University of Staffordshire London

AI-Powered Steganalysis: Detecting Hidden Data in Digital Images Using Deep Learning

A ConvNeXt-Based Hybrid Detection System with Explainable Visual Analysis and Gradio Interface

BSc (Hons) Cyber Security

Final Year Project

May 1st 2026

1 PROJECT OVERVIEW

This project presents the design, implementation and evaluation of an AI-powered steganalysis system capable of detecting hidden data in digital images with high accuracy. The system uses a fine-tuned ConvNeXt-Base model with 3×3 Laplacian High-Pass Filter (HPF) preprocessing to enhance subtle noise residuals introduced by steganographic embedding.

The tool provides a binary cover/stego prediction with confidence scores, six explainable visual outputs, and an LSB extraction module to recover hidden text (e.g., Steghide payloads). It is delivered as a user-friendly Gradio web application that runs locally with no internet connection.

2 SYSTEM PIPELINE

Input Image → HPF Filtering (3×3 Laplacian) → ConvNeXt-Base Model → Prediction (Cover / Stego) → Explainable Visual Outputs → LSB Extraction (Text Recovery)

3 RESEARCH AIM

To develop an accurate, robust and explainable AI-powered steganalysis tool that can detect hidden data in digital images and assist cybersecurity and digital forensics operations.

4 RESEARCH OBJECTIVES

- To review and analyse existing steganography and steganalysis techniques and identify research gaps.
- To design and implement a deep learning-based steganalysis model using ConvNeXt-Base and HPF preprocessing.
- To integrate classical image analysis and LSB extraction for explainable results.
- To evaluate system performance using ALASKA2 dataset and benchmark against state-of-the-art methods.
- To develop a user-friendly Gradio interface for practical deployment.

5 DATASETS

ALASKA2
Large-scale benchmark dataset with 242,424 images covering 25 steganographic algorithms and multiple embedding rates.

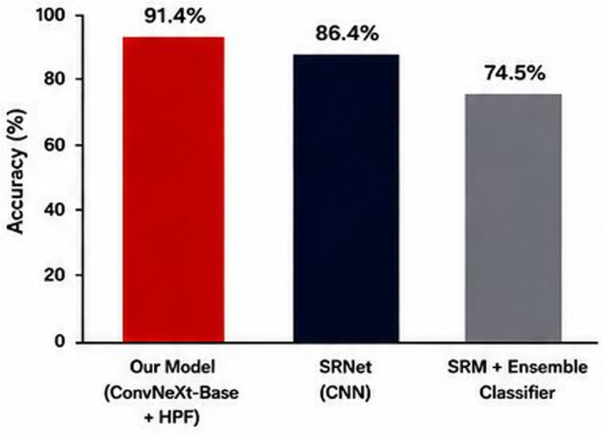
BOSSBase
Classic dataset used for comparative analysis with traditional steganalysis methods.

6 KEY RESULTS (ALASKA2 TEST SET)

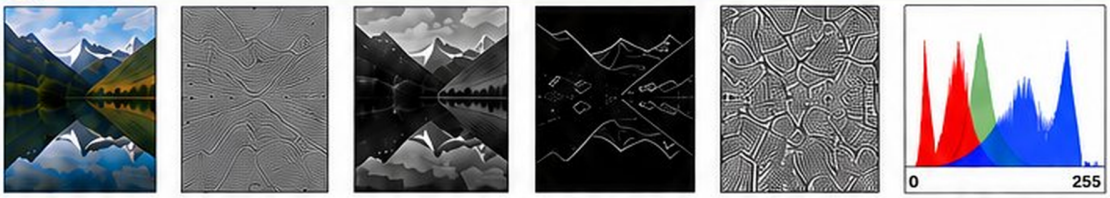
Accuracy	91.4%
Precision	90.2%
Recall	92.8%
F1-Score	91.5%
P_e (Probability of Error)	0.086

Outperforming SRNet by 5.0% and SRM+Ensemble by 16.9% in accuracy.

7 PERFORMANCE COMPARISON



8 EXPLAINABLE VISUAL OUTPUTS



Original Image | HPF Filtered (Laplacian) | Grayscale Image | Canny Edge Detection | LSB Bit-Plane (Bit Plane 0) | RGB Channel Histogram

9 LSB EXTRACTION (TEXT RECOVERY)

Successfully extracts hidden text payloads embedded using tools like Steghide. Supports text output with automatic decoding and error handling.

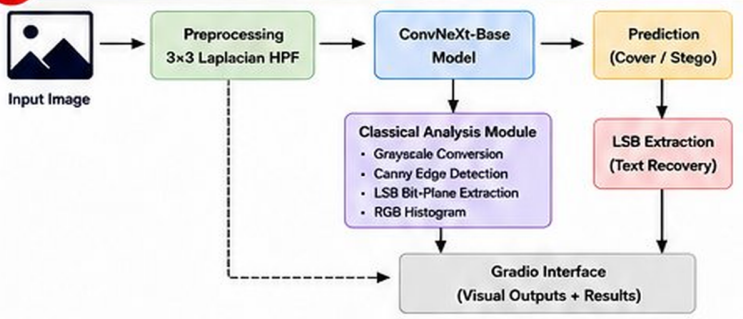
Extracted Text H

This is a secret message hidden in the image using Steghide.

10 ROBUSTNESS ANALYSIS

Attack Type	Settings	Accuracy (Our Model)
JPEG Compression	(QF 70)	86.1%
JPEG Compression	(QF 85)	89.3%
Gaussian Noise	(σ = 0.02)	83.7%
Gaussian Noise	(σ = 0.05)	82.7%
Bilinear Resizing	(50%)	84.2%

11 SYSTEM ARCHITECTURE



12 KEY FEATURES

- High detection accuracy with ConvNeXt-Base + HPF
- Comprehensive visual analysis for explainability
- LSB bit-plane extraction and text recovery
- User-friendly Gradio web interface
- Runs fully locally – no internet required

13 APPLICATION AREAS

- Digital Forensics
- Cybersecurity Operations
- Malware & Forensic Analysis
- Covert Channel Detection
- Data Exfiltration Prevention

14 TECHNOLOGY STACK

PyTorch · timm · OpenCV · NumPy · PIL Pillow · Gradio

15 STUDENT DETAILS

Name: Guevara Ayos Mahinga
Roll Number: 23031403
Degree: BSc (Hons) Cyber Security
Level: 6

16 SUPERVISOR DETAILS

Supervisor Name: Samuel Onalo

17 CONCLUSION

The developed system achieves 91.4% accuracy on ALASKA2 test set, outperforming SRNet by 5.0% and SRM+Ensemble by 16.9%. It provides an explainable, robust and practical solution for real-world steganalysis tasks with strong potential for deployment in cybersecurity and digital forensics environments.

AI-Powered Steganalysis: Detecting Hidden Data in Digital Images Using Deep Learning

A ConvNeXt-Based Hybrid Detection System with Explainable Visual Analysis and Gradio Interface