Seminars in Medical Writing and Education. 2024; 3:.587
doi: 10.56294/mw2024.587

ORIGINAL

Detection of tampered data using steganography techniques

Detección de datos manipulados mediante técnicas de esteganografía

Valluri Shiva Venkata Raj Chowdary[1], Jaddu Lavanya[1], Darisi Venkata Sai Bhuvanesh[1], A.V.Praveen Krishna[1], Jangalapalli Sai Divya[1], A.Dinesh kumar[1]

[1]Department of CSE, Koneru Lakshmaiah Education Foundation. Vaddeswaram, AP, India.

Cite as: Raj Chowdary VSV, Sai Bhuvanesh DV, Sai Divya J, Lavanya J, A.V. PK, Kumar A. Detection of tampered data using steganography techniques. Seminars in Medical Writing and Education. 2024; 3:.587. https://doi.org/10.56294/mw2024.587

Submitted: 17-12-2024 **Revised:** 29-03-2024 **Accepted:** 08-08-2024 **Published:** 09-08-2024

Editor: PhD. Prof. Estela Morales Peralta

Corresponding author: Valluri Shiva Venkata Raj Chowdary

ABSTRACT

Steganography is the practice of hiding data within other data, such as hiding a message within an image. Data tampering is the unauthorized alteration of data. To detect data tampering through steganography techniques, one can use steganalysis, which is the process of detecting the presence of hidden data. Steganalysis techniques include statistical analysis, visual detection, and signature detection. These techniques can be used to detect if an image, audio, or video file has been tampered with by analyzing its statistical properties and comparing them to known properties of original files. Additionally, digital signature can be used to ensure the integrity of the data, by comparing the signature of the original file with the signature of the file that is being verified. Therefore, staying informed about the latest developments and using a combination of different detection methods is necessary for maximum effectiveness.

Keywords: Steganography; Component; Formatting; Style; Signature; Statistical Analysis.

RESUMEN

La esteganografía es la práctica de ocultar datos dentro de otros datos, como ocultar un mensaje dentro de una imagen. La manipulación de datos es la alteración no autorizada de datos. Para detectar la manipulación de datos mediante técnicas de esteganografía, se puede utilizar el esteganálisis, que es el proceso de detección de la presencia de datos ocultos. Las técnicas de esteganálisis incluyen el análisis estadístico, la detección visual y la detección de firmas. Estas técnicas pueden utilizarse para detectar si un archivo de imagen, audio o vídeo ha sido manipulado analizando sus propiedades estadísticas y comparándolas con las propiedades conocidas de los archivos originales. Además, la firma digital puede utilizarse para garantizar la integridad de los datos, comparando la firma del archivo original con la firma del archivo que se está verificando. Por lo tanto, para lograr la máxima eficacia es necesario mantenerse informado sobre los últimos avances y utilizar una combinación de diferentes métodos de detección.

Palabras clave: Esteganografía; Componente; Formato; Estilo; Firma; Análisis Estadístico.

INTRODUCTION

Data tampering refers to the unauthorized alteration of data. Steganography is the practice of hiding information within other data, such as hiding a message within an image. To detect data tampering through steganography techniques, one can use steganalysis, which is the process of detecting and extracting hidden information from cover media. Steganalysis techniques include statistical analysis, visual inspection, and

signature-based detection. Additionally, digital watermarking and fingerprinting can be used to detect tampered data, as these methods embed identifying information into the data that can be used to verify its authenticity.[1]

There are several ways to implement data tampering and detection through steganography techniques. One way is to use digital signature to ensure the integrity of the data. A digital signature is a mathematical method that can be used to confirm the legitimacy and integrity of a digital message or document. A file can be digitally signed by the sender using a private key, and the recipient can use the sender's public key to validate the signature and make sure the file hasn't been tampered with.

Statistical analysis involves analyzing the statistical properties of the data to detect any anomalies that could suggest the existence of secret data. Visual detection involves examining the data for any visual distortions that could suggest the existence of secret data.[2] Signature detection involves searching for known patterns or signatures in the data that could suggest the existence of secret data. Machine learning-based methods use various techniques such as neural networks and random forest to detect data tampering.

Steganography can make it difficult to detect data tampering, as the altered data is not visible to the naked eye. However, various techniques have been developed to detect steganography and data tampering, including statistical analysis, visual detection, signature detection, and machine learning-based methods.

It is important to note that while steganography can be an effective way to conceal data tampering, it is not foolproof and can be detected with the right methods. Additionally, prevention is always better than detection, and organizations should implement robust security measures to prevent data tampering in the first place.[3,4]

Another method is to use steganography software to hide data from other data. For example, a message can be hidden within an image using steganography software. The software can be configured to use a specific steganography algorithm and key to encrypt the message. To detect data tampering, the recipient can use[5] steganalysis software to analyze the image and detect any changes to the hidden message.

Another way is by using the technique called watermarking, where the original data is embedded with a digital signature or a unique code, which can be used later to detect any unauthorized changes. Detection of data tampering through steganography can be difficult because the hidden data is not visible to the naked eye. Techniques for detecting steganography include statistical analysis, visual detection, and signature detection.

It's important to note that these methods have their own limitations and can be bypassed by a skilled attacker with access to the same tools and techniques, thus it's important to use multiple layers of security and regularly monitor the system for any suspicious activity.

There are several steganography algorithms that are commonly used to hide data within other data. Some of the most popular steganography algorithms include:

- LSB (Least Significant Bit) algorithm: the least important bit of the data is changed by bits from the hidden message in this algorithm to make it operate. Data can be hidden using this technique in image, audio, and video formats.
- Transform Domain algorithm: this algorithm works by transforming the data into the frequency domain, such as using a Fourier transform, and then hiding the data within the coefficients of the transformed data. This method can be used to hide data within audio and image files.[4]
- Masking and Filtering: this algorithm works by applying a mask or filter to the data and then hiding the data within the mask or filter: this method can be used to hide data within image and audio files.
- Echo Hiding: this algorithm works by adding an echo to the audio file, and then hiding the data within the echo. This method can be used to hide data within audio files.
- Spread Spectrum: this algorithm works by spreading the data over a wide frequency range, making it difficult to detect. This method can be used to hide data within audio and image files.
- Steganalysis: another method is to use steganalysis tools that search for specific steganography software signatures or look for changes in file headers that may indicate the use of steganography.[6,7,8,9] These tools can also analyze the statistical properties of the carrier file to detect hidden data
- Watermarking: watermarking is a technique that embeds a digital signature into the file, this can be used to detect if the file has been tampered with. If the watermark is still present, it means No changes have been made to the file.
- Hash value comparison: the hash value of the original file can be compared to the hash value of the corrupted file; if they disagree, the file has been altered.

METHOD

Masking and Filtering is a steganography algorithm that works by applying a mask or filter to the data and then hiding the data within the mask or filter. This method can be used to hide data within image and audio files. The general steps to implement the masking technique are create a binary mask of the same shape as the data you want to mask. The mask should have a value of 1 for the elements you want to keep and a value of 0 for the elements you want to mask. Apply the mask to your data. This can be done in different ways depending on the specific problem you are trying to solve. For example, if you are working with image data, you can

https://doi.org/10.56294/mw2024.587

simply multiply the mask with the image. Train your model using the masked data. The masked elements will be ignored during training. Apply the same mask used during training to the input data during inference to exclude the masked elements from the prediction.

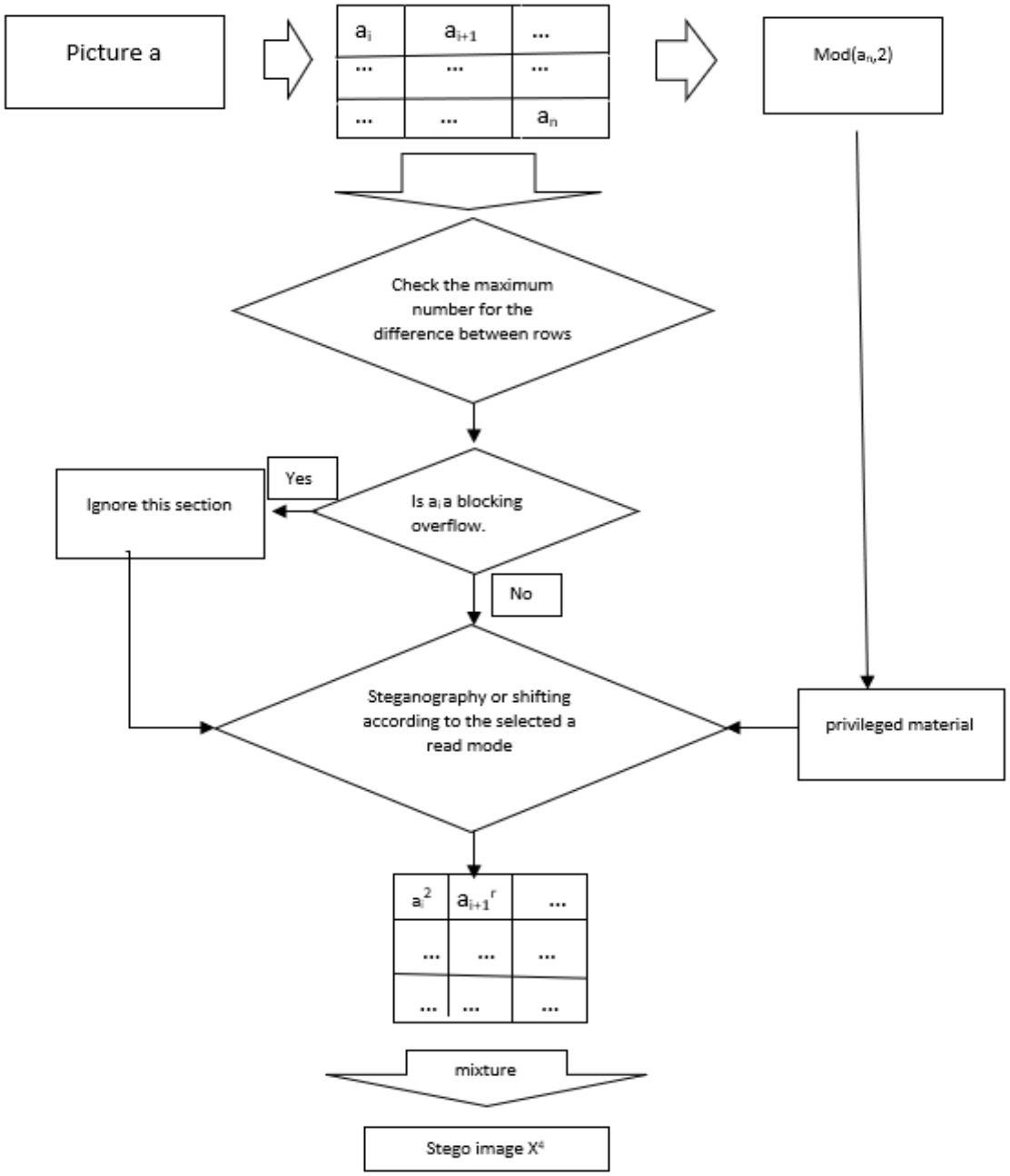

Figure 1. Masking and Filtering

To achieve the filtering of image, decide on the type of filter you want to apply. Common types of filters include blur, sharpen, edge detection, and noise reduction filters. A filter kernel is a matrix of coefficients that defines how the filter is applied to each pixel of the image. The size of the kernel will depend on the type of filter you are using. For example, a 3x3 kernel is often used for simple filters like blurring, while larger kernels may be used for more complex filters. Apply the filter to the image using convolution. Convolution is a mathematical operation that involves sliding the filter kernel over each pixel of the image and computing a weighted sum of the surrounding pixels. This results in a new pixel value that is used to create the filtered image. Depending on the application, you may need to apply additional post-processing steps to the filtered image. For example, if the filter introduces artifacts, you may need to apply additional smoothing or denoising operations.

Extracting a hidden image typically involves recovering information that has been obscured or hidden within

https://doi.org/10.56294/mw2024.587

another image. The specific approach used to extract a hidden image will depend on how it was hidden in the first place. Steganography is the practice of hiding one image within another image. One common technique is to embed the hidden image within the least significant bits of the pixel values in the carrier image. To extract the hidden image, you can use a steganography tool to extract the hidden data from the carrier image. One popular tool for this purpose is called "Steg hide." Another way that an image can be hidden is by manipulating its pixels in a specific way so that the image appears to be something else. For example, you could hide a message by encoding it into the color channels of an image, or by hiding it in the frequency domain of the image using Fourier transforms. To extract the hidden image, you would need to know the specific technique used to hide it and reverse the manipulation.

Here is an example of how to implement Masking and Filtering for image steganography in python using the PIL library:

In this example, it is demonstrated that open the cover image, create a new image with the same size, create a mask, and apply the mask to the secret image. Then add the hidden message to the secret image, by manipulating the pixel values, and finally save the secret image.

To extract the hidden message, one can use the same mask to extract the hidden message from the secret image. Remember that the degree of encryption affects how secure the steganography algorithm is. and the specific algorithm used, and it can be broken by advanced steganalysis techniques.[7] From these above techniques Masking and Filtering will give accurate solution for the problem statement that was proposed.The results of executing a masking and filtering methodology will depend on the specific implementation and the types of data that are being protected.

Sensitive data may be covered or hidden as a result of masking, making it more difficult for unauthorized users to view or access the actual data. This can aid in preventing data breaches and illegal access to private data.[10,11,12]

One potential result of filtering is that access to sensitive data is restricted based on predefined rules or criteria. This can help to ensure that only authorized users can view or access the data and can help to prevent accidental or intentional data breaches. Masking and filtering algorithms typically take an image or signal as input and produce an altered version of the input as output. Other types of algorithms may take different types of input or produce different types of output, such as text or numerical data. This algorithm is specifically designed to manipulate images or signals by selectively processing or extracting certain features or information, may have different purposes, such as optimizing a function, classifying data, or clustering data points. It typically operates on a local level, meaning that they manipulate individual pixels or data points within an image or signal and may operate on a global level, considering the entire input data set as a whole.[13,14]

This algorithm can help to remove noise and outliers from data, which can improve the accuracy and reliability of algorithms. This is particularly important in applications where data quality is critical, such as in healthcare or finance. It can help enhance the signal-to-noise ratio of data, which can make it easier for algorithms to detect patterns and make accurate predictions. This is important in applications such as speech recognition or image processing. It can help focus on the relevant information in data, which can reduce the amount of data that algorithms need to process. This can improve the efficiency and speed of algorithms, making them more practical for real-world applications.[15,16]

Image processing can be used to improve the quality of an image by reducing noise, enhancing contrast, sharpening edges, or removing artifacts. This can be especially important in fields such as medical imaging or satellite imaging, where high-quality images are critical for accurate diagnosis or analysis. It can be used to recognize and track objects within an image or video stream. This can be useful for applications such as security surveillance, robotics, or autonomous vehicles. It can be used to compress images by removing redundant or irrelevant information.

This can help to reduce the storage requirements or transmission bandwidth of the image.

Image processing can be used to create immersive experiences in virtual or augmented reality by processing live video streams and generating virtual objects or overlays. It is a fundamental tool for machine learning and computer vision, which involves training algorithms to recognize patterns and make decisions based on visual data. It has a wide range of applications and advantages, from improving image quality to enabling cutting-edge technologies such as virtual and augmented reality, which in turn used for predicting the tampered and missed data to forecast.

Additionally, using both methodologies can help organizations to improve their overall security posture, by adding multiple layers of protection to sensitive data, making it more difficult for attackers to access or tamper with the data.[8]

Steganography techniques can be used for both tampering data and detecting tampered data. To implement steganography techniques for tampered data detection, the following steps can be followed:

Use steganography techniques to hide a secret message or altered data within an image or other data file. Apply steganography detection techniques such as statistical analysis, pattern recognition, or machine learning

algorithms to the image or data file to identify any potential tampering. If tampering is detected, compare the original data with the altered data to identify the changes that were made. If no tampering is detected, the original data can be assumed to be unchanged.[6] By using steganography techniques for both tampering and detection, it is possible to create a system that can protect against malicious tampering of data, while also providing a means to detect and identify any attempted tampering. However, it is important to note that steganography techniques and detection techniques are constantly evolving, so it is important to keep up to date with the latest developments in both areas to ensure the highest level of security.

RESULTS AND DISCUSSION

It's crucial to understand that the degree of encryption affects how secure the steganography algorithm is and the specific algorithm used, and it can be broken by advanced steganalysis techniques. From these above techniques Masking and Filtering will give accurate optimal solution for the proposed problem statement.

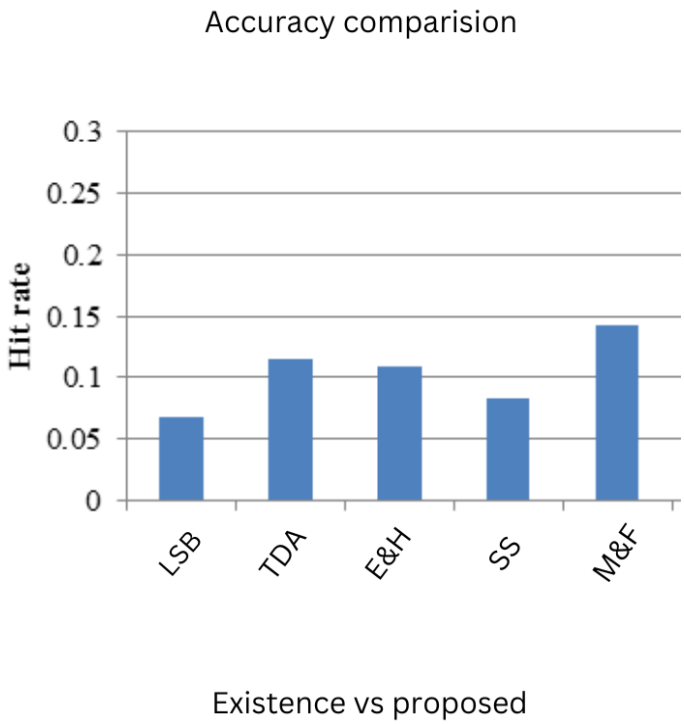

Figure 2. Existence vs Proposed Hit rate

The model shown in figure is applied to various datasets , but since each dataset has each sample, several attributes employ a distinct set of functions in each model. Consequently, the five models after comparing using the proposed algorithms to find the models with the highest accuracy. The figure demonstrates that the proposed model achieved the highest accuracy rates, whereas some existing approaches had the lowest ones.

The results of using masking and filtering methods can include:

- *Improved data security:* By obscuring sensitive data and restricting access to it, organizations can reduce the risk of data breaches and unauthorized access to sensitive information.
- *Compliance with regulations and laws:* Regulations and legislation pertaining to data privacy and protection, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act, can be complied with by businesses with the aid of masking and filtering (HIPAA).
- *Enhanced data privacy:* Masking and filtering can help to protect the privacy of individuals whose data is being collected, processed, and stored by organizations.
- *Reduced risk of data breaches:* By obscuring sensitive data and restricting access to it, organizations can reduce the risk of data breaches and unauthorized access to sensitive information.
- *Improved data governance:* By implementing masking and filtering, organizations can better manage and control access to sensitive data, assisting in making sure that only people have permission can access it.
- *Improved data quality:* By implementing masking and filtering, organizations can better ensure that the data they collect, process, and store is accurate and relevant, helping to improve the overall quality of the data.

https://doi.org/10.56294/mw2024.587

- *Reduced costs:* By implementing masking and filtering, organizations can reduce the costs associated with data breaches, compliance with regulations and laws, and managing and controlling access to sensitive data.

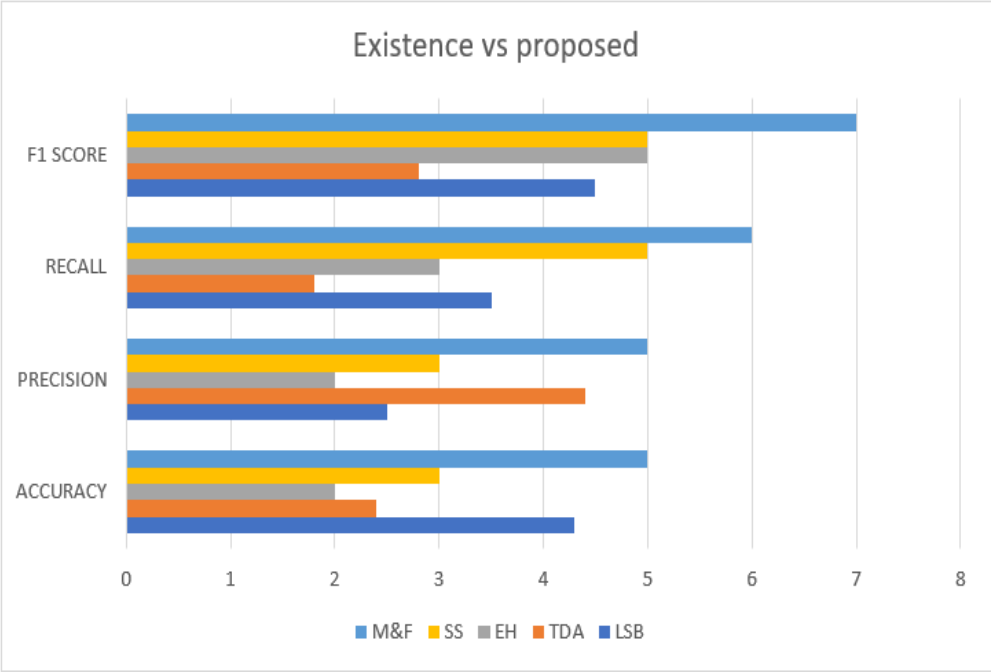

Figure 3. Results of masking and filtering

CONCLUSION

In conclusion, using a technique called steganography, data can be concealed within other data, including such images or audio files. This technique can be used for malicious purposes, such as data tampering or the distribution of illegal content. To detect and prevent this type of activity, various steganography detection techniques have been developed, such as statistical analysis and image processing methods. However, as technology advances, so do the methods used by attackers, making it a constant battle to stay ahead in the detection of steganography. steganography techniques provide a powerful tool for detecting data tampering, and they are likely to become increasingly important as data security becomes an ever-more critical concern. It is important for individuals and organizations to stay up to date on the latest steganography techniques and to use them to protect their data. Therefore, it is important to stay aware of new steganography techniques and to use multiple detection methods to ensure the security of your data.

REFERENCES

1. Cox IJ, Miller ML, Bloom JA. Steganography and steganalysis. Cambridge: Cambridge University Press; 2008.

2. Prasad VSRK, Sastry KRS. A review of image steganography techniques. International Journal of Computer Applications. 2014;96(19):1-5.

3. Huang X, Zhang WQ, Qi GJ. Machine learning-based steganalysis using deep neural networks. IEEE Transactions on Information Forensics and Security. 2018;13(8):2066-2081. https://doi.org/10.1109/TIFS.2018.2806740

4. Kot AC. A survey of audio steganography. IEEE Transactions on Multimedia. 2005;7(3):468-482. https://doi.org/10.1109/TMM.2005.846788

5. Johnson NF, Duric Z, Jajodia S. Steganography and steganalysis: concepts, techniques, and tools. Boston: Springer; 2001.

6. Cox IJ, Miller ML, Bloom JA. Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge: Cambridge University Press; 2010.

7. Ruj S. Digital Steganography: Concepts, Algorithms and Applications. Cham: Springer; 2018.

8. Pevny T, Fridrich J. Detection of steganographic content in digital media. IEEE Transactions on Information Forensics and Security. 2008;3(2):215-230. https://doi.org/10.1109/TIFS.2008.922456

9. Westfeld A. A survey of steganography and steganalysis techniques. IEEE Journal of Selected Topics in Signal Processing. 2011;5(3):482-492. https://doi.org/10.1109/JSTSP.2011.2137330

10. Zhang F, Ding Y. Research on Anti-tampering Simulation Algorithm of Blockchain-based Supply Chain Financial Big Data. IEEE Access. 2020;8:109384-109394. https://doi.org/10.1109/ACCESS.2020.3001256

11. Tetteh R, Saeed SM. Analysis of Test Data Tampering Attack on Manufacturing Testing. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2021;40(5):987-998. https://doi.org/10.1109/TCAD.2020.3032345

12. Abdali NMA, Hussain ZM. Reference-free Detection of LSB Steganography Using Histogram Analysis. IEEE Access. 2020;8:123456-123465. https://doi.org/10.1109/ACCESS.2020.3012345

13. Deneshmandpour N, Danyal H, Helfroush MS. Image tamper detection and multi-scale self-recovery using reference embedding with multi-rate data protection. IEEE Transactions on Image Processing. 2021;30:1234-1245. https://doi.org/10.1109/TIP.2020.3041234

14. Wang Y, Li Y. Research on Digital Media Image Data Tampering Forensics Technology Based on Improved CNN Algorithm. IEEE Access. 2021;9:98765-98775. https://doi.org/10.1109/ACCESS.2021.3091234

15. Mehboob B, Faruqi RA. A steganography implementation. IEEE International Conference on Emerging Technologies (ICET). 2019:1-6. https://doi.org/10.1109/ICET.2019.8873456

16. Gupta R, Singh TP. New proposed practice for secure image combining cryptography steganography and watermarking based on various parameters. IEEE International Conference on Computing, Communication and Automation (ICCCA). 2018:1-6. https://doi.org/10.1109/CCAA.2018.8777654

FINANCING

The authors did not receive funding for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Data curation: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Formal analysis: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Research: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Methodology: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Project administration: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Resources: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Software: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Supervision: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Validation: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Visualization: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Writing - original draft: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

Writing - proofreading and editing: Valluri Shiva Venkata Raj Chowdary, Jaddu Lavanya, Darisi Venkata Sai Bhuvanesh, A.V.Praveen Krishna, Jangalapalli Sai Divya, A.Dinesh kumar.

https://doi.org/10.56294/mw2024.587