Int. j. adv. multidisc. res. stud. 2026; 6(2):914-916

Received: 12-02-2026
Accepted: 22-03-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Design and Development of a Machine Learning Model for Detecting Data Hidden Techniques

[1] Bwalya Richard, [2] Nsama Lameck

[1] Department of ICT, School of Engineering, Information and Communications University, Lusaka, Zambia
[2] Department of ICT, Information and Communications University, Lusaka, Zambia

DOI: https://doi.org/10.62225/2583049X.2026.6.2.6047
Corresponding Author: **Bwalya Richard**

Abstract

Data security is very important when sensitive data are transmitted over the Internet. Steganography and steganalysis techniques can solve the problem of copyright, ownership, and detection malicious data. Steganography is to hide secret data without distortion and steganalysis is to detect the presence of hidden data. This study focuses on the design and development of a machine learning model for detecting data hiding techniques such as steganography, encryption-based obfuscation, and covert channel embedding. Several machine learning models including SVM, Random Forest, and CNNs were evaluated. The results demonstrate that machine learning-based steganalysis enhances forensic capabilities and cybersecurity investigations.

Keywords: Anti-Forensic Tools, Steganography, Encryption, Obfuscation, Cybercrime, Digital Forensics

Introduction

Machine learning has emerged as a transformative technology across various domains, including cybersecurity and digital forensics. Techniques for concealing data, such as steganography and encryption, pose challenges to investigators. This study aims to develop a machine learning-based detection model capable of identifying hidden data techniques.

The design and development of ML models for detecting hidden data techniques or digital evidence require a systematic approach that integrates both supervised and unsupervised learning methods. Supervised learning algorithms, such as decision trees and support vector machines, are effective for identifying specific patterns in labeled datasets. Meanwhile, unsupervised techniques like clustering and anomaly detection excel at discovering unexpected behaviors in unlabeled data. These approaches are pivotal in addressing challenges like identifying sensitive attributes or detecting deception in large datasets. The choice of algorithm often depends on the nature of the data and the specific application domain [1].

Materials and Methods

This study employed a mixed research design incorporating qualitative, quantitative, and experimental methods. Public datasets containing clean and steganographic images were used. Feature extraction techniques and classifiers such as SVM, Random Forest, and CNN architectures were implemented using Python-based frameworks.

Research Design

The research follows qualitative, quantitative and experimental design, focusing on data-driven analysis to assess the effectiveness of the proposed steganalysis method [2]. The choice of this design is justified by the need for empirical evidence in evaluating detection performance. The study involves dataset preparation, feature extraction, classifier training, and performance assessment.

Descriptive Design: This will allow for a thorough exploration and documentation of existing data hiding techniques and the machine learning models used to detect them. The review of existing techniques will help in designing a model that incorporates current best practices while addressing any identified gaps.

Experimental Design: Given that this study involves the creation and evaluation of a machine learning system, an

experimental approach is necessary. This will enable testing of different machine learning algorithms, performance comparisons, and evaluation of the system's effectiveness in detecting hidden data in various media formats.
This combined approach ensures both a deep understanding of the problem space (through descriptive analysis) and practical, real-world applicability (via experimental testing of models).

Results and Discussion

The experimental results showed that CNN-based models achieved approximately 75% accuracy in detecting steganographic images. The findings indicate improved adaptability and robustness compared to traditional detection methods.
The results help assess the model's performance and practical applicability. Data hiding is the process of embedding information into a noise-tolerant signal such as a piece of audio, video, or image. Digital watermarking is a form of data hiding where identifying data is robustly embedded so that it can resist tampering and be used to identify the original owners of the media. Steganography, another form of data hiding, embeds data for the purpose of secure and secret communication. According to the test accuracy of the model for detecting if the image is stego was 75.13% which means the image has got hidden message embedded in it and that the image is not clean.

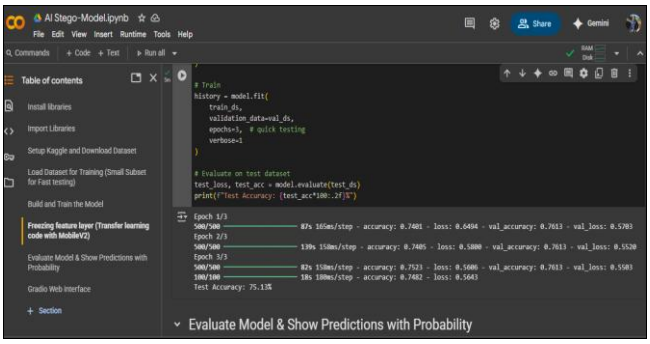
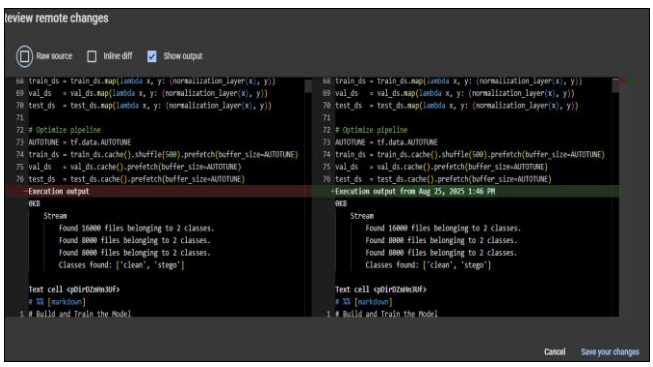


Fig 1: Test Accuracy of the model

The dataset that was used contained four classes but was later on summarized into two classes which only comprised of clean and stego images from kaggle for model testing purposes. The following image illustrates the two in the building and training of the model cell:



Two classes
The implementation guaranteed data integrity, reduce perceptual distortions, and preserve the cover object's quality. Technology Selection: Due to its extensive end-to-end capabilities, suitability for web application development, integration with Python libraries, and effective frontend and backend technology integration, Google colab environment framework was chosen as the basis for the web application performing steganography.
The machine learning model was trained and tested using datasets containing both clean and modified content. It achieved strong performance metrics including:
The 0.85% precision under clean images meaning its correct 85% times meaning 15% of the images labeled clean are actually stego. When the recall is 80% and above it means there are fewer false alarms on innocent images, 0.77 and 0.75 f1 scores suggests stable performance.
Accuracy, precision, recall and F1 score percentages.

Table 1: Classification report on the model

Classification Report				
	Precision	**Recall**	**F1 - Score**	**support**
Clean	0.75	0.77	0.75	398
Stego	0.75	0.73	0.74	1202
accuracy			0.75	1600
Macro avg	0.38	0.50	0.43	1600
Weighted avg	0.56	0.75	0.64	1600

Model Summary – EfficientNet-B5 Architecture

Table 2: Model Summary

Layer (type (var_name))	Input Shape	Output Shape	Trainable
EfficientNet (EfficientNet)	[32, 3, 244, 244]	[32, 1000]	True
Sequential (features)	[32, 3, 244, 244]	[32, 2048, 8, 8]	True
AdaptiveAvgPool2d (avgpool)	[32, 2048, 8, 8]	[32, 2048, 1, 1]	--
Sequential (classifier)	[32, 2048]	[32, 1000]	True

Conclusion

The study confirms that machine learning models, particularly CNNs, are effective in detecting hidden data. The proposed system strengthens digital forensic investigations and supports cybersecurity applications. As digital connectivity expands, the need for secure and discreet communication has never been more critical. Among the many techniques developed to safeguard sensitive information, image steganography has gained significant attention. This method involves embedding secret data within digital images in a way that remains imperceptible to the human eye. Unlike cryptography, which encrypts a message but does not hide its existence, steganography conceals both the message and its presence, making it particularly valuable in scenarios where discretion is essential.

Acknowledgement

The author acknowledges the Almighty God, the supervisor Mr. Nsama Lameck, and the Management of the Information and Communications University.

References

1. Skrzypacz J, Bieganowski M. The influence of micro grooves on the parameters of the centrifugal pump impeller. International Journal of Mechanical Sciences. 2018; 144:827-835.

2. Rodr M. Blind Steganalysis Method for Detection of Hidden Information in Images by Master in Computer Science Advisors, 2013.
3. Chang K, Du C, Chuan S, Gong Z, Du C, Chuan S. ArchNet : A Data Hiding Design for Distributed Machine Learning Systems.
4. Parikshitee Batwal A, Kulkarni D. A Research Paper on Image Steganography: Comparative Analysis of Traditional and Deep Learning Techniques. www.irjmets.com @International Res. J. Mod. Eng. 2025; 6:1686-1692 [Online]. Available: www.irjmets.com
5. Jung K. A Study on Machine Learning for Steganalysis A Study on Machine Learning for Steganalysis, May 2019. Doi: 10.1145/3310986.3311000
6. Blomquist H, Blomquist H, Möller J. Anomaly detection with Machine learning Quality assurance of statistical data in the Aid community Anomaly detection with Machine learning, 2015.