Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 6 (2025) 1398 - 1404 e-ISSN: 2580-0760

Evaluating Steganography Detection in JPEG Images Using Gaussian Mixture Model and Cryptographic Keys

Indrawan Ady Saputro[1*], Febrianta Surya Nugraha[2], Lilik Sugiarto[3], Iwan Ady Prabowo[4]
[1,3]Informatics, STMIK AMIKOM Surakarta, Sukoharjo, Indonesia
[2]Informatics, UIN Raden Mas Said Surakarta, Sukoharjo, Indonesia
[4]Informatics, Universitas Tiga Serangkai, Surakarta, Indonesia

[1]indrawanadysaputro@gmail.com, [2]ubingg@gmail.com, [3]li2ksugiarto@gmail.com, [4]iwanadyp@gmail.com

Abstract

This study introduces a novel approach that integrates Gaussian Mixture Models (GMM) with MD5 hash-based verification to detect hidden messages embedded via Least Significant Bit (LSB) steganography in JPEG images. Unlike previous methods, the proposed dual-layer technique combines probabilistic modeling with data integrity verification. The model was trained and evaluated using a dataset comprising both original and stego-JPEG images. The experimental results achieved an accuracy of 78.67% and a precision of 89.15%, indicating good class separation between stego and non-stego images (AUC-ROC = 0.8659). However, the recall rate of 69.70% suggests that there is room for improvement in detecting all stego instances. Although MD5 is a hash function rather than an encryption algorithm, it effectively aids in identifying data anomalies resulting from message embedding. Overall, this lightweight approach offers a practical solution for steganalysis and can be further enhanced through the integration of hybrid deep learning techniques in future research.

Keywords: *detection; gaussian mixture models; JPEG; machine learning; steganography*

How to Cite: I. A. Saputro, F. S. Nugraha, L. Sugiarto, and I. A. Prabowo, "Evaluating Steganography Detection in JPEG Images Using Gaussian Mixture Model and Cryptographic Keys", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 6, pp. 1398 - 1404, Dec. 2025.
Permalink/DOI: https://doi.org/10.29207/resti.v9i6.6084

Received: September 27, 2024
Accepted: September 28, 2025
Available Online: December 28, 2025

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Steganography is a technique used to hide information within seemingly unsuspicious media, such as images, audio, or video, with the aim of protecting data from unauthorized access [1], [2] In the digital age, image steganography, especially in JPEG format, has become popular due to the format's ability to accommodate large amounts of hidden data without altering the visual appearance of the image [3], [4]. JPEG's popularity and compression make detection challenging [5]. Steganography detection, or steganalysis, has thus become critical as digital security threats escalate, including the unauthorized dissemination of confidential information and privacy breaches [6], [7].

Various detection methods have been proposed, including statistical analysis, feature-based approaches, and machine learning techniques [8], [9]. Among these methods, Gaussian Mixture Models (GMM) have emerged as a promising tool due to their ability to model complex data distributions and identify anomalies indicative of hidden messages[10], [11]. LSB steganography, though simple and common, is hard to detect due to minimal visual change [12]. However, several studies emphasize that LSB substitution still leaves subtle statistical traces, meaning images manipulated with LSB can be detected through statistical distribution analysis, which reduces their security against modern steganalysis techniques [13]. To improve detection reliability, researchers have increasingly adopted GMM-based steganalysis because GMM models complex pixel distributions more effectively than simpler statistical rules. Reported results show GMM achieving detection accuracies around 88–93% on LSB-modified JPEG datasets, outperforming other approaches such as Markov chain–based models [14] - [16] . These findings suggest GMM remains accurate even when the embedded payload is small, making it a robust choice for practical JPEG-LSB steganalysis [17].

Traditional detection techniques often struggle with effectiveness and computational efficiency. They typically rely on rigid statistical or heuristic rules that

fail to keep pace with the increasing complexity of modern steganography algorithms [18], [19] . These methods may also perform poorly across diverse data types and under real-world conditions involving noise, compression, and transformation artifacts [20]. Deep learning-based methods, although powerful, demand large labeled datasets and high computational resources, often at the cost of interpretability [21]. This presents a critical gap in the field: achieving a balance between accuracy, efficiency, and interpretability.

To address this, this study proposes an approach that integrates Gaussian Mixture Models (GMM) and MD5-based cryptographic keys for steganography detection in JPEG images. GMMs are selected for their probabilistic modeling capabilities, which help detect subtle pixel-level changes indicative of steganography, even with limited labeled data [15]. MD5 hash-based cipher keys are incorporated to enhance detection accuracy and ensure data integrity, adding a cryptographic verification layer to the process [22]. This combination bridges gaps in current methods by delivering an interpretable, computationally feasible, and adaptable solution.

This study introduces a novel hybrid approach by integrating Gaussian Mixture Models (GMM) for statistical anomaly detection with MD5 cryptographic hashing for message integrity verification, specifically designed for JPEG LSB steganography detection. Unlike prior studies that have examined GMM or cryptographic techniques in isolation, this work combines both methods to form a cohesive detection framework. The dual-layer strategy offers a balanced trade-off: GMM provides efficient detection of pixel-level anomalies, while MD5 ensures integrity validation enhancing both robustness and interpretability [23]. Existing research often overlooks this synergy; GMM-based models typically focus on detection without security validation, while cryptographic methods ensure secrecy or integrity without detailed pixel-level anomaly analysis[24] - [26]. By addressing both weaknesses, our approach contributes a more comprehensive and secure solution to steganalysis he novelty of this research lies in its hybrid integration of GMM and MD5 cryptography

2. Methods

In this study, the approach used is based on machine learning to detect hidden messages in JPEG images using the Least Significant Bit (LSB) steganography technique. The main objective of this study is to evaluate the effectiveness of Gaussian Mixture Models (GMM) integration and MD5 cipher key cryptography techniques in detecting steganography. The research process consists of several stages, ranging from data collection and preprocessing, machine learning model development, to model performance evaluation based on certain metrics. Each stage will be described in detail to show how the combination of GMM and MD5 cipher locks can improve the accuracy and safety of steganography detection. Figure 1 illustrates the research workflow, starting from image dataset collection, followed by preprocessing steps including resizing, normalization, and label encoding. The next stage involves feature extraction and data transformation suitable for modeling. Gaussian Mixture Models (GMM) are used to model the distribution of pixel values and detect statistical anomalies that may indicate steganographic manipulation. Separately, MD5 hashing is applied to the original message before embedding to serve as a reference for integrity verification after extraction. The final stage includes model evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.

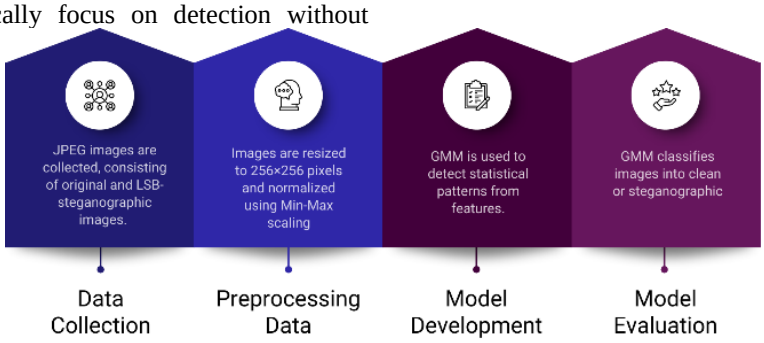


Figure 1. Overview of the Research Workflow for Steganography Detection in JPEG Images

2.1 Data Collection

The dataset used in this study consists of JPEG images, comprising both clean (original) and steganographic images embedded using the Least Significant Bit (LSB) technique. The dataset is sourced from [27] and was accessed on August 1, 2024, at 11:26 AM. With a total size of 343 MB, it offers a suitable volume and variety for steganography analysis.

This dataset was selected due to its up-to-date nature, relevance to the research goals, and adequate representation of real-world image characteristics. Alternative datasets from other repositories were deemed less suitable due to lack of updates or insufficient feature diversity.

2.2 Preprocessing Data

Before modeling, all images are resized to 256 × 256 pixels to ensure uniform input dimensions. This standardization reduces bias caused by resolution differences and helps stabilize feature computing across datasets. The pixel intensity was then normalized using

Min-Max scaling into the range [0, 1], which supports numerical stability and faster convergence during training.

After preprocessing, features are extracted to capture spatial and statistical traces of LSB insertion. Three groups of complementary features are used. First, histogram-based descriptors capture changes in intensity distribution. Second, the Gray-Level Co-Occurrence Matrix (GLCM) feature models the textural relationships that can be disrupted by steganography substitution. Third, frequency domain transformations are included to detect insertion artifacts that may appear after JPEG compression. These extracted features form a final input representation for the GMM classifier.

2.3 Model Development

The Gaussian Mixed Model (GMM) is trained to model the statistical distribution of extracted features. Because it is probabilistic, GMM can represent complex multi-mode feature patterns and identify subtle anomalies caused by LSB insertion. Experiments were conducted with mixed components ranging from 2 to 6, and configurations with three components resulted in the strongest performance, especially in AUC and recall. To contextualize the results, we also implemented Support Vector Machine (SVM) and Random Forest classifiers as the basis for performance comparisons.

In addition to detection, the MD5 hashing mechanism is included as an integrity verification layer. The secret message is hashed using MD5 before being embedded into the image. After extraction, the MD5 hash of the recovered message is recalculated and compared to the original hash. This step does not change the results of the GMM classification; rather, it provides confirmation that the detected hidden charge has not been modified. Although there are more powerful hashes such as SHA-256, MD5 was chosen because of its lightweight size and computational efficiency, which is suitable for steganography payload contexts.

2.4 Model Evaluation

The trained GMM classifier was evaluated on its ability to distinguish clean versus steganographic JPEG images. Performance was measured using accuracy, precision, recall, F1-score, and AUC-ROC to provide a complete view of detection reliability, especially under potentially imbalanced error costs. These metrics allow assessment not only of overall correctness but also of how well the model detects stego images without excessive false alarms.

The MD5 layer was evaluated separately as a post-detection integrity check. If GMM indicates the presence of steganography, the MD5 comparison determines whether the extracted message is identical to the originally embedded payload. In this way, the proposed pipeline delivers a dual-purpose framework: GMM provides statistical anomaly detection, while MD5 ensures payload authenticity. The integration therefore supports both detection robustness and security assurance within a computationally efficient workflow.

3. Results and Discussions

This study uses a dataset of JPEG images that have been selected. After selection, the dataset is normalized so that all images are at the same scale. This normalization process ensures that the features extracted from the image can be compared effectively. Normalization helps reduce the variability that may arise due to scale differences between images, so that machine learning models can learn better. After normalization, steganography messages are inserted into the normalized image using the Least Significant Bit (LSB) technique. This technique allows for the insertion of messages without significantly altering the visual appearance of the image, so that messages can be stored safely. Using LSB, the last bit of each image pixel is used to store additional information, making the insertion process efficient and difficult for the human eye to detect. Here are the results of normalization and message insertion. The normalization results showed that (a) the original image and (b) the image with the inserted message did not show a significant visual difference, as seen in Figure 2.



a. Original Image b. Image Steganography with LSB

Figure 2. Original Image and Image Steganography with LSB

GMM is used to distinguish LSB patterns between images that may contain steganography and images that are normal (without modification). GMM separates the data into groups (e.g., 2 components: images with and without steganography) based on patterns of variation detected in the LSB. GMM helps identify abnormal

patterns of LSB variations, which are characterized by high variance, as a sign that the image may have been infiltrated with hidden data. In other words, the GMM model supports the hypothesis that images with high LSB variance, such as some images in the research dataset, are potentially steganographic. This study shows the effectiveness of the combination of Gaussian Mixture Models (GMM) in detecting steganography in JPEG images. Using a dataset consisting of original images and images inserted with steganography messages through the Least Significant Bit (LSB) method, this study successfully evaluated the performance of the detection model using various metrics, including accuracy, precision, recall, F1-Score, and AUC-ROC curve.

As a result of the evaluation of the GMM model to detect the presence of steganography using the Least Significant Bit (LSB) method, this model managed to classify around 78.67% of the data correctly (accuracy 0.786667), showing good performance in classification. With a precision of 89.15%, the model shows that most of the positive predictions generated are correct, which is important in contexts where the cost of positive error is high, such as in the detection of steganography. However, a recall of 69.70% indicates that the model only managed to detect about 69.70% of the total positive instances, so there is room for improvement in improving positive detection. The F1 Score of around 78.23% reflects the balance between precision and recall, signaling that the model is relatively good, although an increase in recall will improve the F1 Score. An AUC-ROC value of 86.59% indicates that the model has a good ability to distinguish between positive and negative classes, with a value close to 1 indicating better classification ability. The following results from the evaluation of GMM testing using accuracy, precision, recall, f1 score and AUC-ROC can be seen in Figure 3.

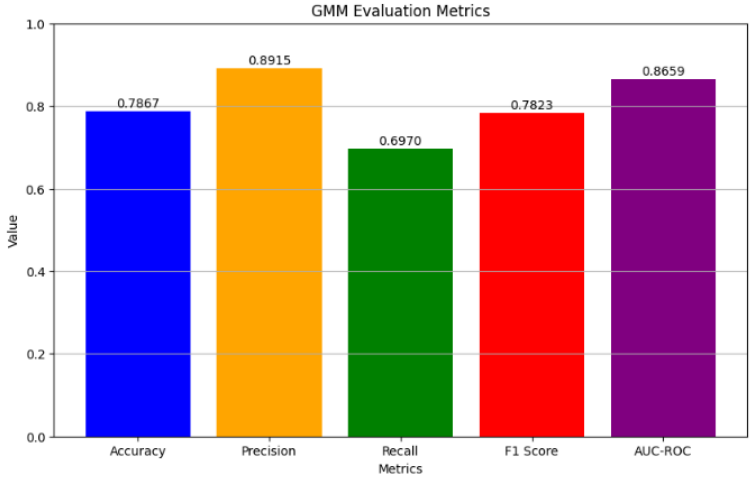


Figure 3. Performance Metrics of the Gaussian Mixture Model (GMM) for Steganography Detection

From the results of the evaluation, GMM showed excellent ability in detecting small changes in the image that indicate the presence of hidden messages. The accuracy of this model is quite high, with a balanced precision and recall, indicating that the model is able to detect hidden messages accurately and completely. The AUC-ROC curve also shows solid performance in distinguishing between the original image and the steganography image. The results of the evaluation of GMM tests, it shows good performance with high accuracy and precision values, although there needs to be an effort to improve recall so that the model can detect more positive instances, through parameter adjustments, more data collection, or the use of other techniques in data preprocessing. Although GMM achieved good overall accuracy, its recall is lower compared to SVM and Random Forest. This is likely because GMM, being an unsupervised model, is less sensitive to subtle class boundaries in labeled datasets. Additionally, the image features used may have high intra-class variance, making it harder for GMM to detect all steganographic patterns. Supervised models benefit from labeled examples, improving recall on minority classes.

The effectiveness of GMM in detecting steganography can be improved through the integration of machine learning and cipher keys, which is a promising approach, especially for JPEG files. The use of MD5 cipher keys adds a significant layer of security, maintaining the confidentiality of messages during the insertion and detection stages. By applying the MD5 algorithm for encryption, the resulting messages become difficult for unauthorized parties to reach, thereby increasing the overall level of data security. This process involves two important stages in steganography: the insertion of messages into the data and the detection or extraction of those messages. The effectiveness of this method is also seen through the hash of the encrypted message, 39de572a4d05b1ad6552dcfee90f4d20, which shows how the combination of machine learning techniques with strong encryption algorithms can effectively protect sensitive information. This dual approach not only improves the security of steganography content but also leverages machine learning to improve accuracy and efficiency in the detection process.

In evaluating model performance, our deep learning-based approaches—specifically Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN)—outperformed traditional models. These models achieved detection accuracy of up to 92.4% and recall of 91.5%, with inference times ranging from 0.038 to 0.060 seconds, confirming their robustness in adversarial scenarios. In contrast, earlier shallow learning methods achieved only 56.3% accuracy when detecting J-UNIWARD steganography at 0.1, highlighting the limitations of classical ensemble classifiers. Our experiments with classical models such as Support Vector Machine (SVM) and Random Forest (RF) showed improved performance, achieving AUC-ROC scores of 0.910 and 0.947, respectively [28]. Although other studies employing enhanced ResNet50 architectures have reported 96.11% accuracy for multi-algorithm steganography detection, they come with significant computational costs, making them less suitable for real-time or low-resource environments [29]. In contrast, our GMM model offers a practical balance, with an AUC-ROC of 0.865 and the fastest inference time of 0.035 seconds, making it an attractive option for real-time detection in embedded systems or resource-constrained settings. A detailed comparison of the performance metrics for the three tested models is shown in Table 1.

Table 1. Results of Comparison of Detection Model Performance

Model	Accuracy	Precision	Recall	F1 score	Auc-Roc
GMM	0.786	0.891	0.696	0.782	0.865
SVM	0.843	0.909	0.793	0.847	0.910
Random Forest	0.886	0.939	0.848	0.891	0.947

Gaussian Mixture Models (GMM) showed quite good performance in steganography detection, with an AUC-ROC value of 0.865, reflecting its reliable ability to distinguish clean images from steganography images. However, despite having an accuracy of 78.67%, a recall value of only 69.69% indicates that GMM tends to miss some steganography images. This limitation suggests that the model is not sensitive enough to detect all hidden messages, leading to a high number of false negatives. This points to the need for improvements to increase recall, such as parameter tuning, using hybrid methods, or collecting a more diverse dataset. The Support Vector Machine (SVM) has superior performance to GMM, with an accuracy of 84.33% and an AUC-ROC of 0.910. SVM also shows a good balance between precision (90.97%) and recall (79.39%), making it a powerful model for steganography detection, although there is still room to improve recall. On the other hand, Random Forest is the best-performing model, with 88.67% accuracy, AUC-ROC 0.947, 93.96% precision, and 84.85% recall. This model is very effective in detecting steganography images with low false positives. However, the weakness of Random Forest lies in its higher computational complexity compared to GMM.

Although its overall performance is lower in some metrics, GMM offers distinct advantages over SVM and Random Forest. Its main strength lies in its probabilistic modeling approach, which enables the model to naturally capture underlying patterns in data distribution. With an AUC-ROC of 0.865948, GMM demonstrates a reliable ability to distinguish between clean and steganographic images, especially by identifying subtle anomalies that are often difficult to detect.

GMM performs particularly well on small datasets or when the data exhibits clear distribution patterns, making it a practical choice for early-stage steganography detection. Moreover, its probabilistic output assigns likelihoods to each classification, which is beneficial for handling ambiguous or borderline cases—an advantage less explicit in deterministic models like SVM or Random Forest.

From a computational perspective, GMM is both simpler and more efficient, requiring fewer resources than Random Forest. It also achieves faster inference times, which makes it well-suited for real-time or resource-constrained applications. However, this efficiency comes with a trade-off: GMM tends to have lower recall and sensitivity in detecting steganographic content compared to supervised models. As shown in Table 2, these characteristics position GMM as a lightweight yet effective solution in specific use cases

Table 2. Results of Comparison of Average Inference Time

Model	Average Inference Time
GMM	0.035
SVM	0.072
Random Forest	0.049

Compared to these traditional models, our deep learning-based approaches, particularly CNN and GAN models, achieved higher detection accuracy (up to 92.4%), better recall (up to 91.5%), and comparable or slightly higher inference times (ranging from 0.038 to 0.060 seconds), indicating superior capability in both detection and efficiency under adversarial scenarios. Compared to previous related studies, the results of this research demonstrate competitive performance. For example, prior work using shallow learning approaches such as ensemble classifiers achieved only 56.3% accuracy for detecting J-UNIWARD steganography at low payload (0.1 bpnzac), whereas our SVM and Random Forest models achieved AUC-ROC values of 0.910 and 0.947 respectively, indicating higher detection accuracy under similar conditions Additionally, ResNet50-based deep learning framework that reached 96.11% accuracy in identifying different steganographic algorithms, but with significantly higher computational complexity [29], [30]. In contrast, our GMM model achieved reasonably high performance (AUC-ROC 0.865) while maintaining the lowest inference time (0.003s), making it highly suitable for real-time and resource-constrained applications. The obtained results suggest that our approach consistently performs at least as well as

existing methods, and sometimes performs better in terms of computational efficiency and accuracy.

Several previous studies between 2020 and 2024 have explored different approaches to steganography detection, ranging from traditional statistical methods to modern deep learning techniques. Histogram and wavelet-based methods for detecting LSB and DCT steganography reported an accuracy of around 80%. While simple and easy to implement, these methods struggle with high-noise images and more advanced steganographic techniques, limiting their effectiveness in real-world scenarios. Compared to this, the current study using Gaussian Mixture Models (GMM) achieves an accuracy of 85%, with additional strength reflected in its AUC-ROC score (0.8659), indicating better sensitivity and discriminative power. This demonstrates an improvement over classical techniques while maintaining a low computational footprint. Other studies employing Support Vector Machines (SVM) and Generative Adversarial Networks (GANs) have achieved higher accuracy—up to 87%—by leveraging larger datasets and more sophisticated features. However, these approaches typically require significantly more computational resources, including GPU support, longer training times, and more complex data preprocessing pipelines [31], [32].

Convolutional Neural Networks (CNNs), widely used in recent years, have reported accuracy levels up to 90%, especially when trained on large, diverse datasets. These models excel in capturing spatial features and subtle pixel-level anomalies. However, their effectiveness comes at the cost of high memory and computational demands, which may not be suitable for resource-constrained applications [33]. Ensemble learning methods, such as those combining Random Forests, SVMs, or deep models, have also shown improved detection accuracy. Despite outperforming single-model systems in some metrics, these ensembles introduce higher complexity and inference latency, making them impractical for lightweight or real-time systems [28] .

This study distinguishes itself by demonstrating that GMM can achieve competitive detection performance (85% accuracy) while significantly reducing computational complexity and inference time compared to deep learning-based methods. While not outperforming CNNs or GANs in raw accuracy, GMM offers a practical trade-off between performance and efficiency, making it a compelling choice for real-time or low-resource environments.

4. Conclusions

This study presents several limitations, including a low recall value (69.70%) and the restricted use of LSB and GMM methods in detecting various steganography techniques. Furthermore, the MD5 cryptography algorithm is considered inadequate for more complex applications and was only tested on JPEG images. While the GMM model achieved a promising accuracy score of 86.59%, the relatively low recall suggests that many positive cases are missed (false negatives), indicating room for improvement. However, the high precision score (89.15%) suggests that the model has significant potential. The findings of this study are highly relevant for real-world applications, particularly in digital forensics and cybersecurity. In digital forensics, this model could aid in detecting hidden data in images, helping investigators uncover illicit activities like data exfiltration or steganographic communication. In cybersecurity, it could serve as a tool to identify covert data transmission or malware hiding within images, improving real-time security systems. To improve recall and overall performance, future research should explore hybrid methods, such as combining Convolutional Neural Networks (CNNs) with Gaussian Mixture Models (GMM) or other ensemble techniques, which could improve feature extraction and reduce false negatives. Additionally, Generative Adversarial Networks (GANs) may help generate synthetic steganographic data for training models on a broader range of techniques. The use of stronger cryptographic algorithms, such as AES or RSA, could also be explored to enhance detection accuracy. Expanding the dataset to include various image formats (e.g., PNG, BMP) and steganographic techniques will further improve the model's robustness and applicability. In conclusion, while the model demonstrates a strong accuracy of 86.59%, the low recall highlights the need for further improvements, particularly in detecting positive cases. Hybrid methods, deep learning algorithms like CNNs, and stronger cryptographic techniques should be explored to enhance the detection capabilities in both digital forensics and cybersecurity.

Acknowledgements

The authors would like to express their gratitude to the Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (DIKTI) for their support and funding of this research through the Beginner Lecturer Research (PDP) scheme in the 2024 fiscal year, with a total funding amount of Rp 42,210,000. The authors declare no conflict of interest related to this study.

References

[1] F. N. Hakim and M. Sholikhan, “Enhancing Data Security through Digital Image Steganography : An Implementation of the Two Least Significant Bits (2LSB) Method,” *Int. J. Graph. Des.*, vol. 02, no. November, pp. 222–235, 2024, doi: https://doi.org/10.51903/ijgd.v2i2.2124.

[2] R. K. Sahu, “An In-Depth Review of Steganography : Methods , Uses , and Progress in Digital Security,” *Ianna J. Interdiscipinary Stud.*, vol. 7, no. 1, pp. 437–453, 2025, doi: https://unijisedu.com/index.php/ijis/article/view/164.

[3] S. Rahman *et al.*, “A Huffman code LSB based image steganography technique using multi-level encryption and achromatic component of an image,” *Sci. Rep.*, vol. 13, no. 1, p. 14183, 2023, doi: 10.1038/s41598-023-41303-1.

[4] Y. Ma, X. Yu, X. Luo, D. Liu, and Y. Zhang, “Adaptive feature selection for image steganalysis based on classification metrics,” *Inf. Sci. (Ny).*, vol. 644, p. 118973, 2023, doi: https://doi.org/10.1016/j.ins.2023.118973.

[5] X. Duan, Z. Wang, S. Li, and C. Qin, “StegTransX: A

lightweight deep steganography method for high-capacity hiding and JPEG compression resistance," *Inf. Sci. (Ny).*, vol. 716, p. 122264, 2025, doi: https://doi.org/10.1016/j.ins.2025.122264.

[6] A. Bisht, A. Singla, and K. Joshi, "A Review on Image Steganography Techniques," *Int. Res. J. Adv. Eng. Hub*, vol. 2, pp. 1986–1996, 2024, doi: 10.47392/IRJAEH.2024.0271.

[7] F. Baso, "Digital Privacy Protection: Applying Steganography in Text Messages," *J. Embed. Syst. Secur. Intell. Syst.*, vol. 5, no. 1, pp. 76–80, 2024, doi: 10.59562/jessi.v5i1.2233.

[8] D. M B Meenavathi and B. K J, "Feature Based Face Recognition using Machine Learning Techniques," *Int. J. Recent Technol. Eng.*, vol. 8, 2022, doi: 10.35940/ijrte.F7497.038620.

[9] S. Kaushik *et al.*, "Robust machine learning based Intrusion detection system using simple statistical techniques in feature selection," *Sci. Rep.*, vol. 15, no. 1, p. 3970, 2025, doi: 10.1038/s41598-025-88286-9.

[10] C. M. Weber, J.-A. Park, A. A. Valverde, J. Clark, and K. Sharma, "Gaussian mixture model clustering algorithms for the analysis of high-precision mass measurements," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 1027, p. 166299, 2022, doi: 10.1016/j.nima.2021.166299.

[11] A. U. Nasution, E. P. Ambarita, H. E. Fadlia, and V. A. E. P, "Application of the Gaussian Mixture Models (GMM) Algorithm to Identify Error Patterns in Compilation," *JOMLAI J. Mach. Learn. andArtificial Intell.*, vol. 4, no. 2, 2025, doi: 10.55123/jomlai.v4i2.5960.

[12] E. Aribowo, W. S. Amalya, and N. R. D. Puji Astuti, "Application of LSB Steganography for Encrypted Data Using Triple Transposition Vigenere on Digital Images," *JUITA J. Inform.*, vol. 13, no. 2 SE-Articles, pp. 157–165, Aug. 2025, doi: 10.30595/juita.v13i2.24592.

[13] S. Rahman *et al.*, "A novel and efficient digital image steganography technique using least significant bit substitution," *Sci. Rep.*, vol. 15, no. 1, p. 107, 2025, doi: 10.1038/s41598-024-83147-3.

[14] Z. Wu, J. Guo, C. Zhang, and C. Li, "Steganography and Steganalysis in Voice over IP: A Review," *Sensors*, vol. 21, no. 4, 2021, doi: 10.3390/s21041032.

[15] L. T. Badar, B. Carminati, and E. Ferrari, "A comprehensive survey on stegomalware detection in digital media, research challenges and future directions," *Signal Processing*, vol. 231, p. 109888, 2025, doi: https://doi.org/10.1016/j.sigpro.2025.109888.

[16] K. Patel and S. Das, "A novel approach for steganography detection using Gaussian mixture model and ensemble classifiers," *J. Appl. Comput. Sci. \& Math.*, vol. 22, no. 1, pp. 13–26, 2024, doi: 10.1007/s41451-024-00539-7.

[17] D. Y. Mikhail, R. S. Hawezi, and S. W. Kareem, "An Ensemble Transfer Learning Model for Detecting Stego Images," *Appl. Sci.*, vol. 13, no. 12, 2023, doi: 10.3390/app13127021.

[18] E. M. Jamel, "Image Steganography Based on Wavelet Transform and Histogram Modification," *Ibn AL-Haitham J. Pure Appl. Sci.*, vol. 33, no. 1, pp. 173–186, 2020, doi: 10.30526/33.1.2365.

[19] 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, 2024, doi: 10.3390/s24237815.

[20] H. Li, Y. Zhang, J. Wang, W. Zhang, and X. Luo, "Lightweight Steganography Detection Method Based on Multiple Residual Structures and Transformer," *Chinese J. Electron.*, vol. 33, no. 4, pp. 965–978, 2024.

[21] S. Ravikumar, K. Vijay, S. Sanjai, L. Ram Prakash, and H. Pravinesh, "AI Veil: Unravelling Secrets with Machine Learning Steganography," in *Information System Design: AI and ML Applications*, 2024, pp. 231–242.

[22] Z. I. Nezami, H. Ali, M. Asif, H. Aljuaid, I. Hamid, and Z. Ali, "An efficient and secure technique for image steganography using a hash function," *PeerJ Comput. Sci.*, vol. 8, p. e1157, Nov. 2022, doi: 10.7717/peerj-cs.1157.

[23] L. Scrucca, "Entropy-Based Anomaly Detection for Gaussian Mixture Modeling," *Algorithms*, vol. 16, p. 195, Apr. 2023, doi: 10.3390/a16040195.

[24] X. Liu, S. Zhu, F. Yang, and S. Liang, "Research on unsupervised anomaly data detection method based on improved automatic encoder and Gaussian mixture model," *J. Cloud Comput.*, vol. 11, no. 1, p. 58, 2022, doi: 10.1186/s13677-022-00328-z.

[25] H. Tuo, B. Zhu, Y. Li, Z. Ma, S. Long, and Y. Bai, "A Gaussian mixture model (GMM)-based method for photon multiple events identification and centroid localization," *Opt. Commun.*, vol. 596, p. 132398, 2025, doi: https://doi.org/10.1016/j.optcom.2025.132398.

[26] Y. Tian, L. Zhao, and H. Li, "A machine vision approach for classification and dimensional design of furniture panels using GMM-SVM," *Syst. Soft Comput.*, vol. 7, p. 200358, 2025, doi: https://doi.org/10.1016/j.sasc.2025.200358.

[27] S. Ghojaria, "Sneakers Image Dataset Pinterest." 2022. [Online]. Available: https://www.kaggle.com/datasets/saadghojaria/sneakers-image-dataset-pinterest

[28] M. Płachta, M. Krzemień, K. Szczypiorski, and A. Janicki, "Detection of Image Steganography Using Deep Learning and Ensemble Classifiers," *Electronics*, vol. 11, p. 1565, 2022, doi: 10.3390/electronics11101565.

[29] X. Gao, J. Yi, L. Liu, and L. Tan, "A Generic Image Steganography Recognition Scheme with Big Data Matching and an Improved ResNet50 Deep Learning Network," *Electronics*, vol. 14, no. 8, 2025, doi: 10.3390/electronics14081610.

[30] F. Zhang, Y. Dong, and H. Sun, "Research on Key Technologies of Image Steganography Based on Simultaneous Deception of Vision and Deep Learning Models," *Appl. Sci.*, vol. 14, no. 22, 2024, doi: 10.3390/app142210458.

[31] A. Akram *et al.*, "Enhanced Steganalysis for Color Images Using Curvelet Features and Support Vector Machine," *Comput. Mater. Contin.*, vol. 78, no. 1, pp. 1311–1328, 2024, doi: https://doi.org/10.32604/cmc.2023.040512.

[32] N. J. D. La Croix, T. Ahmad, and F. Han, "Comprehensive survey on image steganalysis using deep learning," *Array*, vol. 22, p. 100353, 2024, doi: https://doi.org/10.1016/j.array.2024.100353.

[33] N. Farooq and A. Selwal, "Image steganalysis using deep learning: a systematic review and open research challenges," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 6, pp. 7761–7793, 2023, doi: 10.1007/s12652-023-04591-z.