Engineering and Technology Journal e-ISSN: 2456-3358
Volume 11 Special Issue 06 June-2026, Page No.- 1-6
DOI: 10.47191/etj/v11iSPC6.01, I.F. – 8.595

An Artificial Intelligence-Based Framework for Detecting Hidden Information in Stego Images for Secure Digital Communication

Abhinav Vidwans[1], Dr. Manoj Kumar Ramaiya[2]
Research Scholar,Institute of Advance computingSAGE university Indore, MP, India
Professor & Head, Institute of Advance computing SAGE university Indore

ABSTRACT: Cybersecurity and digital forensics face numerous obstacles due to steganography, the practice of concealing a message within a digital image file. However, because of the minimal visible changes in images with the current steganographic techniques, it is difficult to detect the hidden data effectively. The creation of an efficient DL-based steganalysis framework that can enhance the identification of hidden information in pictures in the spatial domain is the primary objective of this study. An enhanced Convolutional Neural Network (CNN) model is proposed and tested on the basis of the BOSS Base 1.01 dataset. The steganographic algorithms used in the study (S-UNIWARD and WOW) were developed to produce stego images having a payload of 0.4 bpp. The proposed CNN model achieved better feature extraction and classification accuracy than the alternatives when asked to distinguish between cover pictures and stego images. The results of the experiments demonstrated that the model was 87.6% accurate for WOW and 84.5% accurate for S-UNIWARD. It was discovered that numerous existing methods, such as ResNet, Alex Net, and CVTStego-Net, performed lower than the benchmark. There was little overfitting and consistent convergence in the validation and training curves. As a whole, the suggested architecture helps with safe digital forensics and cybersecurity applications while also providing a solid, dependable, and efficient solution for image steganalysis.

KEYWORDS*:* Data Hiding, Digital Image Security, Secure Communication, Cybersecurity, Steganography, Digital Forensics, Deep Learning, BOSSBase 1.01.

I. INTRODUCTION

Innovative ideas and technologies have emerged in this digital era, with security being crowned as the paramount concern. In order to prevent unwanted users from accessing the data, it must be kept safe and secure. The first technique of communication was cryptography[1], which included encrypting the message so that only the intended recipient and sender could decipher it [2][3]. The main problem with cryptography was that it made messages more susceptible to decipherment by hackers if they knew they contained secret information [4][5]. The development of steganography allowed for the elimination of this issue [6]. Since steganography used images to conceal data, it was a significant improvement over encryption [7][8]. Steganography is a method of encoding secret data into seemingly innocuous media. To put it another way, it may conceal the presence of a secret communication by concealing it in data that is transmitted or delivered in public[9][10][11]. Users are put at serious risk by steganographic techniques since they may be used to disseminate harmful software or by this kind of malware (so-called stegomalware [12]) to perform tasks such as command and control communications or data leaks.

The most significant issue with identifying picture steganography is that it becomes more difficult as the steganography's effectiveness increases. Visual assaults based on human perceptual irregularities in pictures and statistical approaches that identify abnormalities in the distribution of pixel values are examples of traditional tactics[13][14]. With the rise of AI, a new era has dawned in image steganography detection. AI techniques are better at detecting hidden information than traditional methods, as they learn the subtle changes introduced by steganography. AI is a subset of deep learning, which has yielded remarkable results[15]. Digital image steganography is used for secure communication, and there are growing concerns about security during the transfer of hidden information. This has prompted this research. Conventional methods are very hard to identify the hidden data because modern methods of steganography produce less visual deformation. In recent years, there has been an increasing demand for reliable steganalysis systems because of steganography-related crimes, virus transmission, and illegal data exchange. The study aims to develop an efficient framework with deep learning to improve digital forensics and cybersecurity applications and to facilitate the discovery of hidden information. This paper's main contributions are:

- Created a successful CNN-based steganalysis methodology that can reliably identify hidden information in digital images.

- Improved feature learning and classification performance through the combination of SRM residual preprocessing and multi-scale feature extraction.
- The model's generalizability and robustness were improved by combining dropout, mix pooling, and Batch Normalization techniques.
- Used the BOSSBase 1.01 dataset to show that it could detect recent adaptive steganographic algorithms like WOW and S-UNIWARD.
- Achieved higher accuracy and F1-score compared with several existing deep learning-based steganalysis approaches, highlighting the effectiveness of the proposed architecture.
- Contributed toward secure digital forensics and cybersecurity applications by providing a reliable framework for identifying concealed communication in images.

The suggested work is innovative because it can enhance the performance of image steganalysis for current adaptive steganographic methods like S-UNIWARD and WOW. With fewer misclassifications and increased detection accuracy, the suggested CNN model outperformed several current techniques. In order to improve its reliability for secret data recognition, the study shows that the suggested framework can successfully differentiate between cover and stego images. The resilience and efficiency of the suggested method for digital forensics and cybersecurity applications are supported by the solid experimental findings achieved on the BOSS Base 1.01 dataset.

A. *Organization of paper*

The paper is organized as follows: In Section II, previous research on picture steganography is reviewed. Section III lays out the approach, and Section IV displays the numerical results of an exhaustive performance evaluation campaign. Section V concludes by discussing challenges and future work.

II. LITERATURE REVIEW

This section presents the current literature on steganography techniques for images and steganalysis tools. Moreover, the review reveals the existing gaps in the detection accuracy, feature extraction capability, and robustness against the recent adaptive steganographic algorithms.

Banerjee *et al.* (2026) suggest an MLOps framework that takes drift into account for adaptive content sterilization; this framework enables self-training, robust deployment over extended durations, and continuous monitoring. The experimental results shown by various datasets and steganographic methods validate that MLOps enhance adaptive performance under drift by up to 10%-15% in perceptual quality and 20%-25% in stego removal efficiency [16]. Thapliyal *et al.* (2025) compare the effectiveness of different steganalysis techniques on the ALASKA2 Kaggle dataset. This study starts with basic techniques such as LSB and histogram analysis to set a baseline, and then builds on to more advanced statistical and graphical evaluations. Each method is assessed for its accuracy in detecting modifications indicative of hidden data. Notably, LSB Analysis reveals a 42.52% encoding rate in stego images compared to 4.06% in cover images, and the Chi-Square test shows a 72.21% likelihood of steganographic presence [17].

Dhawan *et al.* (2024) the output is fed into an HFNN that makes use of the direction of alteration and pixel value differences to find the least significant bit replacement of secret data. Utilizing the hidden cameraman picture, the proposed technique achieves a PSNR of 61.963895 and an MSE of 0.041361, according to the experimental data [18]. Melman and Evsutin (2023) provide a research comparing seven metaheuristic optimization techniques for Discrete Fourier Transform (DFT) phase spectrum embedding optimization. In addition to offering error-free embedded data extraction, the new technique stands out due to its high embedding invisibility achieved by metaheuristic optimization. A 2%-6% improvement in PSNR and a 16%-25% improvement in capacity were the results of metaheuristic optimization [19].

Liu, Jiao and Sun (2022) provide an ADM for handling full-resolution feature attention learning and a Feature Enhancement Passing Module (FEPM) for transferring shallow features to deep layers. They created FPNet, a very accurate and compact image steganalysis network with only 0.16 million parameters, by combining these two architectures. Several tests conducted in the same setting demonstrate that approach works better[20].

Research Gap: The dominant research resulted in advancements in steganography detection, embedding quality, and adaptive retraining. There is a lack of generalizability because most approaches only work on certain algorithms or restricted datasets. A comprehensive AI system that integrates efficient computing, flexibility, and accurate hidden information detection has been the subject of very few research efforts. The stability of performance over time and the management of changing steganographic threats are additional issues faced by current techniques. Furthermore, it is still challenging to find a balance between scalability, resilience, and detection accuracy. Secure digital communication systems necessitate an adaptable and intelligent framework for trustworthy stego image detection.

III. METHODOLOGY

The BOSS Base 1.01 dataset is employed to illustrate the complete procedure of the proposed CNN-based steganalysis framework, as illustrated in Figure 1. The input images are first processed using data preparation procedures, which may involve resizing creating stego images, and applying SRM filters. Datasets for training, testing, and validation are created from the pre-processed pictures after normalization. Cover and stego picture classification is the first step in training the suggested CNN model, following dataset compilation. Some classification measures used to measure the model's performance are F1, REC, ACC, and PRE. As a last step in evaluating the suggested method for picture steganalysis, the acquired findings are contrasted with pre-existing models.

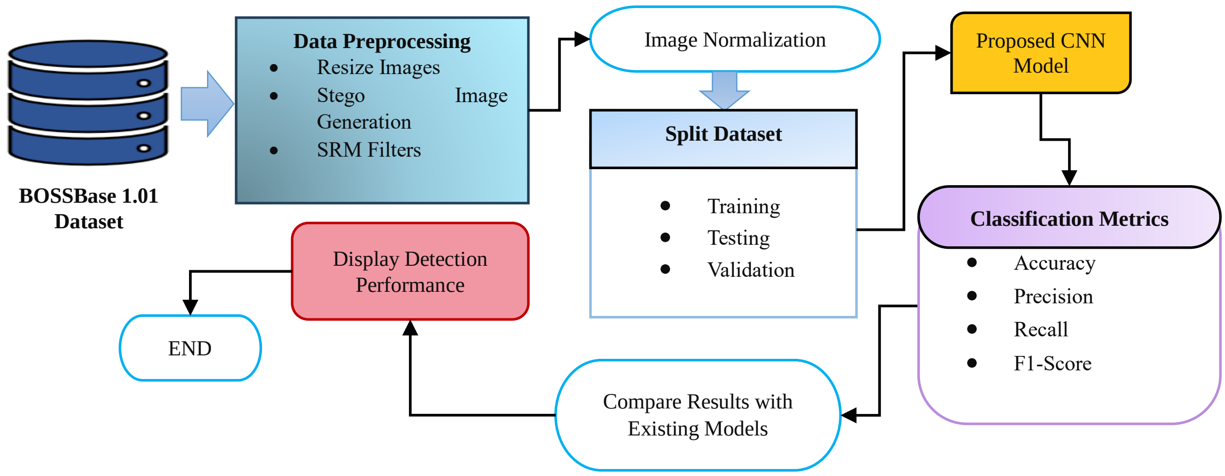


Fig. 1. **Proposed Flowchart for Steganography Detection Each steps of propose flowchart are briefly discussed in next section.**

A. Data Collection and Pre-Processing

This study utilized the freely available Break Our Steganographic System (BOSSBase 1.01) database to train and test the proposed CNN model. This database contains 8-bit grayscale PortableGrayMap (PGM) images. There are 10,000 cover photos in the BOSSBase database, and each one is 512×512 pixels in size. Here are the data preparation processes that were used on the dataset:

Fig. 2. **Sample Images of datasets**

Figure 2 presents sample images from the dataset used for steganalysis experimentation. The images illustrate different versions of the same scene, including original cover images and stego images generated using steganographic embedding techniques. These sample images demonstrate the visual similarity between cover and stego images, where hidden information is embedded without causing noticeable distortion to the human eye. This figure shows how difficult it is to discover hidden information and how important it is to use a strong deep learning model, like CNN, to accurately distinguish between normal and stego images for steganalysis purposes.

- **Image Resizing:** The first images displayed are 512x512 pixels in size, then they are resized to 256x256. This shrinking process minimizes the computational load in the following operations without sacrificing a satisfactory image quality.
- **Stego Image Generation:** Stego images are crafted with a 0.4 bpp payload using the steganographic algorithms S-UNIWARD and WOW.
- **SRM Filters:** The preprocessing stage used a set of 30 Spatial Rich Model (SRM) filters with kernel size 5x5. These filters are employed to remove texture patterns and hidden noise residuals in the images, which are crucial for the detection of hidden information. The SRM filter bank improved the model's capability to learn discriminative features needed for effective steganalysis.

B. Image Normalization

Pre-processing normalizes the picture to make the CNN model more stable and effective. After applying the SRM filter bank and convolution process, Batch Normalization is applied to all the Ffeaturedaps in the network and normalized. This approach improved the gradient flow, accelerated the convergence process, minimized the need for parameter scale, and alleviated the risk of overfitting during training.

C. Data Splitting

Three subsets of the obtained dataset are used for developing and evaluating models. Training uses 8,000 photos, validation uses 2,000, and testing uses 10,000.

D. Proposed Convolutional Neural Network (CNN) Model

To better discover hidden data in spatial-domain pictures, the suggested CNN architecture is created. Separate phases for preprocessing, feature extraction, and classification make up the model. To stabilize training and enhance gradient flow, Batch

Normalization is used in conjunction with the 3TanH activation function. Several layers of convolutional, separable, and depth-wise convolutions—along with layers for ReLU activation, skip connections, mix pooling, and dropout—are integrated into the feature extraction step. In order to concurrently record visual characteristics at several scales, the model makes use of a multi-scale inception module that has parallel convolution filters of sizes 1×1, 3×3, and 5×5. Mix pooling is a hybrid of max pooling and average pooling that improves feature retention and increases the classification accuracy.

Train, validate, and test are the three sections that make up the dataset. During training, the cover-to-stego picture ratio is maintained at 50:50. With a Learning Rate of 0.001, the model was optimized using the Adam Optimizer, which guarantees efficient convergence and stability. The addition of Batch Normalization layers helps with feature distribution and gradient flow, while the addition of Dropout Layers ensures that no features are overfit and improves generalization efficiency. ReLU activation is applied in the feature extraction stage to improve computational efficiency, and the SoftMax classifier is used in the final classification stage for probabilistic prediction of cover and stego images.

E. *Evaluation Criteria*

The proposed CNN model is evaluated using a variety of classification metrics, including ACC, PRE, REC, and F1. These indications are computed using the values of the confusion matrix, which includes metrics for TP, TN, FP, and FN. The following measures are formulated in Equation. (1) to (4)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

The ACC is the ratio of the number of successfully identified images to the total number of predictions. A measure of the ACC rate relative to the overall number of positive samples is known as PRE. A simple definition of the REC would be the ratio of the number of really positive results to the total number of positive results. In terms of harmonic mean, recall and precision make up F1.

IV. RESULTS AND DISCUSSION

A workstation with an Nvidia Repeats-in-Toxins (RTX) 4080 GPU, 64 GB of RAM, a 1 TB hard drive, and an Intel Core i7-12650HX CPU is used to conduct the research. This configuration is strong enough to handle challenging network training activities. The CNN model outperformed the other approach in WOW's classification task, suggesting it is better able to differentiate between stego and cover images (see Table I). The findings reveal that the CNN architecture outperformed both steganography algorithms in discovering hidden information on the WOW dataset.

TABLE I. PERFORMANCE EVALUATION OF CNN MODELS FOR STEGANOGRAPHY DETECTION

Model	Accuracy	Precision	Recall	F1
CNN (S-UNIWARD)	84.5	84.2	84.0	84.1
CNN (WOW)	87.6	87.4	87.2	87.3

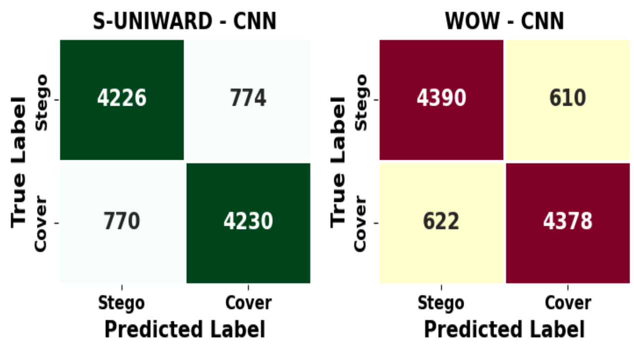
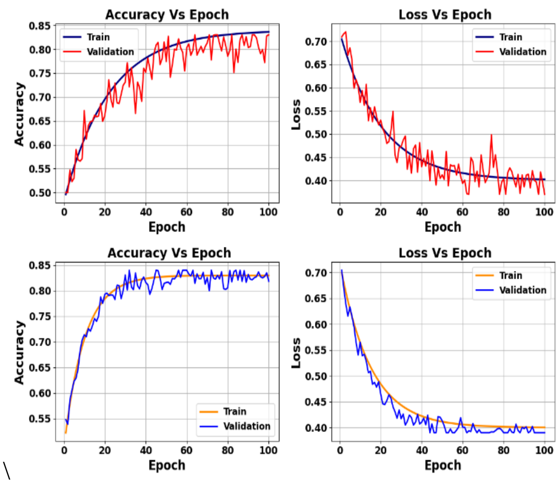


Fig. 3. **Confusion Matrices for CNN-Based Steganography Detection**

Fig. 4.

Figure 3 shows the Confusion Matrices of CNN steganography models tested on images produced by two well-known steganographic methods. The matrices show the outcomes of the Stego and Cover image classifications. While 4230 cover images and 4226 stego images are appropriately categorized in the S-UNIWARD-based model, 774 stego images and 770 cover images are incorrectly classified. The WOW-based algorithm also properly detected 4,390 stego pictures and 4,378 cover images with relatively few misclassifications. The findings show that both steganography strategies are successfully detected by the CNN model, with the WOW method achieving somewhat higher accuracy.



\

Fig. 5. **Training and Validation Performance of CNN Model**

Figure 4 displays the loss curves and validation and training accuracy of the CNN model across 100 training epochs. The left-side plots demonstrate the steady improvement in model accuracy, while the training and validation accuracies converged at 0.82-0.84, indicating effective learning and strong generalization performance. On the right side of the plot, the training and validation loss curves show that the loss values decline continuously over the course of the epochs. It appears that the model trained to a steady convergence with little overfitting, as the validation and training curves are quite near to one another.

A. *Comparative Analysis and Discussion*

Tables II and III make a comparison between the suggested CNN model and current steganalysis methods. ResNet achieved 74.44% accuracy, while AlexNet obtained 76% accuracy and 79% recall. CVTStego-Net reported a 78.225% F1-score. In comparison, the proposed CNN model achieved 87.6% accuracy, showing better overall performance. For WOW and S-UNIWARD at 0.4 bpp, SRM+EC achieved 0.745 and 0.753, IGNCNN obtained 12.3 and 18.4, while SRNet reported 0.933 and 0.929, respectively. The results demonstrate how well and consistently the suggested CNN model detects concealed information encoded using state-of-the-art steganographic methods.

TABLE II. COMPARISON OF CNN MODEL PERFORMANCE WITH EXISTING DEEP LEARNING APPROACHES

Model	Accuracy	Precision	Recall	F1-Score
ResNet [21]	74.44	76.56	70.48	73.41
AlexNet	76	75	79	79
CVTStego-Net [22]	73.697	-	-	78.225
CNN	87.6	87.4	87.2	87.3

TABLE III. COMPARATIVE ANALYSIS BASED ON STEGANALYSIS ALGORITHMS

Algorithms	WOW (0.4 bpp)	S-UNIWARD (0.4 bpp)
SRM+EC [23]	0.745	0.753
IGNCNN [24]	12.3	18.4
SRNet [13]	0.933	0.929
Propose	87.6	84.5

The proposed research has many merits, particularly in enhancing the accuracy and reliability of hidden information detection in digital images, which beneficial for digital forensics, cyber security and secure communication analysis. The model is effectively able to resist the modern steganographic methods and can be used to detect suspicious image content faster and more accurately. Ethically, the research suggests responsible practices in the application of steganalysis in cybercrime investigations, verification of digital evidence, and enhancing data security. The study also ensures ethical compliance by using publicly available datasets such as BOSSBase 1.01 and avoiding misuse of private or sensitive information during experimentation and evaluation.

V. CONCLUSION AND FUTURE WORK

The current digital image steganalysis research has focused on developing a new method based on CNN to extract new information, which is not yet known, from digital images. The WOW and S-UNIWARD steganographic algorithms are used to evaluate the proposed model. The experimental outcomes showed that the proposed CNN model could detect WOW steganography with an acc of 87.6%, a prec of 87.4%, a recall of 87.2%, and an F1score of 87.3%. For S-UNIWARD, the model achieved 84.5% accuracy, 84.2% precision, 84.0% recall, and 84.1% F1-score. Reduced rates of misclassification in cover and stego image categorization are also revealed by the confusion matrix analysis. The results show that the suggested CNN-based design is a strong, effective, and reliable way to find hidden information.

A. *limitation and Future Work*

Important factors that could impact the generalizability of the model in practical settings, such as colour images, compressed image formats, and different payload capacities, are excluded from the suggested research. The model may be enhanced to include colour image steganalysis, various payload levels, and sophisticated adaptive steganographic methods in next research. Lightweight designs and deep learning models based on transformers might be the next big thing for improving detection accuracy and computing efficiency.

REFERENCES

1. B. Singh, M. A. Augie, H. Singh, and T. Banerjee, "Strengthening Modern IAM Authentication with Quantum Cryptography and Anti-Phishing Techniques," *Sarcouncil J. Eng. Comput. Sci.*, vol. 04, no. 10, pp. 17–31, October, 2025, doi: 10.5281/zenodo.17260292.
2. A. Nath, S. Das, R. Sharma, and S. Mandal, "Digital Steganography : A Comprehensive Study on Various Methods for Hiding Secret Data in a Cover file," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 3, pp. 291–300, May 2024, doi: 10.32628/CSEIT24103107.
3. L. Zeng, N. Yang, X. Li, A. Chen, H. Jing, and J. Zhang, "Advanced Image Steganography Using a U-Net-Based Architecture with Multi-Scale Fusion and Perceptual Loss," *Electronics*, vol. 12, no. 18, p. 3808, Sep. 2023, doi: 10.3390/electronics12183808.
4. K. Chinniyan, T. Vani Samiyappan, A. Gopu, and N. Ramasamy, "Image Steganography Using Deep Neural Networks," *Intell. Autom. Soft Comput.*, vol. 34, no. 3, pp. 1877–1891, 2022, doi: 10.32604/iasc.2022.027274.
5. C. Han and T. Xue, "Adaptive network steganography using deep learning and multimedia video analysis for

enhanced security and fidelity," *PLoS One*, vol. 20, no. 6, p. e0318795, Jun. 2025, doi: 10.1371/journal.pone.0318795.
6. A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727–752, Mar. 2010, doi: 10.1016/j.sigpro.2009.08.010.
7. B. Jeganathan, "Exploring the Power of Generative Adversarial Networks (GANs) for Image Generation: A Case Study on the MNIST Dataset," *Int. J. Adv. Eng. Manag.*, vol. 7, no. 1, pp. 21–46, Jan. 2025, doi: 10.35629/5252-07012146.
8. P. M. Kumar and J. A. Renjith, "An image steganographic algorithm on smart mechanism of embedding secret data in images," *Int. J. Electron. Secur. Digit. Forensics*, vol. 8, no. 1, p. 35, 2016, doi: 10.1504/IJESDF.2016.073732.
9. S. K. Davuluri, V. Challagulla, V. Mudapaka, and U. Konka, "Telcoformrix: An AI-Augmented Framework for Declarative and Scalable Provisioning of Real-Time Communication Infrastructure (Work in Progress)," in *2025 IEEE International Conference and Expo on Real Time Communications at IIT (RTC)*, Chicago, IL, USA: IEEE, 2025, pp. 1–4, October. doi: 10.1109/RTC66985.2025.11211725.
10. M. Płachta, M. Krzemień, K. Szczypiorski, and A. Janicki, "Detection of Image Steganography Using Deep Learning and Ensemble Classifiers," *Electron.*, 2022, doi: 10.3390/electronics11101565.
11. M. V. Kumar and P. Pandiaraj, "Content dependent data hiding on GSM full rate encoded image," in *2011 3rd International Conference on Electronics Computer Technology*, IEEE, Apr. 2011, pp. 406–409. doi: 10.1109/ICECTECH.2011.5941781.
12. L. Caviglione *et al.*, "Tight Arms Race: Overview of Current Malware Threats and Trends in Their Detection," *IEEE Access*, vol. 9, pp. 5371–5396, 2021, doi: 10.1109/ACCESS.2020.3048319.
13. O. Kuznetsov, E. Frontoni, K. Chernov, K. Kuznetsova, R. Shevchuk, and M. Karpinski, "Enhancing Steganography Detection with AI: Fine-Tuning a Deep Residual Network for Spread Spectrum Image Steganography," *Sensors*, vol. 24, no. 23, p. 7815, Dec. 2024, doi: 10.3390/s24237815.
14. S. Geetha, V. Kabilan, S. P. Chockalingam, and N. Kamaraj, "Varying radix numeral system based adaptive image steganography," *Inf. Process. Lett.*, vol. 111, no. 16, pp. 792–797, Aug. 2011, doi: 10.1016/j.ipl.2011.05.013.
15. S. Agarwal and K.-H. Jung, "Digital image steganalysis using entropy driven deep neural network," *J. Inf. Secur. Appl.*, vol. 84, p. 103799, Aug. 2024, doi: 10.1016/j.jisa.2024.103799.
16. A. Banerjee, S. Ganguly, I. Mukherjee, and N. Ganguly, "Towards secure digital media: A drift-aware MLOps framework for adaptive stego content sterilization," *Futur. Gener. Comput. Syst.*, vol. 183, p. 108546, Oct. 2026, doi: 10.1016/j.future.2026.108546.
17. K. Thapliyal, S. Mukherjee, R. Gateru, and M. Manchanda, "Decoding Hidden Communication: Image Steganography and Steganalysis via Feature Extraction," in *2025 IEEE Pune Section International Conference (PuneCon)*, IEEE, Dec. 2025, pp. 1–5. doi: 10.1109/PuneCon67554.2025.11378501.
18. S. Dhawan *et al.*, "Secure and resilient improved image steganography using hybrid fuzzy neural network with fuzzy logic," *J. Saf. Sci. Resil.*, vol. 5, no. 1, pp. 91–101, Mar. 2024, doi: 10.1016/j.jnlssr.2023.12.003.
19. A. Melman and O. Evsutin, "Comparative study of metaheuristic optimization algorithms for image steganography based on discrete Fourier transform domain," *Appl. Soft Comput.*, vol. 132, p. 109847, Jan. 2023, doi: 10.1016/j.asoc.2022.109847.
20. J. Liu, G. Jiao, and X. Sun, "Feature Passing Learning for Image Steganalysis," *IEEE Signal Process. Lett.*, vol. 29, pp. 2233–2237, 2022, doi: 10.1109/LSP.2022.3217444.
21. O. A. Alrusaini, "Deep learning for steganalysis: evaluating model robustness against image transformations," *Front. Artif. Intell.*, vol. 8, Mar. 2025, doi: 10.3389/frai.2025.1532895.
22. M. A. Bravo-Ortiz *et al.*, "CVTStego-Net: A convolutional vision transformer architecture for spatial image steganalysis," *J. Inf. Secur. Appl.*, vol. 81, p. 103695, Mar. 2024, doi: 10.1016/j.jisa.2023.103695.
23. Z. Wang, M. Chen, Y. Yang, M. Lei, and Z. Dong, "Joint multi-domain feature learning for image steganalysis based on CNN," *EURASIP J. Image Video Process.*, vol. 2020, no. 1, p. 28, Dec. 2020, doi: 10.1186/s13640-020-00513-7.
24. A. Q. Karamanji, A. S. Ahmed, and A. F. Fadhil, "Comparative Deep Learning Models in Applications of Steganography Detection," *J. Image Graph.*, vol. 12, no. 3, pp. 312–319, 2024, doi: 10.18178/joig.12.3.312-319.