifa and M. G. M. Mostafa, "Video steganography: A comprehensive review," *Multimedia Tools Application*, vol. 34, no. 8, pp. 148–162, 2014.

[6] V. Sarangpure, R. Talmale and G. R. Babu, "Implementation on hiding data and image in audio-video using anti forensics technique," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no. 9, pp. 8159–8164, 2015.

[7] R. Mudusu, A. Nagesh and M. Sadanandam, "Enhancing data security using audio-video steganography," *International Journal of Engineering & Technology*, vol. 7, no. 2, pp. 276–279, 2020.

[8] P. Karthika and P. V. Saraswathi, "IoT using machine learning security enhancement in video steganography allocation for Raspberry Pi," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 5835–5844, 2021.

[9] G. Anvita, D. Singh and M. Kaur, "An efficient image encryption using non-dominated sorting genetic algorithm-III based 4-D chaotic maps," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp. 1309–1324, 2020.

[10] M. Balaji, "Analysis of basic neural network types for automated skin cancer classification using firefly optimization method," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 7, pp. 7181–7194, 2021.

[11] R. Tanwar, K. Singh, M. Zamani, A. Verma and P. Kumar, "An optimized approach for secure data transmission using spread spectrum audio steganography, chaos theory, and social impact theory optimizer," *Journal of Computer Networks and Communications*, vol. 19, no. 3, pp. 382–398, 2019.

[12] S. Kumar, A. Singh and M. Kumar, "Information hiding with adaptive steganography based on novel fuzzy edge identification," *Defence Technology*, vol. 15, no. 9, pp. 162–169, 2019.

[13] Z. S. Younus and G. T. Younus, "Video steganography using knight tour algorithm and lsb method for encrypted data," *Journal of Intelligent System*, vol. 29, no. 1, pp. 1216–1225, 2020.

[14] S. T. Chen and H. N. Huang, "Optimization-based audio watermarking with integrated quantization embedding," *Multimedia Tools and Applications*, vol. 75, no. 8, pp. 4735–4751, 2016.

[15] T. Idbeaa, S. A. Samad and H. Husain, "A secure and robust compressed domain video steganography for intra- and inter-frames using embedding-based byte differencing (EBBD) scheme," *PLoS One*, vol. 11, no. 3, pp. 372–389, 2016.

[16] M. Ramalingam and N. A. Isa, "A data-hiding technique using scene-change detection for video steganography," *Computers and Electrical Engineering*, vol. 23, no. 2, pp. 1–12, 437–446, 2015.

[17] Y. Xue, J. Zhou, H. Zeng, P. Zhong and J. Wen, "An adaptive steganographic scheme for H.264/AVC video with distortion optimization," *Signal Processing: Image Communication*, vol. 76, no. 12, pp. 22–30, 2019.

[18] H. Ramakrishna and S. Jagadeesha, "An optimal modified matrix encoding technique for secret writing in MPEG video using ECC," *Computer Standards & Interfaces*, vol. 48, no. 3, pp. 173–182, 2016.

[19] V. Ulagamuthalvi, G. Kulanthaivel, A. Balasundaram and A. K. Sivaraman, "Breast mammogram analysis and classification using deep convolution neural network," *Computer Systems Science and Engineering*, vol. 43, no. 1, pp. 275–289, 2022.

[20] C. H. Yang, C. Y. Weng, H. K. Tso and S. J. Wang, "A data hiding scheme using the varieties of pixel-value differencing in multimedia images," *The Journal of Systems and Software*, vol. 84, no. 7, pp. 669–678, 2011.

[21] M. Ganga, N. Janakiraman, A. K. Sivaraman, R. Vincent, A. Muralidhar *et al.,* "An effective denoising and enhancement strategy for medical image using Rl-gl-caputo method," *Advances in Parallel Computing (Smart Intelligent Computing and Communication Technology) & IOS Press*, vol. 38, no. 3, pp. 402–408, 2021.

[22] R. Kumar, D. S. Kim and K. H. Jung, "Enhanced AMBTC based data hiding method using hamming distance and pixel value differencing," *Journal of Information Security and Applications*, vol. 47, no. 9, pp. 94–103, 2019.

[23] W. Hong, T. S. Chen and C. W. Luo, "Data embedding using pixel value differencing and diamond encoding with multiple-base notational system," *The Journal of Systems and Software*, vol. 85, no. 17, pp. 1166–1175, 2012.

[24] R. Priya, S. Jayanthi, A. K. Sivaraman, R. Dhanalakshmi, A. Muralidhar *et al.,* "Proficient mining of informative gene from microarray gene expression dataset using machine intelligence," *Advances in Parallel Computing (Smart Intelligent Computing and Communication Technology), IOS Press*, vol. 38, no. 3, pp. 417–422, 2021.

[25] W. Zhao, Z. Jie, L. Xin and W. Qiaoyan, "Data embedding based on pixel value differencing and modulus function using indeterminate equation," *The Journal of China Universities of Posts and Telecommunications*, vol. 22, no. 1, pp. 95–100, 2015.

[26] M. Long, Y. Zhao, X. Zhang and F. Peng, "A separable reversible data hiding scheme for encrypted images based on tromino scrambling and adaptive pixel value ordering," *Signal Processing*, vol. 10, no. 4, pp. 216–228, 2020.

[27] P. Arunachalam, N. Janakiraman, A. K. Sivaraman, A. Balasundaram, R. Vincent *et al.,* "Synovial sarcoma classification technique using support vector machine and structure features," *Intelligent Automation & Soft Computing*, vol. 32, no. 2, pp. 1241–1259, 2021.

[28] Y. Zhile, "A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting," *Neurocomputing*, vol. 397, no. 9, pp. 415–421, 2020.

[29] A. K. Sahu and G. Swain, "High fidelity based reversible data hiding using modified LSB matching and pixel difference," *Journal of King Saud University-Computer and Information Sciences*, vol. 23, no. 2, pp. 529–543, 2019.

[30] F. Yinfeng, "Modelling EMG driven wrist movements using a bio-inspired neural network," *Neurocomputing*, vol. 7, no. 1, pp. 832–854, 2021.

[31] L. Zhibin, "A convolutional neural network using dinucleotide one-hot encoder for identifying DNA N6-methyladenine sites in the rice genome," *Neurocomputing*, vol. 422, no. 3, pp. 214–221, 2021.

[32] Y. P. Lee, J. C. Lee, W. K. Chen, K. C. Chang, I. J. Su *et al.,* "High-payload image hiding with quality recovery using tri-way pixel-value differencing," *Information Sciences*, vol. 191, no. 5, pp. 214–225, 2012.

[33] N. I. Wu, K. C. Wu and C. M. Wang, "Exploring pixel-value differencing and base decomposition for low distortion data embedding," *Applied Soft Computing*, vol. 12, no. 4, pp. 942–960, 2012.

[34] P. Arunachalam, N. Janakiraman, J. Rashid, J. Kim, S. Samanta *et al.,* "Effective classification of synovial sarcoma cancer using structure features and support vectors," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 2521–2543, 2022.

[35] N. Zarmehi and M. A. Akhaee, "Digital video steganalysis toward spread spectrum data hiding," *IET Image Processing*, vol. 23, no. 4, pp. 1–8, 2015.

[36] L. N. Q. Khanh, "Deep ETC: A deep convolutional neural network architecture for investigating and classifying electron transport chain's complexes," *Neurocomputing*, vol. 375, no. 5, pp. 71–79, 2020.

[37] S. Karthik, R. S. Bhadoria, J. G. Lee, A. K. Sivaraman, S. Samanta *et al.,* "Prognostic kalman filter based Bayesian learning model for data accuracy prediction," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 243–259, 2022.

[38] I. Cosimo, "A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, no. 4, pp. 96–107, 2019.

[39] B. Gupta, "A modified approach of video steganography for information hiding," *Compusoft, an International Journal of Advanced Computer Technology*, vol. 7, no. 8, pp. 2803–2806, 2018.

[40] P. R. Deshmukh and B. Rahangdale, "Data hiding using video steganography," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 4, pp. 856–860, 2014.

[41] S. Samanta, V. K. Dubey and K. Das, "Coopetition bunch graphs: Competition and cooperation on COVID-19 research", *Information Sciences*, vol. 589, pp. 1–33, 2022.

[42] S. Samanta, V. K. Dubey and B. Sarkar, "Measure of influences in social networks", *Applied Soft Computing*, vol. 99, pp. 106858, 2021.

[43] S. Samanta and B. Sarkar, "Isomorphism on generalized fuzzy graphs and imagevisualizations", *Soft Computing*, vol. 24, no. 19, pp. 14401–14409, 2020.