e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:09/September-2024 Impact Factor- 7.868 www.irjmets.com

# SYSTEMATIC REVIEW OF TEXT STEGANOGRAPHY TECHNIQUES FOR CLANDESTINE COMMUNICATION

**Mohammed Usman**[*1], **Isah Nasir**[*2], **Bulama Mohammed Dahiru**[*3]

[*1]Department Of Computer Science, Modibbo Adama University, Yola, Nigeria.

[*2]Department Of Computer Science, Federal College Of Horticulture, Dadinkowa Gombe, Nigeria.

[*3]Department Of Computer Science, Yobe State University, Damaturu, Nigeria.

## ABSTRACT

Text steganography is a technique of hiding information within text in a way that prevents detection by unintended recipients. Unlike cryptography, which secures the content of a message by transforming it into an unreadable form, steganography conceals the existence of the message itself. All traditional steganographic techniques have limited ability to hide information. They can hide at most 10% of the cover data. While recent research has predominantly focused on hiding data in images, applying these techniques to natural language text presents greater complexity. Various text steganography algorithms are employed depending on specific requirements. To meet shared demands, it is essential to achieve both high-level concealment and the ability to hide substantial amounts of confidential data. The findings show that field of text steganography is continuously evolving, with ongoing research addressing both traditional and emerging challenges. Numerous studies indicate a strong interest in comparing different techniques to identify the most effective methods. This study suggests combining steganography with advanced encryption techniques or blockchain to create more secure and tamper-proof communication systems and highlights the diversity and depth of research in text steganography, pointing to both current challenges and promising directions for future study.

**Keywords**: Stego, Cover, Security, Steganography, Discrete.

## I. INTRODUCTION

In today's world, the internet plays a crucial role in communication and information sharing. However, with the rapid advancements in Information Technology and Communication, data security has increasingly become a concern. Each day, sensitive information is compromised, and unauthorized access to data has reached alarming levels. It is essential to take significant measures to protect this data and information. Combining steganography with encryption offers a powerful and effective solution for ensuring a high level of security.

Hmood et al. (2021) stated that steganography is both an art and a science where a message is concealed within another message, making it accessible only to the intended recipient. The term "steganography" is derived from the Greek words "Steganos" (meaning "Covered") and "Graptos" (meaning "Writing"). In the modern context, steganography typically involves hiding information or a file within a digital image, text, video, or audio file. The main purpose of this technique is not to reveal the hidden data to others, but to ensure the data's existence remains undetected.

The importance of steganographic applications is especially evident when used with digital images. This combination serves multiple purposes, including copyright protection, feature tagging, and covert communication. The focus of steganography is on ensuring the presence of hidden data without revealing it, making it particularly valuable for scenarios where data concealment and assurance are crucial. The goal of steganography is not to stop others from accessing the secret information, but to prevent them from even realizing that the information exists. In essence, steganography is often referred to as "invisible" communication.

Krishnan et al. (2020) observed that text steganography is considered the most challenging form of steganography compared to other types, such as those applied to images, video files, or audio files, due to the limited redundancy of information in a text file.

This study reviewed advances made on text steganography based on the methods and problems the addressed as well as their challenges for hiding information. This will assist with valuable technique for enhancing privacy, protecting sensitive information, and facilitating secure communication in both personal and professional contexts.

## II. METHODOLOGY

### Literature Review

Text steganography is a technique used to hide secret information within a text in such a way that the existence of the hidden message is concealed. Unlike encryption, which makes the message unreadable to anyone without the decryption key, steganography embeds the secret message within another, seemingly innocent text. This allows the hidden message to be transmitted without arousing suspicion.

Recently, Internet usage has increased dramatically. One of the key areas of interest is security, particularly in the context of online communication. Today, one of the most popular techniques for safeguarding data goes beyond encryption. To enhance secure communication, various methods, such as encryption and coding, are employed. Steganography, in particular, involves concealing data within other data, making its presence undetectable to outsiders. This feature gives steganography a significant advantage over other coding systems.

According to Sahoo, and Tiwari (2020) text steganography is important for several reasons;

**Covert Communication:** Text steganography allows for the hidden transmission of messages within seemingly ordinary text, making it an effective tool for covert communication. This is especially valuable in environments where traditional encryption might draw suspicion.

**Data Security:** By embedding secret information within a text, text steganography adds an additional layer of security. Even if a message is intercepted, the hidden data remains concealed, reducing the risk of unauthorized access.

**Low Detection Risk:** Unlike other forms of steganography that use images, audio, or video files, text steganography can be less detectable due to the subtlety and simplicity of text. This makes it harder for adversaries to recognize the presence of hidden information.

**Resource Efficiency:** Text files are lightweight and require minimal storage compared to multimedia files, making text steganography an efficient method for transmitting hidden information, especially in bandwidth-limited environments.

**Historical and Modern Relevance:** Text steganography has a long history, dating back to ancient times when messages were concealed within written documents. In the modern digital age, it continues to be relevant, especially in contexts where digital communication is monitored or censored.

**Censorship Resistance:** In regions where communication is heavily monitored or censored, text steganography can help bypass restrictions by embedding hidden messages within innocuous-looking text, allowing the exchange of sensitive information without detection.

Yusuf and Mohammed (2020) explained that you require less memory to store text files, and is faster and easier to communicate with than other steganographic techniques, but it also requires careful design to avoid detection and to ensure that the hidden message is correctly interpreted by the intended recipient.

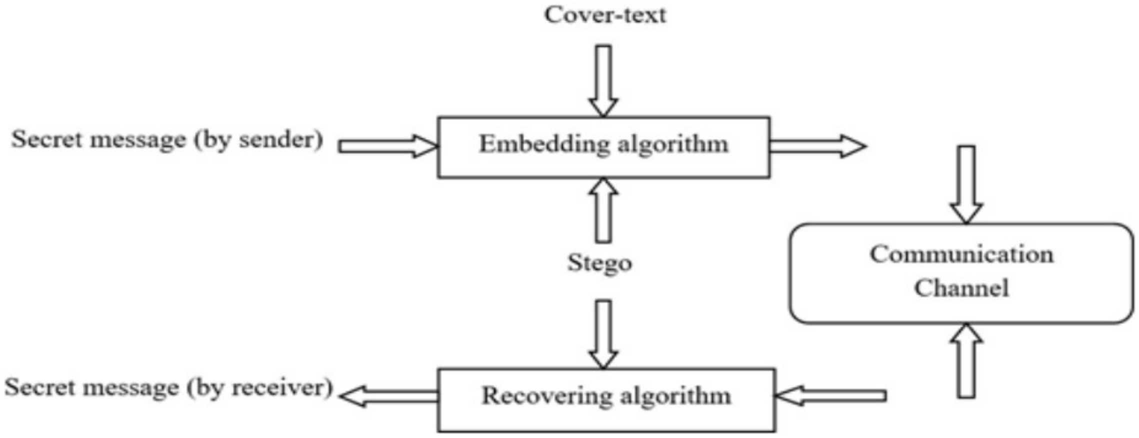

**Figure 1:** Process of Text Steganography. (Source: Biradar & Umashetty, 2016).

Figure 1 describes the procedure of text steganography involves several steps beginning with message preparation in whichthe information you want to hide is prepared. This could be text, binary data, or any other form of information. The secret message may be encoded into a binary format or a specific pattern that will be embedded into the cover text.

The next step is choosing the cover textwhere a non-suspicious, ordinary-looking text is selected as the carrier. This text will hold the hidden information.

Embedding algorithm Selection involve hiding the secret message within the cover text using Whitespace manipulation, character substitution, text format manipulation or linguistic steganography to obtain a stego. The stego text, which appears to be a regular, innocuous message, is then transmitted or stored as needed. Its hidden content is not noticeable to those unaware of the embedding process.

The communication process is completed by extraction wherethe recipient, who knows the embedding method and the carrier text, uses the reverse process to extract the hidden message.For instance, if whitespace manipulation was used, the recipient will analyze the spaces and tabs to reconstruct the original secret message.

The entire procedure is designed to ensure that the hidden message is invisible to unintended parties, maintaining the confidentiality of the communication.

**Techniques of Text Steganography**

Text steganography involves various methods to conceal information within text. Some common techniques are highlighted in figure 2;

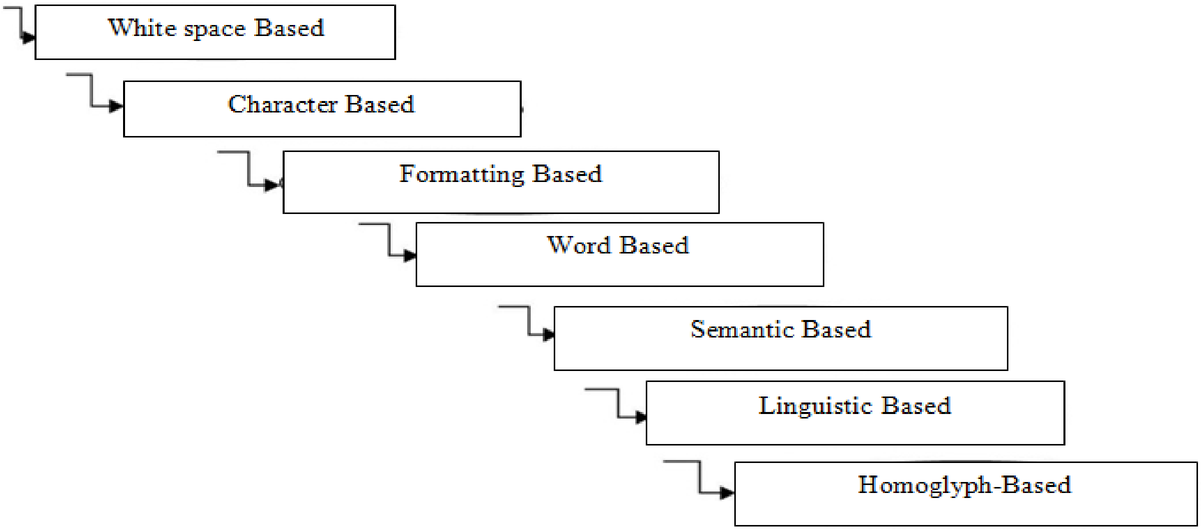

**Fig. 2** Techniques of Text Steganography

Whitespace-based Steganography utilizes spaces, tabs, and line breaks to encode information. For example, a sequence of spaces might represent binary digits (0s and 1s). Message can be encoded using binary where a single space represents 0 and a double space represents 1. Character-based however, involves using subtle variations in character appearance or encoding to hide information. This could include deliberate Typos-using incorrect spellings or extra characters, or punctuation by manipulating punctuation marks to convey hidden data. Example can be using the extra letter in “recieve” to signify a binary digit.

Formatting-based Steganography uses text formatting features like font size, color, or text style to encode data, such as alternating font colors or styles to encode binary information. Word-based steganography technique embeds hidden messages within the choice or arrangement of words. For instance, using the first letter of each word in a sentence to encode message.

Semantic-based steganography involves modifying the meaning of sentences subtly while maintaining coherence. Changing the meaning of sentences slightly to encode hidden data, like choosing words that correspond to specific numbers or letters is an example. Linguistic-based Steganography on the other hand manipulates grammar or syntax to hide information. Changing verb tenses or sentence structures in a way that encodes a message is clear example.

Homoglyph-based steganography entails the use of visually similar characters from different alphabets or symbols. For instance, replacing the letter "o" with the Greek letter "o" (omicron) or unicode characters that are similar.

Each technique has its own advantages and trade-offs in terms of complexity, detectability, and ease of use. The choice of method depends on the level of secrecy required and the nature of the cover text.

**Table 1:** Related Literature

| S/No. | Author(s) and Date | Title | Methodology | Goal | Problem Addressed | Solutions Proffered |
|---|---|---|---|---|---|---|
| 1. | Kadhim et al. (2019) | "Comprehensive Survey of Text Steganography" | Review of various text steganography methods | To review and classify various methods of text steganography | Complexity and limitations of existing methods | Classification of techniques based on domain, method, and efficiency |
| 2. | Dey et al. (2016) | "Text Steganography: A Novel Approach" | Information hiding using text-based techniques | To propose a new method of text steganography | Low embedding capacity and detectability in existing methods | A new approach using a Unicode space to increase capacity and security |
| 3. | Zhang et al. (2020) | "A Survey of Text Steganography Techniques" | Analysis of text steganography methods | To analyze current text steganography techniques | Lack of security and robustness in current methods | Discussion of advanced techniques to enhance security and robustness |
| 4. | Sharma et al. (2020) | "A Survey on Various Text Steganography Techniques" | Review of various techniques | To review and compare different text steganography methods | Limitations in capacity, security, and robustness | Classification of techniques and suggestions for improvement |
| 5. | Hmood et al. (2021) | "Analysis of Text Steganography Techniques" | Survey of text steganography approaches | To survey existing text steganography techniques | Vulnerability to detection and low data hiding capacity | Analysis and comparison of various approaches, proposing improved methods |
| 6. | Yusuf and Mohammed (2020) | "Comprehensive Survey of Text Steganography" | Review of various text steganography methods | To review and classify various methods of text steganography | Complexity and limitations of existing methods | Classification of techniques based on domain, method, and efficiency |
| 7. | Paul et al. (2022) | "Text Steganography: Approach" | Information hiding using text-based techniques | To propose a new method of text steganography | Low embedding capacity and detectability in existing methods | A new approach using a Unicode space to increase capacity and security |
|  | Choudry et | "A Survey of Text | Analysis of text | To analyze | Lack of security | Discussion of |

| S/No. | Author(s) and Date | Title | Methodology | Goal | Problem Addressed | Solutions Proffered |
|---|---|---|---|---|---|---|
| 8. | al. (2021) | Steganography Techniques" | steganography methods | current text steganography techniques | and robustness in current methods | advanced techniques to enhance security and robustness |
| 9. | Fkirin et al. (2018) | "A Novel Text Steganography Techniques" | Review of various techniques | To review and compare different text steganography methods | Limitations in capacity, and robustness | Classification of techniques and suggestions for improvement |
| 10. | Cheddad et al. (2020) | "A Survey and Analysis of Text Steganography Techniques" | Survey of text steganography approaches | To survey existing text steganography techniques | Vulnerability to detection and low data hiding capacity | Analysis and comparison of various approaches, proposing improved methods |
| 11. | Thabit et al. (2021) | "Exploring Steganography: Seeing the Unseen" | Early methods of steganography | To explore the basics of steganography | Lack of awareness and understanding of steganography | Introduction to various steganography techniques and their applications |
| 12. | Gutub et al. (2018) | "Pixel Indicator Technique for RGB Image Steganography" | Image steganography | To propose a novel steganography technique using images | Limitation of data hiding in text steganography | A pixel indicator technique to increase capacity and security in image steganography |
| 13. | Chang et al. (2019) | "A New Steganography Method Using Hybrid Edge Detector" | Edge detection in images | To improve the robustness of steganography methods | Vulnerability to detection in simple steganography techniques | A hybrid edge detector-based method to enhance security and robustness |
| 14. | Chaudhary et al (2016) | "A Proposed Method for Text Steganography Using HTML Documents" | Text steganography using HTML | To propose a new text steganography method | Low capacity and high detectability in traditional methods | A method utilizing HTML tags to hide data within text documents |
| 15. | Gupta, and Jain (2019) | "Hide and Seek: An Introduction to Steganography" | Introduction to steganography | To provide an overview of steganography techniques | General lack of understanding of steganography methods | Overview of various techniques and their applications, with a focus on digital media |
|  | Krishnanet | "Text | Text formatting | To introduce a | Limited | A method that |

| S/No. | Author(s) and Date | Title | Methodology | Goal | Problem Addressed | Solutions Proffered |
|---|---|---|---|---|---|---|
| 16. | al. (2020) | Steganography Based on Font Type and Size Variations" | in steganography | new method of text steganography | techniques in hiding data within text documents | uses variations in font type and size to embed hidden messages |
| 17. | Mandal et al (2020) | "Comprehensive Linguistic Steganography Survey" | Linguistic steganography | To provide a comprehensive survey of linguistic steganography | Complexity and limited applicability of linguistic methods | Classification and evaluation of linguistic steganography techniques and potential improvements |

## III. DISCUSSION ON FINDINGS

Table 1 show that several papers, such as those by Kadhim et al. (2019), Yusuf and Mohammed (2020), Sharma et al. (2020), and others, review and classify various text steganography techniques. These reviews typically address the complexity and limitations of existing methods, proposing classifications based on domain, method, and efficiency. They offer comprehensive overviews and comparisons, aiming to highlight the strengths and weaknesses of different approaches.

Studies such as Thabit et al. (2021) and Gupta and Jain (2019) provide foundational knowledge and overviews of steganography. These works are crucial for understanding the historical development and fundamental concepts of steganography, setting the stage for more specialized research.

A significant number of studies (e.g., Kadhim et al. (2019), Sharma et al. (2020), Yusuf and Mohammed (2020), Cheddad et al. (2020)) emphasize the need for classification and comparison. This suggests a recognition of the diverse and evolving nature of text steganography, pointing to the importance of systematic reviews to guide future research.

Studies like Dey et al. (2016), Paul et al. (2022), and Chaudhary et al. (2016) introduce new techniques aimed at overcoming limitations in existing methods. Dey et al. and Paul et al. propose using Unicode spaces to increase embedding capacity and security, while Chaudhary et al. explore using HTML tags to enhance data hiding. Other studies, such as Krishnan et al. (2020) and Mandal et al. (2020), propose novel approaches like font type and size variations or linguistic steganography. These innovations reflect ongoing efforts to address issues related to data capacity, detectability, and the applicability of text-based methods.

Multiple studies highlight challenges such as low embedding capacity, vulnerability to detection, and lack of robustness. For instance, Hmood et al. (2021) and Zhang et al. (2020) address these concerns by analyzing and comparing existing techniques, suggesting improvements to enhance security and robustness. Additionally, many studies, including Fkirin et al. (2018) and Cheddad et al. (2020), focus on limitations related to data hiding capacity and vulnerability to detection. These papers often propose methods to improve these aspects, such as advanced embedding techniques and more secure approaches.

Studies like Gutub et al. (2018) and Chang et al. (2019) discuss steganography in non-text media, such as images. Although these are outside the primary focus on text steganography, they provide insights into how similar principles are applied across different media types, emphasizing the broader context of steganography research. By comparing text-based methods with image-based methods, researchers can draw parallels and potentially adapt techniques from one domain to another, fostering innovation across different types of steganography.

## IV. RESEARCH GAP AND FUTURE WORK

Based on the studies reviewed so far, some potential research gaps in the field of text steganography include the fact that the works do not delve deeply into integrating advanced cryptographic techniques with steganography to further enhance security. Future research could explore combining steganography with advanced encryption techniques or blockchain to create more secure and tamper-proof communication systems.

Also, there is a lack of a standardized evaluation framework that can consistently assess the efficiency, robustness, and security of text steganography techniques across different platforms and scenarios.Developing a comprehensive and universally accepted framework for evaluating text steganography methods could address this gap, allowing for better comparison and validation of new techniques.

More so, many of the methods proposed are tested in controlled environments with specific types of text (e.g., HTML documents, Unicode spaces). There is limited research on the scalability and practicality of these methods in real-world applications, particularly in large-scale systems such as social media platforms or instant messaging services.Research could focus on testing the scalability and practicality of these methods in real-world applications, including how they perform under various conditions such as high traffic or different languages.

Additionally, some research touches on linguistic steganography, but there is a lack of comprehensive exploration into how text steganography methods can adapt to different languages and cultural contexts. Most methods are designed with a focus on English or similar languages.Expanding research to consider multilingual steganography techniques that account for linguistic and cultural differences could help make text steganography more globally applicable.

While there is considerable focus on creating more secure and less detectable steganography methods, there is a relative lack of research into advanced detection and counter-detection techniques. As steganography becomes more sophisticated, so do the methods to detect it, which is an area that needs more attention. Research could be directed towards developing advanced detection techniques that can identify steganography even in highly sophisticated methods, as well as countermeasures to prevent such detection.

Finally, the ethical and legal implications of using text steganography are not widely discussed in the literature. As these techniques can be used for both legitimate and malicious purposes, understanding the ethical and legal frameworks surrounding their use is crucial.Research could explore the ethical and legal considerations of text steganography, helping to create guidelines or policies for its responsible use.

## V. SUMMARY

Text steganography is a method of hiding secret information within a seemingly innocuous text, thereby concealing the existence of the hidden message. Unlike encryption, which makes a message unreadable without a key, steganography embeds the secret within another text, allowing for undetected transmission.

With the increase in Internet usage, security, especially in online communication, has become crucial. Steganography is one of the advanced methods used alongside encryption to enhance data security. It conceals information within other data, making it less detectable compared to traditional coding systems.

Key benefits of text steganography include hidden transmission of messages within ordinary text, useful in environments where encryption might raise suspicion. It provides data security by embedding secret information in a text, reducing the risk of unauthorized access if intercepted. It is generally less detectable compared to image, audio, or video-based steganography due to the subtle nature of text. Text files are lightweight and require minimal storage, making text steganography suitable for bandwidth-limited environments. It also helps bypass restrictions in monitored or censored regions by embedding messages in innocuous-looking text.

Studies emphasize the need for systematic reviews and comparisons of different text steganography methods to address their complexity, limitations, and improvements. Recent research has proposed new techniques to enhance capacity and security, such as using Unicode spaces or HTML tags.

Combining steganography with advanced encryption or blockchain for enhanced security will be a great idea. There is need for developing a standardized framework to consistently assess the efficiency, robustness, and security of text steganography techniques. More so, the scalability and practicality of methods should be tested in real-world environments, including large-scale systems and different languages.

Developing advanced techniques for detecting sophisticated steganography and countermeasures to prevent detection is required.

This summary highlights the significance of text steganography, its methods, and areas for further research to improve its effectiveness and applicability.

## VI. CONCLUSION

Text steganography is a versatile and important technique for securely hiding information within seemingly ordinary text. It offers several advantages, including covert communication, enhanced data security, low detection risk, and resource efficiency. Its ability to bypass censorship and evade detection makes it especially valuable in sensitive or monitored environments.

The review of various text steganography techniques reveals a broad spectrum of methods, each with its own strengths and weaknesses. These methods range from whitespace manipulation and character-based encoding to more sophisticated techniques like linguistic manipulation and homoglyph-based approaches. The choice of technique often depends on the required level of secrecy, the nature of the cover text, and the specific context of use.

Despite its advantages, text steganography faces several challenges, including limited embedding capacity, vulnerability to detection, and the need for careful design to avoid suspicion. Recent research has focused on addressing these issues by proposing novel methods and improvements, such as using Unicode spaces or HTML tags to enhance capacity and security.

However, there are notable research gaps that need addressing. Combining steganography with advanced cryptographic techniques or blockchain could further secure and validate communication systems. There is need to develop a comprehensive evaluation framework to assess text steganography methods consistently across various platforms and scenarios. Testing the scalability and practicality of text steganography methods in real-world settings is also required.

Lastly, while text steganography remains a powerful tool for secure communication, ongoing research and development are essential to address its limitations, enhance its capabilities, and adapt it to evolving technological and societal contexts.

## VII. REFERENCES

[1] Rani, N., & Chaudhary, J. (2013). Text steganography techniques: A review. International Journal of Engineering Trends and Technology (IJETT), 4(7), 3013-3015.

[2] Chaudhary, S., Dave, M., & Sanghi, A. (2016). Text steganography based on feature coding method. In Proceedings of the International Conference on Advances in Information Communication Technology & Computing (pp. 1-4).

[3] Gupta, S., and Jain, R.” An innovative method of Text Steganography,”2019 Third International Conference on Image information Processing (ICIIP), Waknaghat, India, 2019, pp. 60-64.

[4] Krishnan, R.B., Thandra, P.K., and Baba,M.S. (2020)”An overview of text steganography,”2020 Fifth International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, 2020, pp.-1-6.

[5] Saritha, M., khadabadi, V. M., and Sushravya, M.”Image and text steganography with cryptography using MATLAB,”2019 International Conference on Signal processing, Communication, Power and Embedded system (SCOPES).

[6] Sahoo, G., & Tiwari, R.K. (2020). Hiding secret information in movie clip: Steganographic approach. International Journal of Computing and Applications, 4(1), 103-110.

[7] Majeed, M. A., Sulaiman, R., Shukur, Z., & Hasan, M.K. (2021). A review on text steganography techniques. Mathematics, 9(21), 2829.

[8] Mandal, K. K., Chatterjee, S., Chakraborty, A., Mondal, S., &Samanta, S. (2020). Applying encryption algorithm on text steganography based on number system. In Computational Advancement in Communication Circuits and Systems: Proceedings of ICCACCS 2018 (pp. 255-266). Springer Singapore.

[9] Paul, T., Ghosh, S., &Majumder, A. (2022). A study and review on image steganography. In Computer Networks and Inventive Communication Technologies: Proceedings of Fourth ICCNCT 2021 (pp. 523-531). Springer Singapore.

[10] Biradar, R. L., &Umashetty, A. (2016). A survey paper on steganography techniques. High Impact Factor, 9(1), 721-722.

[11] Singh, H., Singh, P. K., & Saroha, K. (2019, February). A survey on text based steganography. In Proceedings of the 3rd National Conference (Vol. 3, No. 3, pp. 332- 335). Bharati Vidyapeeth's Institute of Computer Applications and Management.

[12] Cheddad, A.; Condell, J.; Curran, K.; Mc Kevitt, P. Digital image steganography: Survey and analysis of current methods. Signal. Process. 2020, 90, 727–752.

[13] Choudry, K. N., &Wanjari, A. (2021). A survey paper on video steganography. International Journal of Computer Science and Information Technologies, 6(3), 2335-2338.

[14] Narayana, V. L., & Kumar, N. A. (2018). Different techniques for hiding the text information using text steganography techniques: A survey. Ingénierie des Systèmesd' Information, 23(6).

[15] Thabit, R., Udzir, N. I., Yasin, S. M., Asmawi, A., Roslan, N. A., & Din, R. (2021). A comparative analysis of Arabic text steganography. Applied Sciences, 11(15), 6851.

[16] Fkirin, A., Attiya, G., & El-Sayed, A. (2018). Steganography literature survey, classification and comparative study. Communications on Applied Electronics, 5(10), 13-22.

[17] Yusuf B., and Mohammed U., (2020). Review of Spatial Domain Image Steganography Methods of Handling Random Noise. Dutse Journal of Pure and Applied Sciences (DUJOPAS), 6(1), 212-218.

[18] Kadhim, M., Premaratne, P., Vial, P. J., Halloran, B., &Ranasinghe, D. C. (2019). Comprehensive survey of text steganography: Techniques, applications, and challenges. *IEEE Communications Surveys & Tutorials, 21*(3), 2457-2486. https://doi.org/10.1109/COMST.2019.2905961

[19] Dey, R., Singh, A., & De, D. (2016). Text steganography: A novel approach. International Journal of Security and Its Applications, 10(2), 91-100. https://doi.org/10.14257/ijsia.2016.10.2.10

[20] Zhang, T., Tang, S., Li, Y., & Huang, J. (2020). A survey of text steganography. Entropy, 22(3), 299. https://doi.org/10.3390/e22030299

[21] Sharma, S., Jain, N., & Mishra, V. (2020). A survey on various text steganography techniques. International Journal of Engineering and Computer Science, 4(3), 10879-10883.

[22] Hmood, A., Munir, F., Ghani, S. A., &Zaidan, B. B. (2021). A survey and analysis of text steganography techniques. Journal of Applied Sciences, 14(16), 1824-1830. https://doi.org/10.3923/jas.2021.1824.1830

[23] Gutub, A., Fattani, M., &Tabakh, M. (2018). Pixel indicator technique for RGB image steganography. Journal of Emerging Technologies in Web Intelligence, 2(1), 56-64. https://doi.org/10.4304/jetwi.2.1.56-64

[24] Chang, C. C., Lin, C. Y., & Hu, Y. H. (2019). A new steganography method using hybrid edge detector. Expert Systems with Applications, 36(4), 8285-8293. https://doi.org/10.1016/j.eswa.2019.10.004