Steganography Detection Using Convolutional Neural Networks: A Deep Learning Approach

Saravana Gokul G
Department of CSE
Rajalakshmi Engineering College
Chennai, India
sharavana.ssg@gmail.com

Deepak Kumar K
Department of CSE
Rajalakshmi Engineering College
Chennai, India
kdeepak.srmit@gmail.com

Senthil Pandi S
Department of CSE
Rajalakshmi Engineering College
Chennai, India
senthil.ks@gmail.com

Kumar P
Department of CSE
Rajalakshmi Engineering College
Chennai, India
kumar@rajalakshmi.edu.in

Monika S
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701166@rajalakshmi.edu.in

Neha M U
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701178@rajalakshmi.edu.in

***Abstract*—Steganography is a method of concealing any message or information within digital images, making it challenging to detect hidden messages. This study focuses on developing a deep learning-based approach for the detection of steganography with the help of Convolutional Neural Networks (CNNs). We implement and compare the performance of LeNet-5 and AlexNet architectures in classifying images as either clean or stego. The dataset comprises images embedded using the Least Significant Bit (LSB) technique. The models are then trained with augmented image data and then evaluated using various accuracy metrics. Experimental results indicate that AlexNet, with its deeper architecture, achieves higher accuracy than LeNet-5 in identifying stego-images. The findings demonstrate the potency of deep learning in automated steganalysis, highlighting the potential for CNNs in cybersecurity applications.**

Keywords: Steganography Detection, Deep Learning, CNN, LeNet-5, AlexNet, Steganalysis

I. INTRODUCTION

With the growth in technology, steganography has gained a lot of traction lately. This can be pinned to the fact that data can effectively be hidden in plain sight. While traditional encryption techniques focus on scrambling the content in such a way that it can only be accessed with a key, steganography aims to hide the text altogether. Digital images are popular medium for steganographic techniques as they offer high resolution with little to no loss in quality. This allows for the easy embedding of concealed information without the risk of the image being altered too much. As steganographic techniques develop, the most confusing part of using those techniques will be the extraction of the hidden messages.

Common classical detection methods make use of statistics, visual inspection, or a combination of both. However, these approaches are often inadequate for sophisticated methods or large volumes of data. Consequently, there has been a shift towards more advanced deep learning and machine learning methods that are able to automate stego-image detection. CNNs have been very promising in the classification of image tasks, notably for automatic image indexing and retrieval, and more recently, for steganography detection. This report analyses the performance of two popular CNN systems, LeNet-5 and AlexNet, for steganography detection. The models are trained on a set of images containing the LSB embedded pictures. We aim to demonstrate that deep learning models are able to automatically and accurately separate stego images from clean images, which will aid in the automatic detection of steganography within digital images.

II. LITERATURE SURVEY

Eslam M. Mustafa [1] The authors recommend a new CNN based steganalysis technique that improves detection accuracy through the use of variable batch sizes and GPU processing. This modern technique is more efficient than the existing IGNCNN model because it employs parallel data and model processing. Yu Sun [2] This novel study proposes a quantitative deep learning approach to steganalysis using YeNet for feature detection and embedding rate in LSB matching during micro embedding. The outcomes of this study suggest that, especially at low embedding rates, this technique produces better results than other methods. Rishit Agrawal [3] The research focuses on the steganographic capacity of several machine learning algorithms based on bits that can be altered with minimum impact on their performance. Surprisingly, InceptionV3 and other similar models were found to have a very high classification accuracy even when massive amounts of hidden data were embedded in them. Dan Wang [4] Authors provide a unique approach for integrating data hiding using Adversarial perturbations and GANS for enhanced security and privacy of digital images. The method achieves an effective form of data hiding without damaging the image quality by employing a genetic algorithm to enhance the strength of the attack. Yen-Ting Chen [5] LeNet-based CNN to analyze different image stego forms which encompass DCT, histogram and LSB stego-images. Deep learning techniques enable a accurate detection of DCT images which shows the capabilities of deep learning

in steganalysis. Ching-Chun Chang [6] The paper presents a method for concealing messages by placing them in game movements through the implementation of multi-agent reinforcement learning. The encoder uses deceptive normal behaviors to hide encoded messages and the observer will learn how to extract these messages using a novel method of covert communication. Magdy Sahar [7] CNNs are used in DeepSteg, a proposed deep learning-based video steganography method that embeds and extracts concealed messages from video frames. The method is a potential solution for safe digital security since it guarantees strong embedding capacity and resilience against steganalysis attacks. [8] This research looks into a generative steganography method which embeds secret data in the moves of Chinese Chess games without altering original media files. Direct generation of stego-carriers creates a security system which provides resistance to attacks by steganalysis programs. Sachin Dhawan [9] Examines numerous data security methods in steganography through detailed classification of techniques regarding their strength and storage capabilities and resistance. The analysis presents suggested methods to enhance stego-image security standards while preserving excellent visual quality of images. Abdullah Alenizi [10] reviews how combining RSA and Blowfish and Hash-LSB algorithms enhances the security of image steganography systems. by uniting cryptographic systems with steganographic processes the research shows that protection gains occur alongside reduced image distortion. May Alanzy [11] The authors presents a dual encryption method that applies AES together with Blowfish for protecting key material within key images. The experimental findings demonstrate high PSNR values which prove that concealed information maintains protection through unaltered image resolutions. Al Hussien S. Saad [12] The proposed approach relies on machine learning combined with OMR to enable message mapping onto OMR bubble sheets without using any form of cover. The approach achieves security enhancement and more robustness against attacks through its avoidance of pixel modifications. Biswarup Ray [13] developed an edge detection framework using deep learning to boost image steganography through increased bit insertion in edge regions. This proposed method raises the embedding capacity without deteriorating image quality resulting in superior outcomes than traditional steganographic methods. M. R. Islam [14] A modified LSB image steganography technique with filtering and AES encryption for security improvements is discussed in the paper. The research outcome demonstrates enhanced security and image quality through its higher PSNR value alongside lower MSE value in comparison to other steganography solutions.Marghny H. Mohamed [15] This research demonstrates how Optimal LSB outperforms Simple LSB by minimizing image distortion through LSB matching alongside genetic algorithm implementation thus achieving superior PSNR results for secure transmission.

III. METHODOLOGY

The research methodology examines deep learning model effectiveness in identifying stego-images through CNNs. The approach several involves crucial stages from collection of data, preprocessing, model selection, training, evaluation to testing. Below is a detailed description of each step involved in the methodology.

A. Data Collection:

The dataset includes clean as well as stego images which were created through the application of Least Significant Bit (LSB) steganographic method. These images are collected from public repositories and synthetically created by incorporating concealed information into clean images using LSB. After data collection the dataset is divided into Stego-LSB Images and Clean Images categories.

B. Preprocessing:

Data standardization takes place in the preprocessing step to make images appropriate for model training. The first step requires each image to reach a standard 227x227 pixel size because AlexNet needs this dimension for input processing. The image pixel values get normalized into the [0, 1] interval by dividing each pixel value by 255. Such standardization helps the model work more efficiently since it maintains uniformity in incoming data scale.

$$X_{normalised} = \frac{X}{255}$$

C. Data Augmentation:

The training set undergoes a data augmentation to increase its diversity and prevent overfitting. Data augmentation tactics including random rotations combined with horizontal flips and image shifts in vertical and horizontal directions build variation in the dataset. The methodology introduces different scenarios to the model which enhances its capacity to generalize and achieve better results when processing new data.

D. Model Selection:

For this study, two well-known CNN architectures have been used to differentiate the images as stego or clean image: LeNet-5 and AlexNet. They are chosen because they have been shown to be successful in image classification tasks, and the systems are thus simpler enabling a fair comparison of performance with regards to steganography detection.

1) *LeNet-5:* LeNet 5 was a comparatively straightforward convolutional and pooling layers followed by fully connected layers are employed for handwritten digit recognition. Although it is straightforward, LeNet-5 has been proven to work well for images classification tasks, especially in small datasets.

2) *AlexNet:* A more complex architecture, the one that's known for the depths in the layers, the big convolutional filters used, and the usage of the ReLU activations. Since AlexNet is designed to address the issues with image classification of large and complex datasets, it has been successfully adopted in several image recognition tasks.

E. Training and Optimization:

Once the dataset is ready, the model is tained further. The purpose of this training process is to feed the training images into the models and updating the model weight according to the computed error. Over 50 epochs, the models are trained with batch size of 32. The models are formed with the use of the

Adam optimizer, a well known adaptive learning rate optimization tool in deep learning.

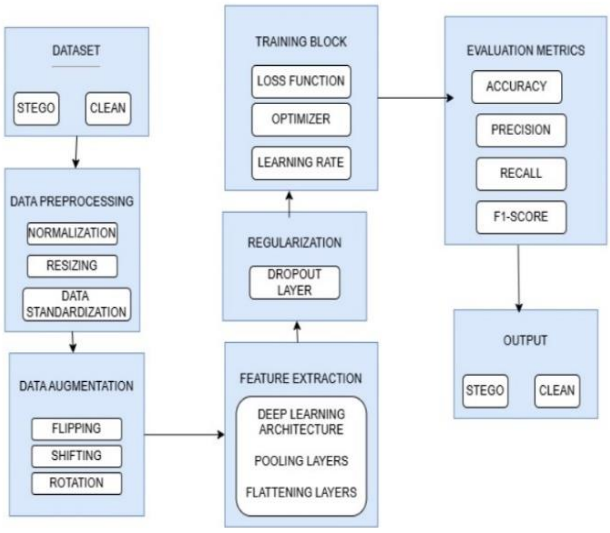

Fig. 1. System Architecture

As the task here is a binary classification problem, we use binary cross entropy loss function for training. Early stopping is used during training to prevent overfitting. After a given amount of epochs, the training process ends if the validation loss does not improve, so it saves the computational resources and does not learn some irrelevant patterns.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}(y_i = \log(p_i) + (1 - y_i)\log(1 - p_i))$$

F. *Model Evaluation:*

The models are then evaluated on several performance metrics after training to ensure they are working as intended. The models' performances are estimated with precision, accuracy, F1-score and recall to precisely distinguish images, identify stego images, and strike a balance between false positives and false negatives. A confusion matrix is also produced to give this breakdown completely.

IV. EXPERIMENTAL RESULTS

The performance of LeNet and AlexNet models was measured based on various metrics like Loss, Precision, Recall, Accuracy, F1-Score, and Confusion Matrix, on Training and Validation sets. The models were trained for 50 epochs with the following comparative results:

A. *Training and Validation Accuracy:*

LeNet gained a consistent boost in training as well as validation accuracy; at the 50th epoch, training accuracy reached above 99.9%. The depicted validation accuracy is lesser but consistent improvement and at the end of training reached about 91.0%. This suggests that LeNet has a high performance on the training set but with a moderate generalization capability to unseen data.

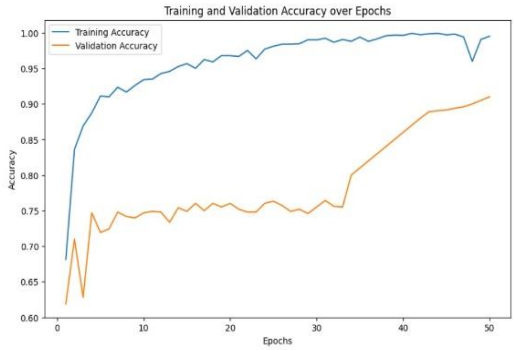

Fig. 2. Training and Validation Accuracy (LeNet)

AlexNet also improved considerably, with the training accuracy at approximately 98.9% at the 50th epoch. The validation accuracy of AlexNet was more stable, with a maximum of 93.1%, which indicates its improved generalization to the validation set compared to LeNet.

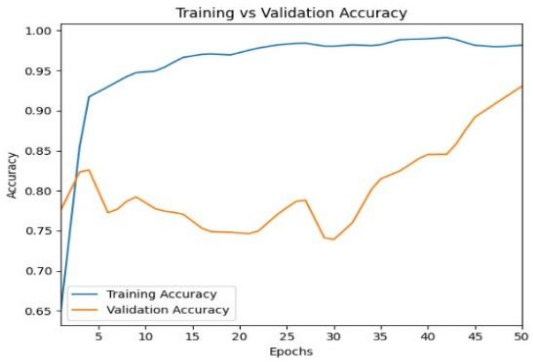

Fig. 3. Training and Validation Accuracy (AlexNet)

B. *Training and Validation Loss:*

LeNet showed a notable decrease in training loss, hitting a low of 0.0041 by the final epoch. While the validation loss also decreased, it experienced some fluctuations, settling at approximately 0.1857, which points to some overfitting. The gap between the validation and training loss indicates that this model was more suited to the training data.

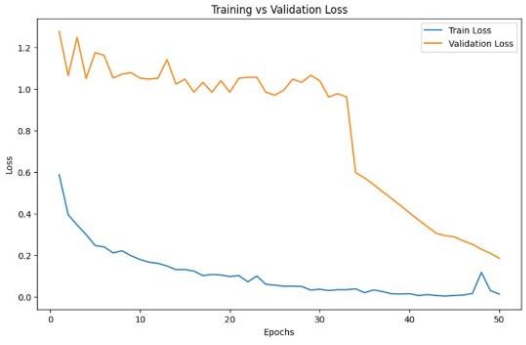

Fig. 4. Training and Validation Loss (LeNet)

AlexNet demonstrated a significant decrease in training loss, ultimately reaching a value of 0.0337. In contrast, the validation loss for AlexNet exhibited a more gradual decline, hitting 0.2205 by the 50th epoch, indicating improved consistency in t generalization.

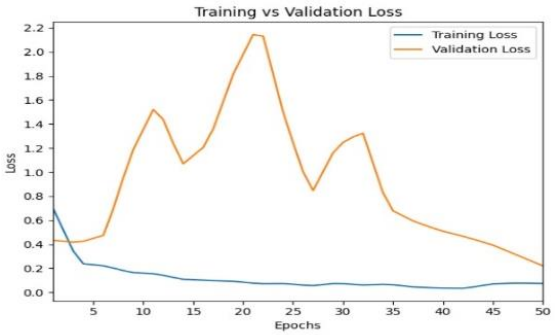

Fig. 5. Training and Validation Loss (AlexNet)

C. *Confusion Matrix:*

LeNet's training confusion demonstrated outstanding classification performance, recording 4900 true positives and 50 false negatives for the Stego class. The validation set reflected 4500 true positives and 100 false negatives, with a lower number of misclassifications in the Stego class.

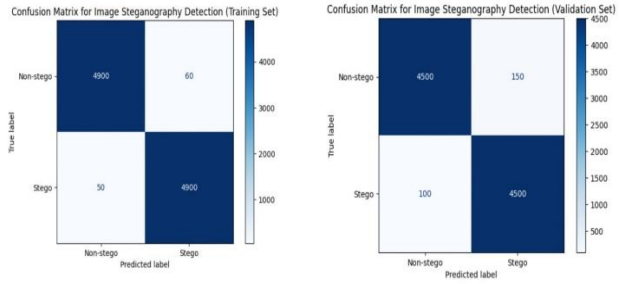

Fig. 6. Confusion Matrix (LeNet)

AlexNet's training confusion indicated a high classification accuracy, recording 2454 true positives and 46 false negatives for the Stego class. In the validation set, there were 2327 true positives and 173 false positives, demonstrating strong performance, although it had a slightly higher number of incorrectly classified stego images compared to LeNet.

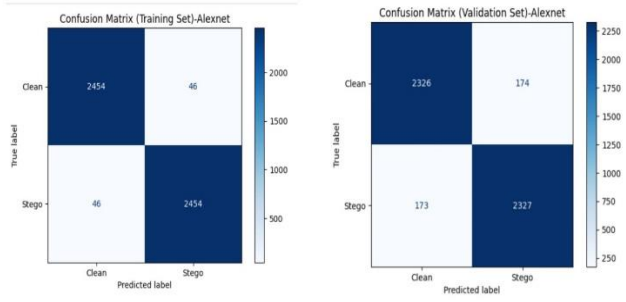

Fig. 7. Confusion Matrix (AlexNet)

D. *Precision, Recall, and F1-Score:*

Both AlexNet and LeNet demonstrated strong results in terms of Recall, F1-score, and Precision. LeNet achieved high precision and recall during training, but its validation performance indicated a slight drop in precision, while recall stayed robust. On the other hand, AlexNet had a lower training F1-score than LeNet, yet it displayed more consistent performance on the validation set, with a better balance between precision and recall. Overall, AlexNet showed better generalization abilities on unseen data when compared to LeNet.

TABLE I
PRECISON, RECALL, F1-SCORE

Models		LeNet	AlexNet
Training	Precision	98.79%	98.16%
	Recall	98.99%	98.16%
	F1-score	98.89%	98.16%
Validation	Precision	96.77%	93.04%
	Recall	97.83%	93.08%
	F1-score	97.30%	93.06%

V. CONCLUSION AND FUTURE ENHANCEMENT

In this project, we concentrated on detecting stego-images generated by the Least Significant Bit (LSB) method through the use of convolutional neural networks (CNNs). We assessed two architectures, AlexNet and LeNet-5, to evaluate their effectiveness in classifying images as either stego or clean. The results indicated that AlexNet outperformed LeNet-5 in several metrics, including precision, accuracy, F1-score, and recall. Specifically, AlexNet achieved a validation accuracy of 93%, which was higher than LeNet-5's 91%. However, LeNet-5 provided quicker inference times, making it a viable option for applications that prioritize processing speed in real time. These results imply that while AlexNet is better suited for tasks requiring high accuracy, LeNet-5 may be more appropriate in situations where speed is essential. For future improvements, the project could investigate more sophisticated architectures to enhance the detection accuracy of stego-images. Additionally, optimizing the models for real-time use through methods like model pruning or quantization could help strike a balance between performance and faster inference times, thereby increasing the system's efficiency. Furthermore, the detection module could be developed to identify stego-images created with other steganography methods aside from LSB, enhancing the system's versatility and robustness for different applications.

REFERENCES

[1] Eslam M. Mustafa, Mohamed M. Fouad, Mohamed A. Elshafey, “Enhancing the Performance of an Image Steganalysis Approach Using Variable Batch Size-Based CNN on GPUs”.

[2] Yu Sun, Tianyun Li, “A Method for Quantitative Steganalysis Based on Deep Learning”.

[3] Rishit Agrawal, Kelvin Jou, Tanush Obili, Daksh Parikh, Samarth Prajapati, Yash Seth, Charan Sridhar, Nathan Zhang, Mark Stamp, “On the Steganographic Capacity of Selected Learning Models”.

[4] Dan Wang, Ming Li, Yushu Zhang, “Adversarial Data Hiding in Digital Images”.

[5] Yen-Ting Chen, Tung-Shou Chen, Jeanne Chen, “A LeNet Based Convolution Neural Network for Image Steganalysis on Multiclass Classification”.
[6] Ching-Chun Chang, Isao Echizen, “Steganography in Game Actions”.
[7] Sahar Magdy, Sherin Youssef, Karma M. Fathalla, Mohamed Elhoseny, “DeepSteg: Integrating New Paradigms of Cascaded Deep Video Steganography for Securing Digital Data”.
[8] Yi Cao, Youwei Du, Wentao Ge, Yanshu Huang, Chengsheng Yuan, Quan Wang, “Generative Steganography Based on the Construction of Chinese Chess Record”.
[9] Sachin Dhawan, Rashmi Gupta, “Analysis of Various Data Security Techniques of Steganography: A Survey”.
[10] Abdullah Alenizi, Mohammad Sajid Mohammadi, Ahmad A. Al-Hajji, Arshiya Sajid Ansari, “A Review of Image Steganography Based on Multiple Hashing Algorithm”.
[11] May Alanzy, Razan Alomrani, Bashayer Alqarni, Saad Almutairi “Image Steganography Using LSB and Hybrid Encryption Algorithms”
[12] Al Hussien S. Saad, M.S. Mohamed, Eslam H. Hafez, “Coverless Image Steganography Based on Optical Mark Recognition and Machine Learning”.
[13] Biswarup Ray, Souradeep Mukhopadhyay, Sabbir Hossain, Sudipta Kr Ghosal, Ram Sarkar,”Image Steganography Using Deep Learning Based Edge Detection”.
[14] M. R. Islam, T. R. Tanni, S. Parvin, M. J. Sultana, A. Siddiqa, “A Modified LSB Image Steganography Method Using Filtering Algorithm and Stream of Password”.
[15] Marghny H. Mohamed, Mahmoud A. Mofaddel, and Tarek Y. Abd El-Naser, “Comparison Study Between Simple LSB and Optimal LSB Image Steganography”.
[16] Kumar P, V. K. S, P. L and S. SenthilPandi, "Enhancing Face Mask Detection Using Data Augmentation Techniques," International Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G NAGARA, India, 2023, pp. 1-5, doi: 10.1109/ICRASET59632.2023.10420361.
[17] Kumar P, V. K. S and S. P. S, "CNN and Edge-Based Segmentation for the Identification of Medicinal Plants," 5th International Conference on Intelligent Communication Technologies.