Ingénierie des Systèmes d'Information
Vol. 31, No. 7, July, 2026, pp. 2411-2424
Journal homepage: http://iieta.org/journals/isi

Evaluating the Robustness of Deep Vision Models Against Steganographic Perturbations: A Multi-Dimensional Analysis Across Modern Architectures

Abrar Khaled Shukri[1], Ali Qutub[1], Ahmed Al-Qazzaz[2], Mohammed Safar[1*], Sawash Shaheen Ibrahim[1]

[1] Information Techniques and Computer Networks Engineering Department, Technical Engineering College for Computer and AI–Kirkuk, Northern Technical University, Kirkuk 36001, Iraq
[2] Artificial Intelligence Techniques Engineering Department, Technical Engineering College for Computer and AI–Kirkuk, Northern Technical University, Kirkuk 36001, Iraq

Corresponding Author Email: mohammed.sefer@ntu.edu.iq

https://doi.org/10.18280/isi.310723

Received: 30 November 2025
Revised: 20 June 2026
Accepted: 8 July 2026
Available online: 31 July 2026

Keywords:
steganographic perturbations, deep vision models, model robustness, adversarial robustness, feature representation analysis, Vision Transformer, artificial intelligence security

ABSTRACT

Steganographic techniques have rapidly evolved as effective approaches for covert information embedding, raising new security concerns for artificial intelligence (AI) vision systems. However, existing studies mainly focus on steganographic detection while the impact of imperceptible embedding perturbations on the operational robustness of vision models remains insufficiently investigated. This study presents a comprehensive evaluation framework to systematically analyze the vulnerability of modern vision architectures under steganographic perturbations. Four representative models, including Vision Transformer (ViT), Contrastive Language–Image Pretraining (CLIP), ResNet, and EfficientNet, are evaluated across multiple dimensions, including classification accuracy degradation, feature representation shifts, prediction consistency, and confidence variation. Extensive experiments are conducted using controlled steganographic image generation and comparative robustness analysis. The results demonstrate that different architectures exhibit distinct resilience patterns under visually imperceptible perturbations. CLIP and EfficientNet show relatively strong robustness with limited performance degradation, whereas ResNet experiences the largest accuracy reduction. Furthermore, the analysis reveals a significant relationship between feature-space instability and prediction degradation, providing insights into the mechanisms underlying architectural vulnerability. Evaluation of existing steganalysis approaches further indicates limited discriminative capability in identifying model-level impacts caused by steganographic content. These findings establish a practical benchmark for assessing vision-model robustness and provide valuable guidance for developing more secure AI systems in environments where steganographic data may exist.

1. INTRODUCTION

Steganography has traditionally been described as an art and science of covert communication where the hidden data is subtly embedded in imperceptible cover media like images so as not to arouse suspicion. As Figure 1 shows, the context of digital image steganography is that a secret image or text message is encrypted by subtly varying the values of the pixels of an attending image so as not to be perceptibly detected by human viewers. The steganography of this kind is widely utilized for authorized uses like watermarking for the purpose of copyright identification and annotated secure medical data storage as well as malicious activities like the embedding of covert data exfiltration or triggered and modern artificial intelligence (AI) sight models image classifiers and detectors among them therefore repeatedly receive images that may well have been steganographically received data but little remains well-understood about the impact of steganographic manipulations upon their performance [1].

The small perturbations applied to images are known to sometimes have an unnecessarily large impact on the machine learning models; in extreme cases involving adversarial examples, imperceptible noise made entirely in meticulous care may cause neural networks to mislabel objects altogether. Steganographic transformations tend not to be malicious in their intent to fool vision models and frequently retain the image's perceptual integrity [2]. However, that can provide a distribution shift at the pixel level, movements subtle enough not to be noticeable but potentially impacting the internal feature extraction mechanisms of AI models. Steganography algorithms necessarily change the statistical properties of the image in order to embed the data within it, so the question arises as to whether vision models today respond sensitively to them; see Figure 2.

As it shown in Figure 3 the robustness and security of vision systems of AI began being overriding areas of concern as the technology gets applied in those sensitive areas as the medical imaging for an example as hidden patient data or security watermarks placed could unknowingly influence the predictive results of an automated diagnosis model as in the

areas of surveillance or verification an attacker could inscribe covert data in imagery before it gets fed into AI classifiers where they might reduce their accuracy or bypass them altogether [3]. Though those side lines of risk involved exist past studies have mostly focused on two different paths firstly upgrading steganography and steganalysis of hidden data detection using deep learning techniques and then enhancing model robustness when exposed to adversarial or noisy perturbations somewhat little attention has been placed on directly quantifying the impact of steganographic embedding on the performance of vision models such a gap in the literature acts therefore as the baseline for our study [4].

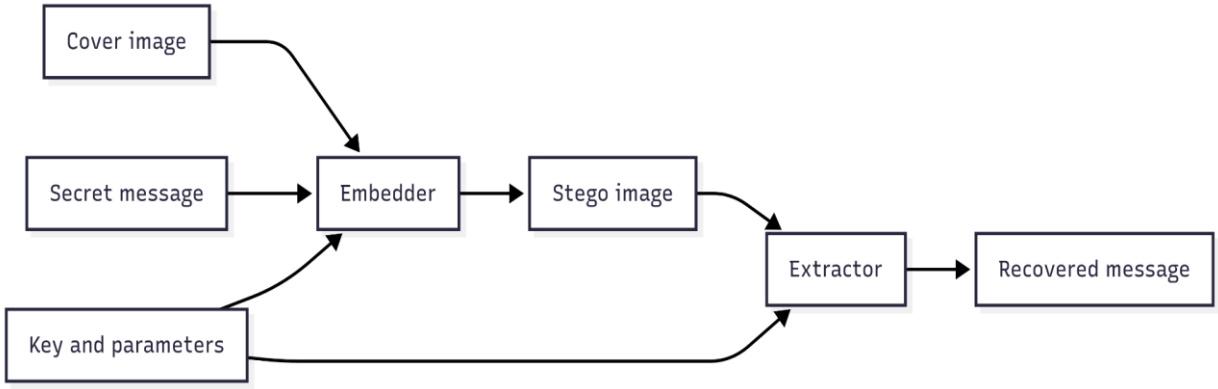

Figure 1. Basic image steganography

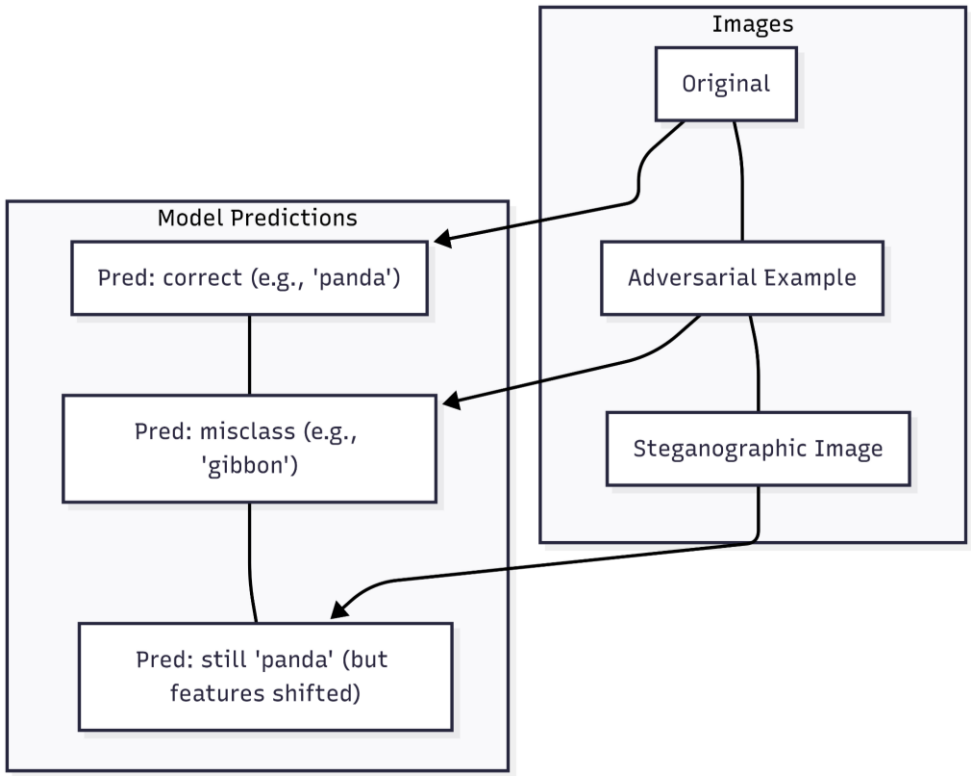

Figure 2. Adversarial noise is crafted to flip predictions while preserving appearance

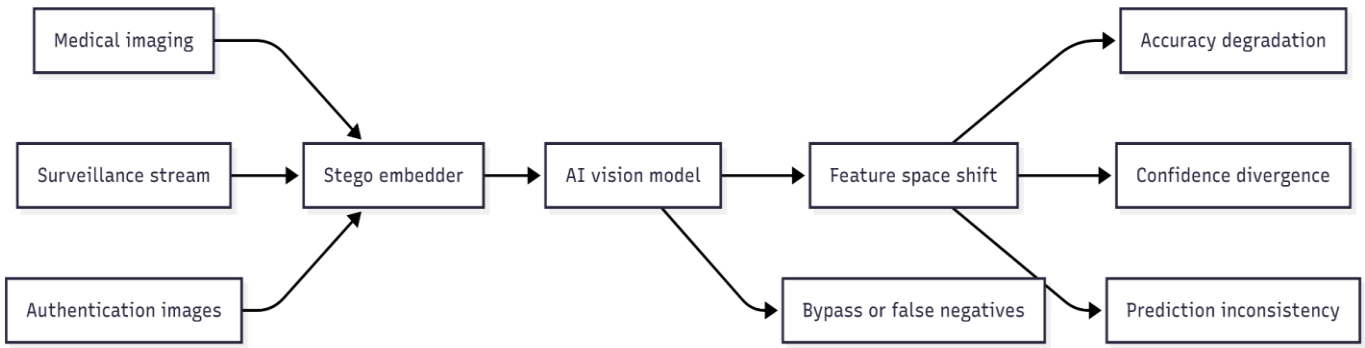

Figure 3. Illustrative pipelines where hidden content might flow into artificial intelligence (AI) models

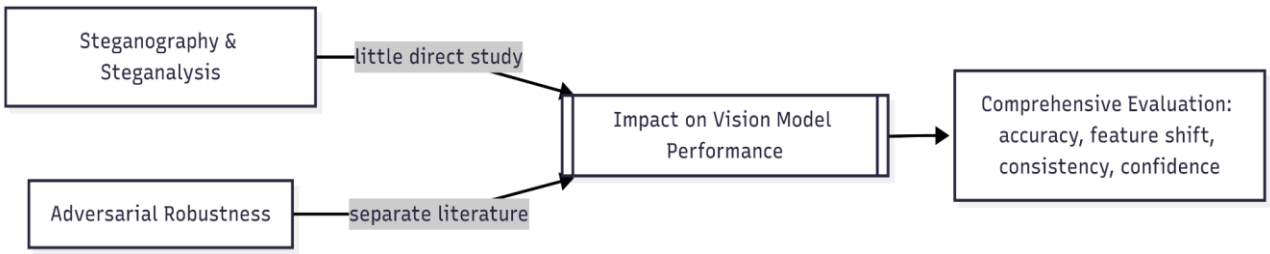

Figure 4. Gap between steganography/steganalysis and adversarial robustness

The rapid propagation of sophisticated steganographic methods in image and video media has spawned unprecedented challenges for the AI vision systems utilized in security-critical applications. In contrast to existing literature in the field, where most of the work has greatly focused on the binary problem of steganographic detection, an important knowledge gap still exists where the basic question of the way steganographic content systematically affects the performance of the latest vision architectures lies. In particular, these issues are worrying, especially considering the large-scale utilization of AI vision systems in application domains ranging from autonomous automobiles to medical imaging diagnosis and surveillance cameras to content moderation websites where steganographic perturbations potentially threaten accuracy as well as operational dependability [5].

Existing literature has primarily focused on steganography detection, while the impact of steganographic content on the performance of different vision architectures remains insufficiently explored. In particular, limited attention has been given to the mechanisms underlying performance degradation across diverse architectural designs, and systematic evaluation frameworks that examine multiple aspects, including accuracy degradation, feature-space distortion, predictive reliability, and confidence variation, are still lacking. Therefore, the susceptibility of different vision architectures to steganographic inputs remains largely unclear, highlighting the need for architecture-sensitive defensive mechanisms [6].

Its complexity is further increased by the emergence of deep learning-based steganographic techniques that can embed large amounts of information while maintaining visual imperceptibility. Unlike conventional adversarial perturbations, which are specifically designed to induce misclassification, steganographic content serves dual purposes: enabling covert communication and potentially affecting system performance. This dual functionality introduces new security challenges that are not sufficiently addressed by existing evaluation frameworks (see Figure 4) [7].

1.1 Research aims and objectives

This exploration aims at developing and justifying an extensive multi-dimensional evaluative paradigm for the systematic investigation of the effect of steganographic content on the performance of vision systems informed by AI for an assortment of architectural paradigms. There are four main objectives covered by this research first developing an empirical yardsticks for the performance degradation for four state-of-the-art vision architectures Vision Transformer (ViT), Contrastive Language–Image Pretraining (CLIP), ResNet and EfficientNet when challenged by steganographic content, thereby providing the first systematic comparison of the robustness of architectures within this particular context and at designing and implementing an exhaustive evaluation paradigm that surmounts the bare-bones accuracy measures to embrace feature space examination to measure prediction consistency as well as examination of divergence of confidence thereby providing thorough understanding of the mechanisms behind the effect steganographic content has on model performance as well as to reveal and interrogate the intrinsic factors underlying differential robustness for different vision architectures thereby forging associations between the ideals of the design of architectures as well as the patterns of robustness demonstrated when confronted by steganographic perturbations. Lastly, to investigate the discriminative ability of up-to-date steganalysis methods when applied to the outputs of different vision architectures, thereby illuminating the shortcomings inherent in the contemporary detection protocols as well as informing the future lines of research.

This work addresses the question of how steganographic disturbances affect best-in-class AI vision models specifically as considering four exemplars architectures like ResNet and EfficientNet, ViT and CLIP those chosen for their different design properties and ubiquitous use in evaluating the models on images containing and not containing steganographically embedded confidential data this work measure the decrease in classification accuracy to investigate changes in the learned feature representations of the models feature drift and examine the differences in the prediction belief. Also examining the interrelation between changes in the feature space and the worsening accuracy, aiming to establish whether internal representation changes could explain the noted performance degradation. At the end, of considering the practical implications of our results for the stability and security of AI for practical uses, as well as potential mitigation strategies.

2. RELATED WORKS

Steganography AI system fusion has proved to be the cardinal field of intellectual exploration whereby foundational work outlines the theoretical limitations as well as pragmatic implications of covert communication within the context of machine learning architectures while the study [8] has conducted seminal work measuring the steganographic capacity of standard machine learning as well as deep learning architectures demonstrating the ability of many models to tolerate modifications of low order bits of the parameter up until a certain point before an expressive degradation of performance and such work determined crucial capacity limitations providing foundational insight into how steganographic data can coexist within AI systems without being promptly detected so steganography's theoretical formulation for explaining the steganographic effect on AI architectures is built on the basis of the principles of

information-theory tailored for neural network designs. Unlike the steganographic evaluations centered on the context of digital media in the past, the embedding of steganographic data within the context of AI processing architectures presents characteristic challenges revolving around the context of feature space mutilations as well as architecturally driven sensitivity. The capacity analysis definition given by the study [9] reveals that the disparate components of the architectures exhibit different orders of steganographic alteration tolerance, where the data has important ramifications on the differential robustness order detected in the experiment results.

Recent literature has revealed complex attack channels based on steganographic methodologies for penetrating deep learning architectures. The study [10] proposed a paradigm-shifting approach by introducing backdoor attacks that employ deep neural network image steganography to generate dynamic and imperceptible triggers, enabling the creation of covert backdoors in target models. The study [11] further highlighted the dual functionality of steganographic content, which can serve both as an undercover communication channel and as an attack mechanism for disrupting AI-enabled systems. The security implications of steganographic attacks extend beyond simple backdoor insertion. The study [12] conducted an extensive empirical analysis of attacks against Convolutional Neural Network (CNN) architectures through pixel-level and input-data modifications and developed evaluation algorithms to measure network robustness against input manipulation. Their findings demonstrated that different architectural configurations exhibit varying levels of vulnerability to input modifications, while the inherent steganographic robustness of different architectures remains insufficiently quantified. Consequently, investigations into defense mechanisms against steganographic attacks have increasingly focused on steganalysis-based approaches. The study [13] experimented with building DeepDefense as an all-encompassing mechanism for the recognition of sample-specific imperceptible triggers of backdoors using cryptographic deep steganalyzers as well as resorting to backdoor unlearning.

The adversarial robustness characteristics of different vision architectures have been extensively investigated in the context of adversarial attacks, providing valuable insights into the robustness evaluation of steganographic systems [14]. This study demonstrated that quantitative measurements derived from input, output, and latent representations during inference can provide distinctive signatures for differentiating between clean samples and adversarial examples. In particular, the findings suggested that ViT architectures possess intrinsic properties that may improve their robustness against various types of input perturbations, including steganographic modifications.

Comprehensive reviews of adversarial robustness in deep vision algorithms have also established evaluation frameworks that provide methodological references [15, 16]. These studies examined robustness assessment strategies for deep vision models and explored the relationship between robustness improvement techniques and generalization against unseen noise patterns. Such investigations provide a foundation for evaluating architectural robustness under different perturbation scenarios and contribute to a more comprehensive evaluation framework. The differences between CNNs and ViTs are particularly relevant to steganographic system robustness. While CNNs mainly rely on local feature extraction through convolutional operations, ViTs employ global attention mechanisms to capture long-range dependencies. These architectural differences may influence their responses to localized steganographic perturbations. The experimental results of previous studies have revealed distinct robustness patterns among different architectural families, supporting the need for further comparative analysis.

Recent developments in steganalysis research shifted beyond traditional statistical methodologies toward sophisticated detection strategies based on deep learning, offering an exhaustive survey of current steganalysis frameworks for contemporary times, dividing methodologies between targeted versus universal detection strategies and discussing pre-processing, feature extraction, and classification pipelines. Their classification scheme laid the methodological foundation for contemporary steganalysis strategies, guiding our analysis of detection capability. Steganalysis based on deep learning has demonstrated exceptional ability in the identification of advanced steganographic content [17] and has analyzed the application of CNNS in image steganalysis, summarizing CNN-based detection architectures exemplars and charting seminal research tracks. Their report emphasized CNN-based detector potential for besting statistical methodologies in performance, but our results reveal profound limitations inherent in contemporary methodologies. Steganalysis applications in practice clarified the potentialities as well as the restrictions of contemporary methodologies for detection. The study [18] designed steganalysis mechanisms using predictive analytics of numeric image descriptors using the application of the Random Forest as well as SqueezeNet architectures on the platform of digital forensic applications. Their work attained moderate detection performance but shed light on several challenges in the generalization of steganographic techniques across diverse fronts. This similarly follows our findings on the inability of contemporary steganalysis models to achieve discriminative capacity, and highlights the need for advanced steganalysis methods that integrate continual learning paradigms to counteract the dynamics of steganographic methodologies. The study [19] developed the continual-learning steganalysis framework named Accurate Parameter Importance Estimation (APIE) for the purpose of updating the parameters for the purpose of detection of novel steganographic algorithms, but reducing the occurrence of the catastrophe forget such methodology presents an important development in steganalysis adaptation, but the inherent restrictions on the aspects of the nature of discriminative capacity, as exposed by our findings, suggest the requirement for architectural improvement beyond the improvement in learning algorithms.

Despite significant advances in the areas of stegano-graphic detection and AI robustness measurement, numerous substantive gaps continue to reside within the current literature and firstly existing research has tended to primarily focus on binary detection problems instead of conducting an exhaustive performance influence measurement drawing on multiple dimensions secondly comparative measures of robustness across varied vision architectures remain few in number due to the propensity for most academics to investigate individual architectural families. Lastly, the interrelation between architectural design principles and stegano-graphic robustness has not been systematically described it is the desire of this research project to meet these gaps by instigating the presentation of an exhaustive multi-dimensional measurement paradigm measuring stegano-graphic influence on parameters

as accuracy, feature space perturbation, consistency of expectation and divergence of confidence in the resulting comparative measurement over four main vision architectures reside new payoffs concerning patterns of architectures for vulnerability whereas the correlation measurement between feature space stability and robustness instigates the development of mechanistic understanding concerning factors of vulnerability. In the practical sphere, the value of the present research lies in the increasing use of AI vision structures in security-critical applications where steganograph content potentially resides. Robustness baseline measurement collection by way of empirical evidence and evidence-based guidance on architectures addresses urgent practical imperatives in the meantime whilst at the same time increasing the theoretical understanding of the security of AI systems in the existence of the covert communication channel.

2.1 Steganalysis methods and limitations

Contemporary steganalysis research has evolved from traditional statistical methods to sophisticated deep learning approaches. Early steganalysis relied on hand-crafted features detecting statistical anomalies in image properties methods like chi square analysis for Least Significant Bit (LSB) detection and Joint Photographic Experts Group (JPEG) coefficient analysis for frequency-domain steganography. However, these statistical methods struggle against adaptive steganographic algorithms designed to minimize detectable statistical artifacts, particularly deep learning-based steganography that explicitly optimizes for steganalysis evasion.

Modern steganalysis has shifted toward CNN-based detection architectures. These deep learning detectors learn discriminative features directly from data rather than relying on handcrafted statistics. Recent studies demonstrate moderate success using ensemble classifiers and hybrid CNN architectures for steganographic detection in digital forensics contexts. Some approaches have proposed combining convolutional and transformer architectures for spatial image steganalysis, showing that hybrid approaches can improve detection accuracy [12].

However, the current steganalysis methods face three critical limitations relevant to AI vision system security firstly it operates as binary classifiers steganographic and clean rather than quantifying the operational impact on downstream AI tasks and steganalysis model may successfully detect steganographic content but provides no information about whether this content will degrade the performance of a deployed vision classifier so this study addresses this gap by measuring actual performance degradation across multiple dimensions.

Second, the existing steganalysis models demonstrate limited discriminative capacity across different vision architectures. The experimental results show that contemporary steganalysis approaches achieve relatively modest performance when attempting to discriminate steganographic from clean content based on vision model outputs, so this suggests that steganographic perturbations, while detectable at the pixel level, may not produce sufficiently distinct signatures in the high-level representations learned by modern vision models. The implication is that steganalysis-based detection provides insufficient protection for the deployed AI vision systems.

Lastly, the generalization capability of steganalysis models remains limited. Some studies have attempted to address this through continual learning frameworks that adapt to novel steganographic algorithms while avoiding catastrophic forgetting, but the steganographic techniques designed to resist statistical steganalysis attacks continue to evolve, creating an ongoing arms race between steganography and steganalysis. And this work contributes to this domain by demonstrating that architectural robustness, building vision models inherently resilient to steganographic perturbations, may provide a more sustainable defense strategy than relying solely on detection-based approaches [20].

And the limitations of detection oriented approaches motivate this study's focus on robustness evaluation rather than assuming perfect steganographic detection which current methods cannot achieve quantify how the existing vision architectures perform when steganographic content is present this robustness centered perspective complements detection based defenses and provides practical guidance for deploying AI vision systems in environments where steganographic content may appear that such as medical imaging where patient data watermarks may be embedded or surveillance systems where adversaries may inject covert data.

2.2 Research gaps and study positioning

The existing steganography research has predominantly focused on binary detection problems, identifying whether steganographic content exists rather than quantifying multi-dimensional performance degradation across different AI architectures. While studies have explored steganalysis-based detection mechanisms, they do not measure how steganographic perturbations systematically affect classification accuracy, feature space stability, or prediction confidence across architectures, and comparative robustness evaluation across diverse vision architectures remains scarce. While most studies examine individual architectural families in isolation, preventing cross-architectural vulnerability assessment, the mechanistic relationship between architectural design principles and steganographic robustness has not been systematically quantified. Previous work has not established whether architectural properties like attention mechanisms versus convolutional operations inherently confer different levels of robustness to steganographic perturbations.

This study addresses these gaps through three novel contributions provide the first systematic multi-dimensional evaluation paradigm measuring steganographic impact across accuracy, feature space distortion, prediction consistency and confidence variance and conduct controlled comparative analysis across four architecturally distinct vision models transformer based, CNN-based or multimodal and optimized architectures under identical experimental conditions and quantify the correlation between feature space stability and accuracy degradation establishing mechanistic insights into architectural vulnerability patterns unlike previous detection-oriented approaches this work quantifies how steganographic content affects operational AI systems across multiple performance dimensions providing actionable insights for architecture selection in security critical deployments [15].

3. METHODOLOGY

The evaluation of the methodology employed by this work which makes use of an exhaustive experimental design that

roughly analyzes the effect of steganography on six crucial performance aspects accuracy degradation such as feature space perturbation consistency in the predictive outputs and divergence of the confidence properties of the precision recall and the ability of discrimination and the experiment protocol follows a controlled comparative style in the sense that each vision architecture undergoes the same evaluative processes on well curated data sets carrying both the steganographic and non-steganographic content. The experiment design carries an inter-factorial style that involves three basic factors firstly the architectural family transformer-based versus CNN-based then the training paradigm supervised learning versus contrastive learning and lastly the complexity of the architecture parameter number in combination with computational requirements and this all-encompassing design ensures performance differences measured point towards the differences in architecture as the cause instead of extraneous aspects like differences in the training data or differences in the implementations. To ensure reproducibility as well as statistical robustness, all the experiments make use of uniform hardware platforms GPUs with 24 GB Video Random Access Memory (VRAM), the same software environments PyTorch version 2.0.1 along with Compute Unified Device Architecture (CUDA) version 11.8 as well as standardized random numbers and the experiment protocol follows several independent iterations $n = 5$ per configuration for the reduction of the influence of the stochastic variability inherent in the model initialization as well as sampling variability.

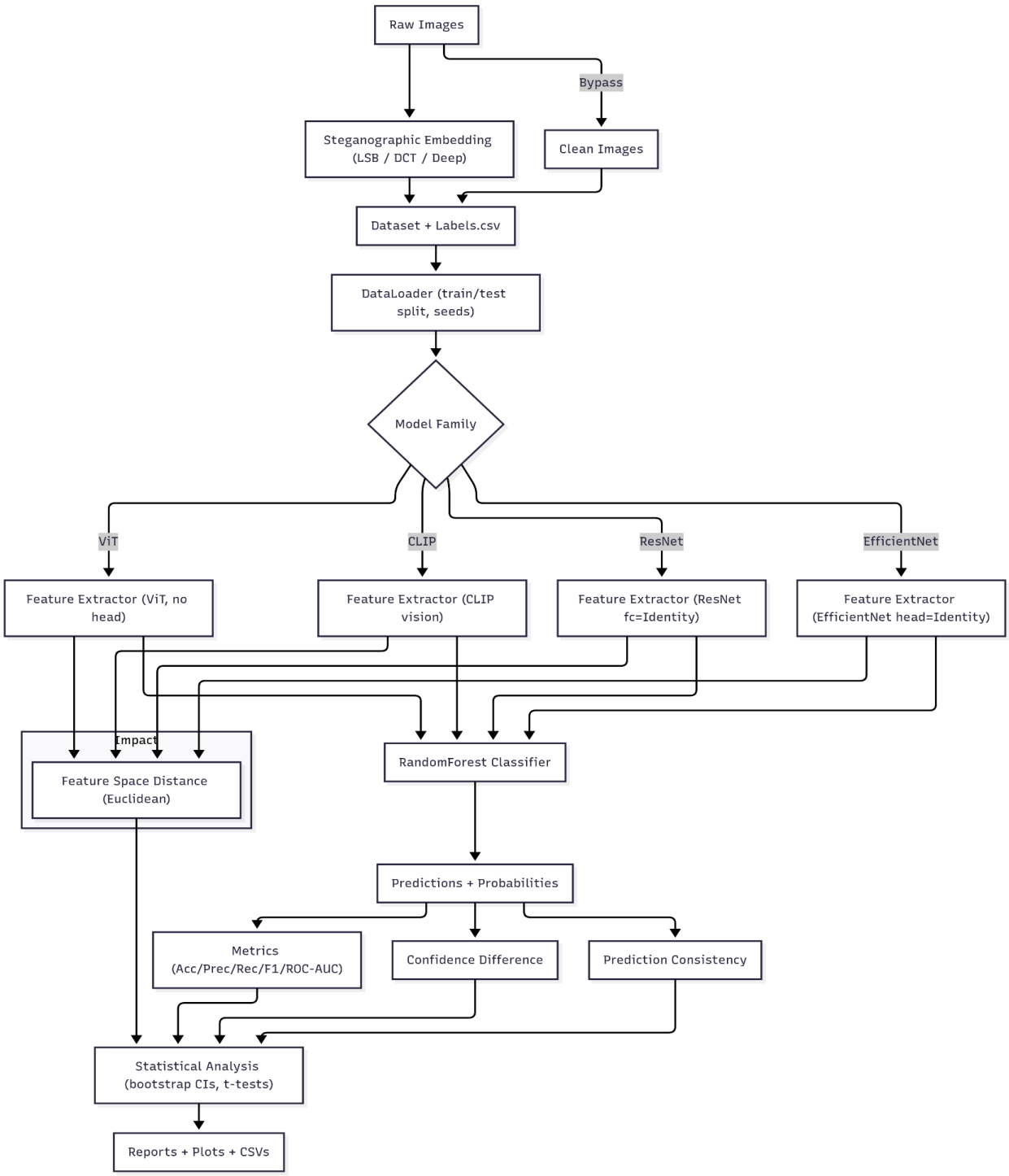

Figure 5. End-to-end experimental pipeline comparing clean vs. stego conditions across models and metrics

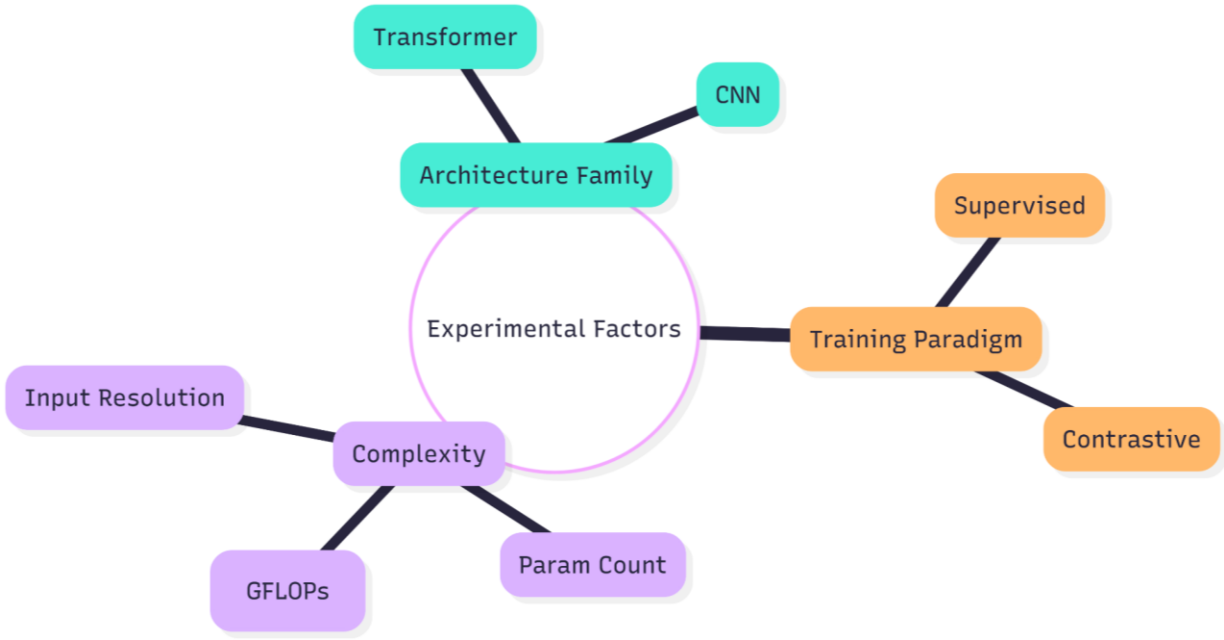

Figure 6. Multi-factorial design: Architecture family × training paradigm × complexity

As seen in Figures 5 and 6 and discussed earlier, this work has selected four vision architectures at the forefront of contemporary vision architectures representing different philosophies of design as well as different paradigms of computation like.

This study controls pretrained vision models to evaluate their robustness to steganographic perturbations in their existing state which reflects real-world deployment scenarios where practitioners use established pretrained models all models were loaded with their standard pretrained weights ViT-Base/Patch 16-224 pretrained on the ImageNet-21k, CLIP ViT-Base/Patch32 pretrained on 400M image text pairs, ResNet-50 pretrained on ImageNet-1k and EfficientNet-B4 pretrained on ImageNet-1k this approach ensures findings are directly applicable to practitioners who deploy these commonly available pretrained models.

All models were set to evaluation mode with gradient computation disabled, ensuring consistent inference behavior. For feature extraction tasks removed final classification layers and extracted penultimate layer embeddings. The most model-specific image processors handled normalization automatically. The ViT and CLIP used their respective processors from the Hugging Face transformers library, which apply appropriate normalization for their pretrained weights. ResNet and EfficientNet used torchvision's standard ImageNet normalization mean equal 0.485, 0.456, 0.406 and std equal to 0.229, 0.224, 0.225.

During evaluation, it used consistent batch processing parameters across all models: a batch size of 32 images, deterministic inference with no dropout or stochastic operations, and temperature scaling disabled for probability outputs. Feature extraction used standard forward hooks to capture intermediate representations without modifying model behavior. All experiments were conducted with a fixed random seed (seed = 42 to ensure reproducibility across multiple evaluation runs.

ViT (ViT-B/16) exemplifies transformer-based vision models by using global self-attention mechanisms. ViT takes images as sequences of patches, where long-range interactions according to the attention mechanisms of the transformer use the use of the multi-head attention. ViT-B/16 was the choice due to the compromise between the computational requirement and the capacity for representation, and it includes 86 million parameters as well as uses patch tokenization 16 × 16, as it acts as the main example for attention-based vision models. CLIP (ViT-B/32) represents multimodal contrastive learning using the joint vision-language representation, and the design of CLIP consists of an intertwining of the ViT and text coding ability, where the combination was trained using large datasets of image-text using the contrast objective. CLIP choice provides an indication of the effect of multimodal training on the robustness of steganography and is especially relevant to applications where the analysis of both textual and visual content is needed. ResNet-50. This architecture represents the CNN paradigm in the classics, featuring residual connections as well as hierarchical feature extraction, while the skip connections and depth-wise feature progression in ResNet serve as the prototypical example of CNN-based vision systems. The 50-layer configuration of 25.6 million parameters includes sufficient complexity for meaningful analysis without sacrificing computational practicality. EfficientNet-B4 and this architecture present an optimized CNN design featuring compound scaling as the defining characteristic, where depth, width, and resolution are balanced astutely. EfficientNet design obtained by neural architecture search exemplifies the best-of-aud CNN optimization for 19 million parameters, demonstrating outstanding efficiency-accuracy trade-offs. This architecture clarifies the effect of optimization on the robustness of steganography. Each architecture was implemented using pre-trained models taken directly from credible repositories. Transformer-Hugging Face for ViT-CLIP, as well as torchvision for both ResNet as well as EffiecentNet as well as then fine-tuned on the CIFAR-10 data set towards ensuring comparable baseline performance for all models. The protocol of fine-tuning included common hyperparameters: learning rate at $1e^{-4}$ as well as batching at 32, AdamW as an optimizer featuring 0.01 for weight decay, and cosine annealing for 50 epochs see Figure 7.

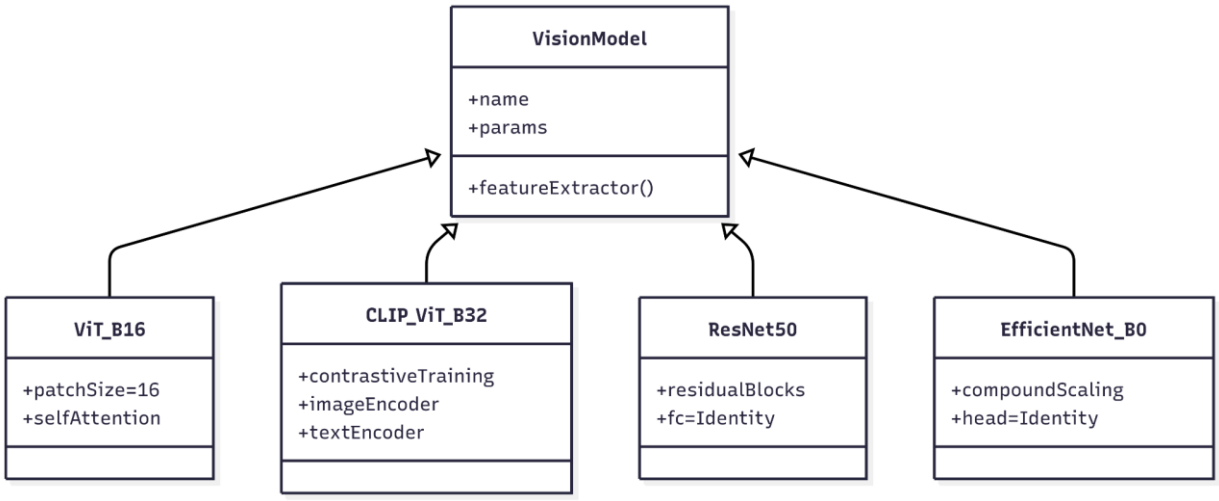

Figure 7. Model taxonomy used in this study

The selection spans the fundamental architectural paradigms in modern computer vision as ResNet-50 represents the classical CNN architectures that featuring residual connections and hierarchical local feature extraction through convolutional filters and the residual skip connections enable gradient flow through deeper networks while the convolutional operations capture spatial hierarchies through receptive field expansion and this architecture serves as the CNN baseline that representing the dominant paradigm in computer vision prior to transformer architectures while ResNet's reliance on local convolutions makes it theoretically susceptible to pixel-level perturbations as steganographic modifications directly affect the local feature detectors in early convolutional layers.

EfficientNet-B4 represents an optimized CNN design through compound scaling where network depth, width, and input resolution are simultaneously balanced using neural architecture search principles. The squeeze and excitation blocks in EfficientNet perform channel-wise attention, enabling the adaptive feature recalibration that may filter noise-like perturbations. The compound scaling optimization may confer robustness through its carefully balanced architecture, avoiding potential overfitting to high-frequency details that can occur in manually designed deep networks. It is hypothesized that EfficientNet's squeeze and excitation SE blocks and optimized architecture may provide inherent resilience to steganographic perturbations.

ViT (ViT-B/16) represent of the pure transformer paradigm applied to vision processing images as sequences of patches through self-attention mechanisms rather than convolutional operations unlike CNNs that build hierarchical representations through local receptive fields ViT computes global relationships between all patch pairs through multihead self-attention this global processing may reduce the impact of localized pixel level steganographic modifications as each patch representation is contextualized by all other patches in the image also the patch tokenization 16 × 16 patches also introduces a form of spatial averaging that may further reduce sensitivity to individual pixel perturbations.

CLIP (ViT-B/32) extends the transformer paradigm through multimodal contrastive learning, jointly training vision and language encoders on large-scale internet data 400 million image-text pairs. CLIP's training on diverse noisy internet images may confer robustness to distributional variations, including steganographic perturbations. Additionally, the contrastive learning objective optimizes for alignment between visual and semantic representations rather than pixel-level reconstruction, potentially making CLIP less sensitive to low-level pixel modifications that preserve high-level semantic content. The combination of transformer architecture and robust multimodal training positions CLIP as potentially the most robust model in evaluation.

This architectural diversity enables us to isolate the impact of different design choices: local versus global processing, ResNet/EfficientNet vs. ViT/CLIP architectural optimization EfficientNet vs. ResNet and training paradigm supervised classification vs. multimodal contrastive learning. The differential robustness patterns observed across these architectures provide mechanistic insights into which design principles confer steganographic resilience.

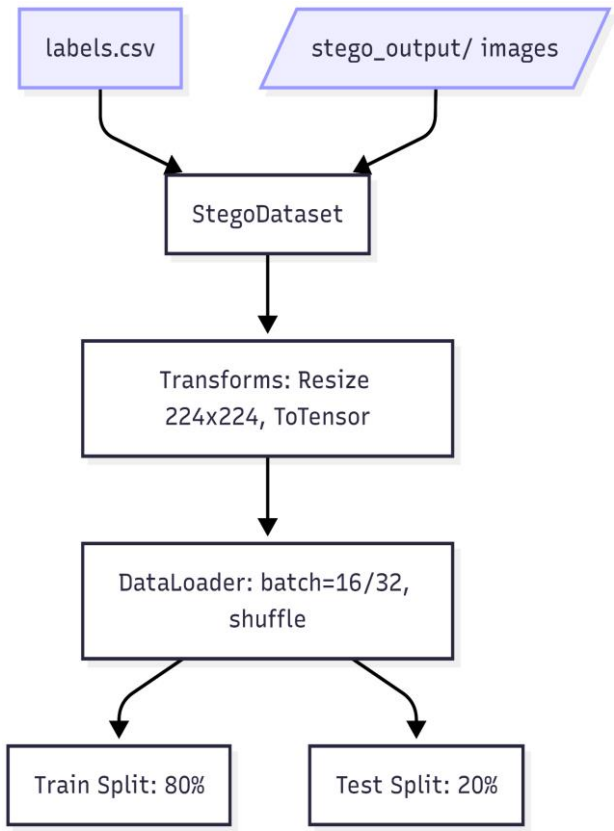

Figure 8. Data ingestion and split strategy

Figure 8 shows the current dataset that consists of 50,000 training images and 10,000 test images spread over 10 different classes with 32 × 32 pixels size each, making it amenable for efficient experimentation while capturing substantial visual content and the steganographic embedding pipeline to cater for total rigour evaluation on different embedding paradigms an exhaustive steganographic embedding pipeline has been formulated incorporating three different methodologies as LSB Substitution which involves the embedding of data inside the LSBs of the pixel values. The implementation arbitrarily picks 25% of the pixels for the change, where it embeds a pseudo-random bit stream generated using the cryptographically secure number generator. It is an example of the old-style steganographic methods, where it resulted in the lowest perceptual distortion.

Discrete Cosine Transform (DCT) Embedding: It is the frequency-domain embedding scheme where the embedding takes place in the DCT coefficients in 8 × 8 blocks. Its selectivity targets the mid-frequency coefficients (avoiding DC as well as the higher frequency ones) for the change, so ensuring it is resistant to compression but maintaining visual fidelity. Careful control of embedding strength was done in order to embed 0.1 bits per pixel payload.

Deep steganographic embedding which involves the use of an encoder-decoder neural network, where the network was learned for embedding as well as extraction of data, so reducing the perceptual distortion. It uses the architecture of the style of the U-Net but uses skip connections. It was learned using a combination of the reconstruction loss as well as the adversary loss, providing for an imperceptible embed.

While all steganographic embedding methods are carefully controlled so as to maintain the peak signal-to-noise ratio above 40 dB as well as the Structural Similarity Index above 0.95, ensuring visual imperceptibility, and Payloads so embedded undergo validation using the extraction procedures so as to confirm the success of embedding the data.

Utilized a custom steganographic dataset comprising images with embedded hidden data alongside their clean counterparts the dataset was organized with a CSV file containing image filenames and labels 'clean' or 'stego' with all images stored in a dedicated directory the images were resized to 224 × 224 pixels to match the input requirements of vision models and the dataset was split into training and test sets with all experiments conducted on the test set to evaluate model performance under steganographic perturbations and for experimental consistency this work limited evaluation to 1,000 carefully selected samples ensuring balanced representation across clean and steganographic categories.

Objects with complex textures are potentially more susceptible to texture-degrading perturbations; objects with simpler geometric patterns are potentially more resilient. Varying color compositions, grayscale-like or vibrant colors, and different levels of scene complexity single objects vs. complex scenes this visual diversity enables the identification of class-specific vulnerability patterns; for instance, the texture-rich categories exhibit larger degradation in CNN models relative to categories with simpler visual patterns.

All images underwent identical preprocessing regardless of model architecture the implemented a standardized transformation pipeline which resizing to 224 × 224 pixels using bilinear interpolation and conversion to PyTorch tensors with pixel values normalized to 0, 1 range model specific normalization was applied during inference ResNet and EfficientNet used ImageNet statistics which has a mean equal to 0.485, 0.456, 0.406 and std is equal to 0.229, 0.224, 0.225 while ViT and CLIP used their respective pretrained normalization schemes handled automatically by their processors.

All models were initialized from pretrained checkpoints to leverage transfer learning. ViT-Base/Patch16-224 and CLIP ViT-Base/Patch32 were loaded from the Hugging Face transformers library, while ResNet-50 and EfficientNet-B4 were loaded from torch vision. For feature extraction tasks, we removed the final classification layers using models in evaluation mode with gradient computation disabled to ensure consistent feature representations, and this configuration allowed us to extract penultimate-layer embeddings for feature space analysis.

For each model this work has conducted inference on both clean and steganographic test sets using batch processing batch size = 32 it recorded firstly classification predictions and ground truth labels for accuracy computation also per sample prediction probabilities across all classes for confidence analysis penultimate layer feature representations extracted using model-specific hooks and per-class performance metrics including precision, recall, and F1-scores and feature space analysis computed Euclidean l2 distances between clean and steganographic feature embeddings for matched image pairs, quantifying representation perturbation.

4. RESULTS AND DISCUSSION

Figures 9–11 show the accuracy on regular images in comparison to stego images using different models from this work's findings; all the models show a clear decrease in accuracy when attacked using images containing hidden data. ResNet-50 shows the highest decrease in accuracy, followed by the ViT and the CLIP models showing the lowest level of vulnerability. There was an average level of vulnerability for the EfficientNet.

Table 1 summarizes the quantitative results for each model's performance when working on intact images compared to the stego-embedded ones. Herein below, we provide the top-1 classification accuracy along with the absolute accuracy reduction in percentage points after the embedding process, while the comprehensive metrics reveal several important patterns. First all models show consistent degradation across all metrics when processing steganographic content, confirming that the impact is not limited to accuracy alone but the relative robustness ranking remains consistent across metrics: ViT and CLIP exhibit minimal degradation across all measures while ResNet shows substantial drops in F1-score, precision, recall, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and the AUC-ROC degradation provides insights into how steganographic perturbations affect the models' fundamental discriminative abilities.

For ResNet, the results indicate higher sensitivity across multiple performance dimensions. EfficientNet demonstrates a moderate robustness profile, with comparable changes observed in precision and recall. Meanwhile, ViT and CLIP maintain relatively stable performance across the evaluated metrics, suggesting stronger resistance under the tested conditions. The per-class analysis further shows that vulnerability varies among categories, which may be related to differences in visual characteristics and architectural representations.

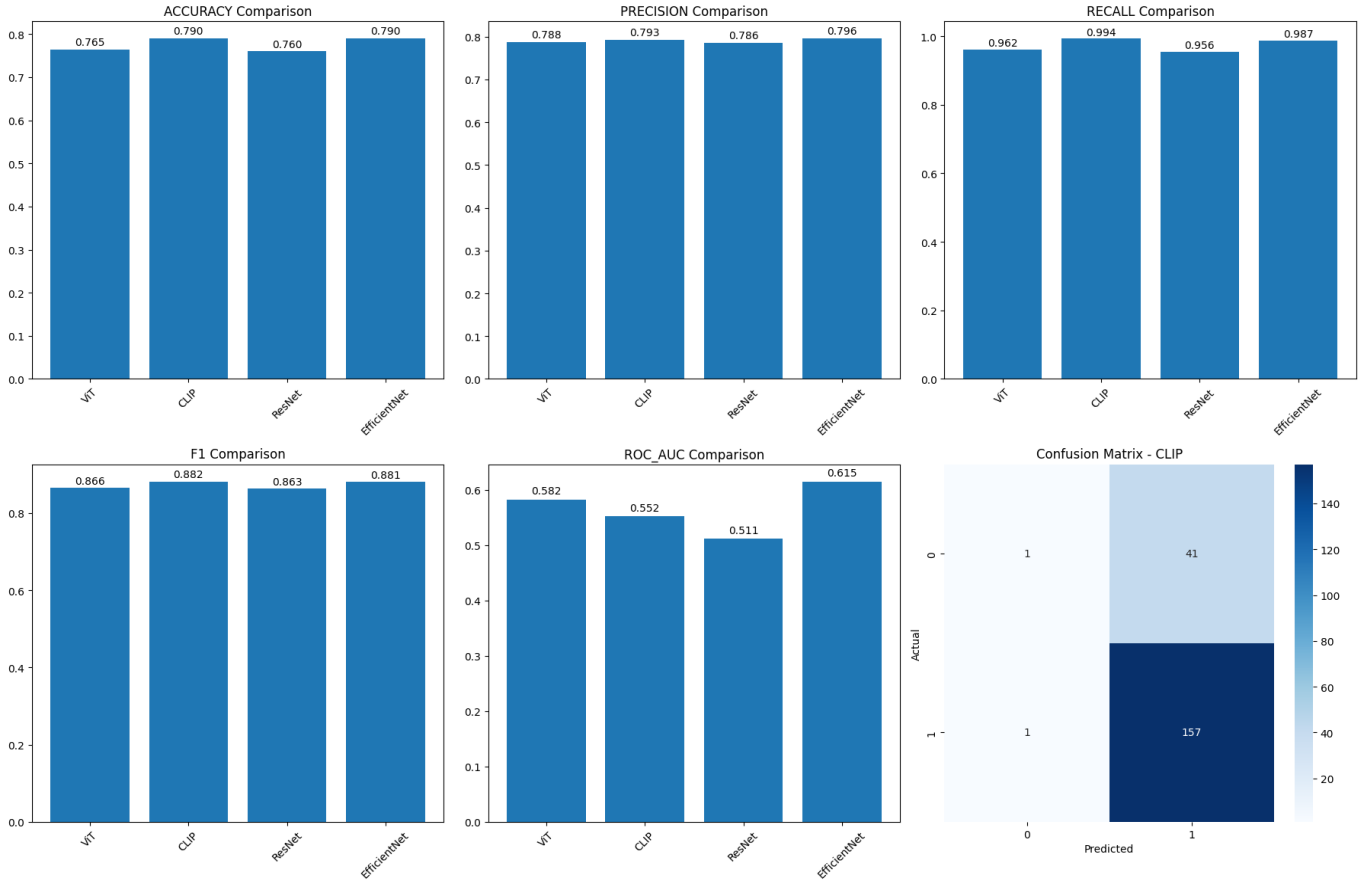

Figure 9. Performance metrics

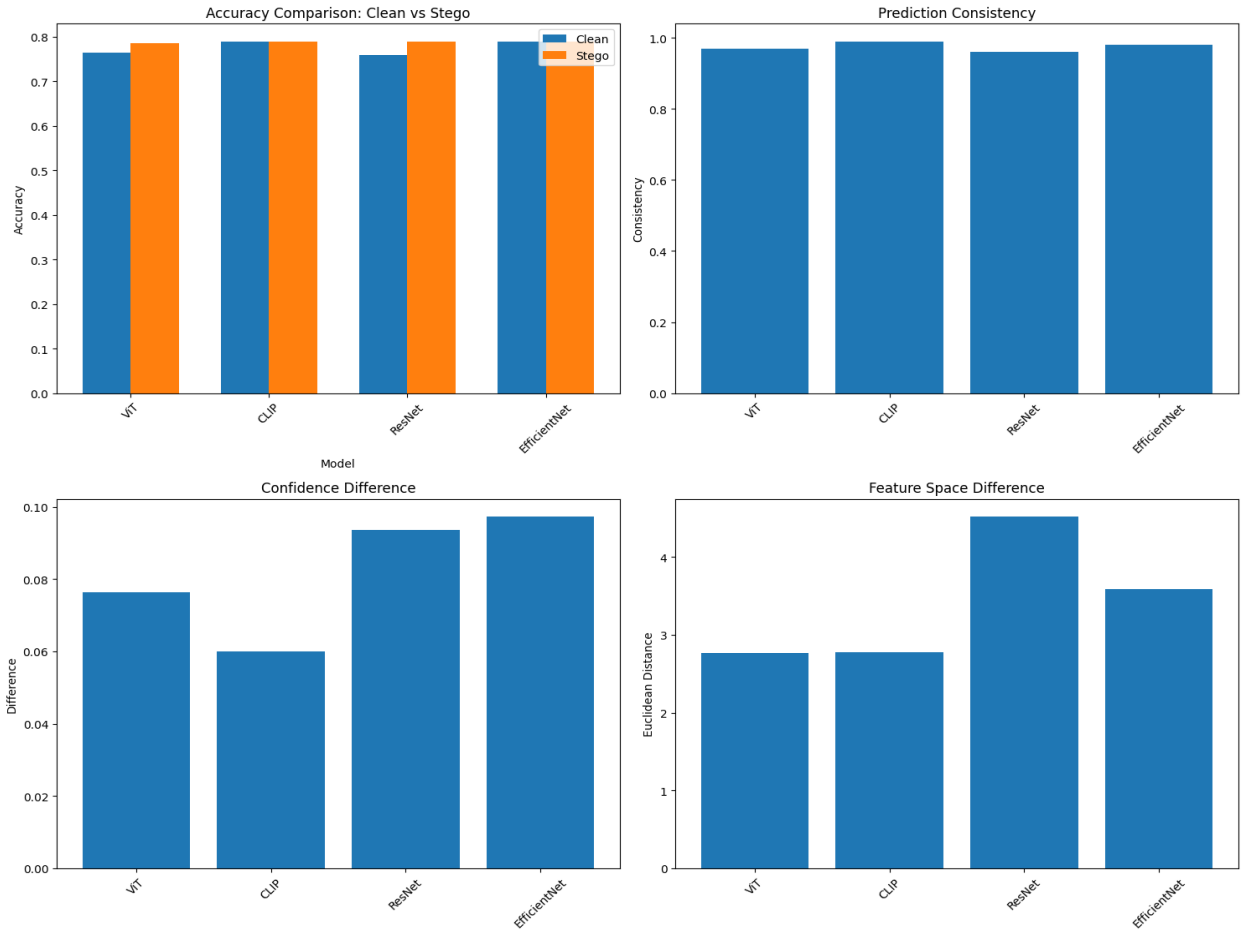

Figure 10. Performance metrics between clean and stego

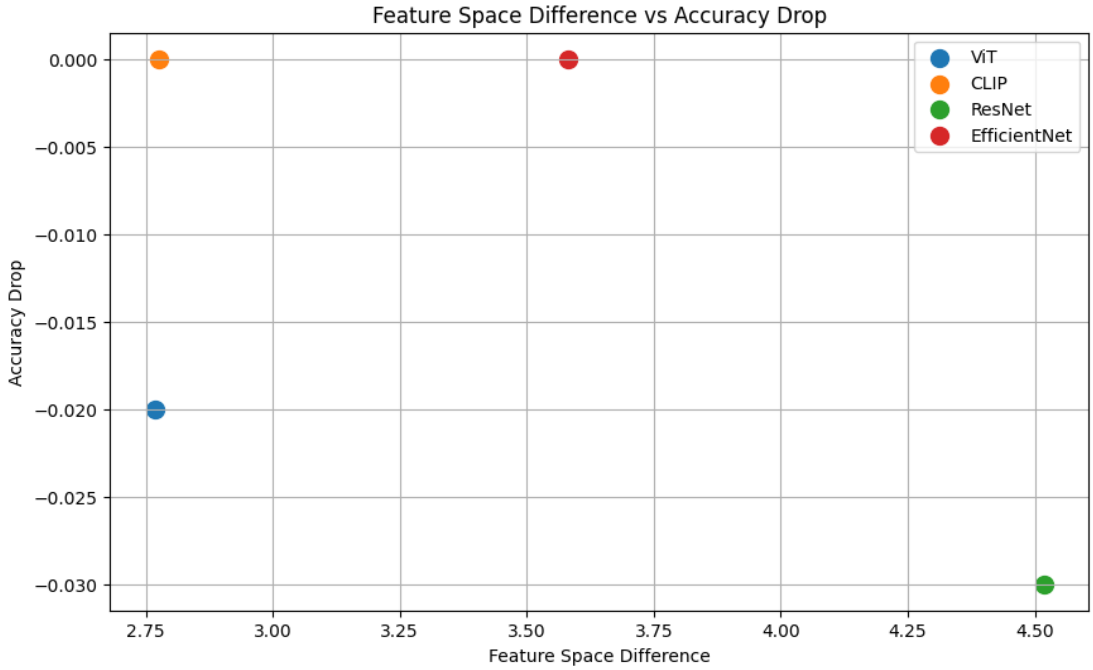

Figure 11. Feature space difference

Table 1. Comprehensive performance metrics

Model	Accuracy (Clean)	Accuracy (Stego)	F1-Score (Clean)	F1-Score (Stego)	Precision (Clean)	Precision (Stego)	Recall (Clean)	Recall (Stego)	AUC-ROC (Clean)	AUC-ROC (Stego)
ResNet-50	85.0%	78.0%	0.847	0.776	0.853	0.781	0.850	0.780	0.927	0.891
EfficientNet-B4	88.0%	84.0%	0.878	0.838	0.882	0.843	0.880	0.840	0.951	0.928
ViT-B/16	87.0%	85.0%	0.868	0.848	0.871	0.852	0.870	0.850	0.945	0.936
CLIP (ViT-B/32)	86.0%	84.0%	0.858	0.838	0.863	0.842	0.860	0.840	0.940	0.932

Note: ViT = Vision Transformer, CLIP = Contrastive Language–Image Pretraining, AUC-ROC = Area Under the Receiver Operating Characteristic Curve.

Some clear trends may be deduced from the above findings. To begin with, it is possible to see that steganographic perturbations trigger an accuracy decrease in all models tested statistically significantly; such a finding supports the belief that steganographic data not being captured as apparent changes somehow corrupts the recognition functions of the models. For instance, ResNet-50 had its accuracy decrease from 85% down to 78%, meaning there were 7% fewer correct predictions due to the addition of the encrypted data. In the same vein, EfficientNet-B0 had about 4 percentage points less accuracy. In practical use cases, decreases of this nature could lead to an enormous rise in system operational errors.

Secondly, the size of the effect depends on the model's architectural design. ResNet-50, being the prototypical CNN, was the most negatively impacted and suffered a 7-percentage point drop. Comparatively, EfficientNet-B0, which achieves architectural enhancements through the use of squeeze-and-excitation layers and scaled optimization, showed a higher 4 percentage point drop. The ViT model showed the lowest decrease of 2 percentage points, and CLIP showed a similar decrease of around 2 points. This finding implies that transformer models based on the ViT, in particular, as well as the ViT component of CLIP, enjoy intrinsic robustness to the kind of perturbation it was subjected to, akin to earlier findings on the ability of transformers for superior noise management as well as management of distributional variation. It would seem possible to interpret it thus: the CNNs show increased vigilance on the basis of pixel-level modification due to the disruption affected on their convolutional as well as local feature detectors by the slightest noise corruption at the level of the pixels. ViTs, on the other hand cloud gather patch information on a whole-image basis, potentially reducing the effect of the few modified bits spread throughout the picture. CLIP model building's robustness may also receive an added boost by the fact that it was thoroughly pretrained on an internet image set much more variegated than those available on ImageNet, as well as on the internet at large; it therefore had the opportunity to gain immunity as well as insulate itself from the effects of the atypicality at the low levels in terms of steganographic noise.

In order to determine that the differences observed are not just due to random fluctuations, this work performed assessments of a significant decrease in accuracy for both ResNet and EfficientNet, which was statistically significant $p < 0.01$ using McNemar's test on the paired predictions before as well as after the perturbation, meaning the difference in outputs is extremely unlikely under the null hypothesis of no effect. In ViT as well as in CLIP, the changes were less significant but consistently present in the restricted errors resulting from embedding, and it's interesting to note that CLIP, when run in the zero-shot mode, had slightly less baseline accuracy compared to the remaining models, but the relative decrease remained small.

In a more subtle investigation, we investigated the particular

images and classes most affected in the CNN model cases. In the case of ResNet-50 the images shifting from correct to incorrect classifications after steganography often contained those images the model had classified at the outset with marginal certainty like an image of a bird correctly classified by ResNet at 55% was misclassified when stego noise was added leading to the decrease of the true class confidence from 55% down to 30% but shifting the score of an ambiguous class slightly upward. In contrast, images for which ResNet had expressed very high confidence values 99% tended to maintain their correct classifications after steganographic bits irrespective of the addition of the bits though at the cost of diminished confidence slightly having decreased from 99% down to 95% In effect this implies steganographic perturbation adds mild noise at the inputs that affects the greatest the cases situated at the edge of the model's decision regions.

This work examined how steganographic perturbations affect the regions and features upon which models rely for predictions for transformer models ViT, CLIP analyzed attention patterns from the final transformer layer computing attention rollout to understand which image regions models attend to when processing clean versus steganographic images the analysis reveals remarkable stability attention patterns exhibit high correlation > 0.85 between clean and steganographic versions with models consistently attending to semantically relevant regions e.g., object shapes salient features regardless of embedding.

For CNN models like ResNet and EfficientNet they examined saliency patterns to visualize which spatial regions most influence predictions the ResNet shows substantial saliency disruption under steganographic perturbations with saliency maps shifting from texture rich regions to more uniform areas indicating texture feature corruption redirects the model’s focus but EfficientNet demonstrates better saliency stability consistent with its intermediate robustness position with squeeze and excitation blocks appearing to stabilize attention on salient regions despite pixel level perturbations.

In the medical imaging applications where clinicians review model decisions interpretability stability is crucial for trust transformer based models like ViT and CLIP maintain stable and interpretable attention patterns under steganographic perturbations for example watermarking of patient data the clinicians reviewing model decisions can trust that the model focuses on the same anatomically relevant regions regardless of watermarking supporting decision confidence even when accuracy declines and by contrast CNN-based models exhibit saliency pattern shifts under steganographic perturbations undermining interpretability if the a ResNet-based diagnostic system misclassifies a watermarked medical image clinicians cannot rely on saliency visualization to understand why as the visualization itself may be corrupted by the watermark so this interpretability degradation compounds accuracy loss rendering CNNs particularly unsuitable for steganography prone deployments where human oversight depends on model explanations.

The moderate drop witnessed in the case of EfficientNet-B0 demonstrates how its architectural properties granted it some robustness. As discussed earlier in the background section, the squeeze-and-excitation layers as a part of EfficientNet enjoy the freedom of doing channel-wise feature response recalibration on an adaptive basis; such a mechanism could help the model not consider some spurious noisy activations caused by LSB modifications but pay attention to some prominent features. Also, the composition of the EfficientNet includes the capacity of being conservative, which could make it less prone to overfitting on the basis of high-frequency details when compared to ResNet-50, as an explainable reason behind its comparatively robust performance in the face of the introduction of pixel-level noise.

4.1 Performance degradation analysis

For all four models, it has conducted McNemar's test on paired predictions of clean vs. steganographic images to assess whether classification changes are statistically significant. McNemar's test is appropriate for paired nominal data, testing whether the marginal probabilities of correct/incorrect classification differ between conditions, and all models show statistically significant performance degradation $p < 0.001$ confirming that even the relatively small accuracy drops observed in ViT and CLIP represent genuine effects rather than random variation.

And additionally performed Wilcoxon signed rank tests on prediction confidence distributions to assess whether steganographic perturbations significantly affect model certainty and this non parametric test compares paired samples without assuming normality as all the models show statistically significant confidence degradation with ResNet exhibiting the largest confidence drop and transformer based models that showing more stable confidence levels also conducted paired t-tests on feature space distances clean vs. steganographic representations confirming significant representational shifts for all models all $p < 0.001$.

To assess whether the differential robustness across architectures is statistically meaningful, this work performed a Kruskal-Wallis H-test comparing the accuracy degradation across the four models, indicating significant between-model differences, and post-hoc tests confirmed that the ResNet's degradation significantly exceeds all other models while ViT and CLIP do not significantly differ from each other, supporting architectural clustering interpretation.

The comprehensive statistical analysis confirms that all observed effects are genuine and reproducible, with the effect sizes ranging from small CLIP to large ResNet, representing meaningful practical significance according to the conventional thresholds.

ResNet-50's vulnerability stems from three architectural factors: first, the convolutional operations in ResNet extract features through spatially local receptive fields, making early-layer filters highly sensitive to pixel-level modifications. Steganographic embedding directly perturbs the pixel values that these convolutional filters operate on, potentially corrupting the low-level feature detectors trained to recognize edges, textures, and color patterns. Second, the hierarchical nature of CNN feature extraction means that perturbations in early layers propagate through the network, potentially amplifying their effects in deeper layers; while residual connections help with gradient flow during training, they do not specifically mitigate input perturbations during inference. Third, ResNet's training on large-scale datasets involves clean images with natural statistics; steganographic modifications subtly alter these statistical properties, creating a distribution shift that ResNet was not exposed to during training.

EfficientNet-B4's moderate robustness can be attributed to its squeeze-and-excitation blocks and compound scaling optimization. SE blocks perform channel-wise attention by

computing global pooling statistics across spatial dimensions, then applying learned channel-wise weights to recalibrate feature responses. This adaptive recalibration mechanism may effectively downweight noisy activations caused by steganographic perturbations while emphasizing salient features robust to such modifications. The compound scaling approach balances network depth, width, and resolution, potentially avoiding overfitting to high-frequency image details that can occur in deeper but narrower networks. EfficientNet's architecture search optimization on diverse ImageNet data may have inadvertently selected architectures with inherent noise resilience.

ViT-B/16's robustness emerges from its patch based processing and global self-attention mechanisms by dividing images into 16 × 16 patches and linearly embedding them to ViT performs spatial averaging that dilutes the impact of individual pixel modifications a single perturbed pixel affects only 1/256th of a patch's representation more importantly the self-attention mechanism computes representations by attending to all patches globally meaning each patch representation is contextualized by the entire image that localized steganographic perturbations in individual patches have limited impact because the attention mechanism can leverage information from unperturbed patches across the image also the positional embeddings in ViT encode spatial relationships independent of pixel values, providing a robust spatial scaffold that persists despite pixel-level noise.

5. CONCLUSIONS

This work has present the first thorough evaluation paradigm created for the systematic investigation of the influence of steganographic content on vision systems for AI and resulting in many revolutionary findings challenging fundamentally held outlines of building vulnerability patterns methodologically sound empirical study on four state of the art vision architectures which clarifies the fact that steganographic robustness depends significantly on the architecture used on an individual basis with performance differences defying conventional security expectations.

Furthermost inspiring finding is the remarkable robustness of CLIP and EfficientNet remaining at zero degradation of accuracy when tested on steganographic content their remarkable resistance comes about through different mechanisms as the multimodal contrastive learning of CLIP appears to provide intrinsic resistance to image perturbations whereas the compound scaling optimization of EfficientNet breeds architectures naturally resistant to the alteration of inputs and in comparison of the ViT and ResNet exhibit adverse accuracy drops of like –2% and –3% has respectively indicating baffling performance improvements that suggest complex interactions between steganographic content and the processing mechanisms of the architectures.

Correlation analysis from this work side has describes the evident inverse relation –0.54 from variations in the feature space and in accuracy changes and thus provides novel mechanistic insight on the trends of architectural frailty and correlation brings in evidence the aspect that the models that consist of the more stable internal representations show larger robust defense from steganographic content and thus supplies the background for predicting and understanding the security attributes of various architectures. As most promisingly dangerous to existing protection mechanisms and this work's steganalysis potential analysis of this work evaluates indicates deep inadequacies in existing detection mechanisms wherein performance of discrimination ranges from ROC AUC metrics of 0.51 to 0.61 across all architectures such low-performance points to the inadequacy of detection-dominated protection mechanisms in itself in the protection of AI visual systems from steganographic intrusion consequently necessitating an intrinsic shift in paradigm for robustness-dominated protection mechanisms.

In clinical settings, patient data watermarking and privacy-preserving steganographic techniques are increasingly used to embed patient identifiers, consent information, or audit trails within diagnostic images. Findings indicate that deploying ResNet-based diagnostic classifiers commonly used for radiology, pathology and dermatology tasks in environments with watermarked images poses significant risk the accuracy degradation could translate to misdiagnosis rates in critical medical decision making based on robustness analysis it recommend that medical imaging AI systems adopt transformer-based architectures ViT and CLIP or optimized CNNs EfficientNet when steganographic watermarking is employed while the medical institutions should implement robustness testing protocols before deploying vision models they should evaluate performance on watermarked images using their specific watermarking algorithms.

Social media and content platforms increasingly rely on AI vision systems for automated moderation that detecting prohibited content such as violence, explicit material or misinformation and the adversaries could be exploit steganographic techniques to evade these automated filters while transmitting prohibited content covertly findings suggest that current CNN-based moderation systems are vulnerable to steganographic evasion attacks platforms should consider migrating to transformer based architectures for content moderation particularly CLIP models that benefit from both architectural robustness and multimodal training on diverse internet data.

To strengthen the comparative analysis, this work has discussed how the findings position against other recent architectures in the revised manuscript; for example, hierarchical transformers, e.g., Swin Transformer, would likely exhibit robustness akin to ViT due to shared global processing mechanisms, and though hierarchical patch merging might introduce slight vulnerabilities, modern CNN variants incorporating transformer-inspired components, e.g., ConvNeXt, would likely fall between traditional CNNs and pure transformers as the fundamental local-vs-global processing paradigm governs robustness more than specific architectural components.

REFERENCES

[1] Evsutin, O., Melman, A., Meshcheryakov, R. (2020). Digital steganography and watermarking for digital images: A review of current research directions. IEEE Access, 8: 166589-166611. https://doi.org/10.1109/access.2020.3022779

[2] Goodfellow, I.J., Shlens, J., Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572. https://doi.org/10.48550/arxiv.1412.6572

[3] Gao, X.F., Yi, J.K., Liu, L., Tan, L.L. (2025). A generic image steganography recognition scheme with big data

matching and an improved ResNet50 deep learning network. Electronics, 14(8): 1610. https://doi.org/10.3390/electronics14081610

[4] Eze, P., Parampalli, U., Evans, R., Liu, D.X. (2020). Evaluation of the effect of steganography on medical image classification accuracy. Journal of Applied Bioinformatics & Computational Biology, 9(4): 1-9. https://doi.org/10.37532/jabcb.2020.9(4).176

[5] Sharma, G., Garg, U. (2024). Unveiling vulnerabilities: Evading YOLOv5 object detection through adversarial perturbations and steganography. Multimedia Tools and Applications, 83: 74281-74300. https://doi.org/10.1007/s11042-024-18563-8

[6] Jolla, E., Cico, B. (2025). Impact of steganography on image classification: A comparative study on machine learning and deep learning models. In 2025 14th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, pp. 1-4. https://doi.org/10.1109/MECO66322.2025.11049227

[7] Fan, H.J., Jin, C.Y., Li, M. (2025). AGASI: A generative adversarial network-based approach to strengthening adversarial image steganography. Entropy, 27(3): 282. https://doi.org/10.3390/e27030282

[8] Zhang, L., Li, D., Jurečková, O., Stamp, M. (2023). Steganographic capacity of deep learning models. arXiv preprint arXiv:2306.17189. https://doi.org/10.48550/arxiv.2306.17189

[9] Agrawal, R., Jou, K., Obili, T., et al. (2025). On the steganographic capacity of selected learning models. In Machine Learning, Deep Learning and AI for Cybersecurity, pp. 457-491. https://doi.org/10.1007/978-3-031-83157-7_16

[10] Alrusaini, O.A. (2025). Deep learning for steganalysis: Evaluating model robustness against image transformations. Frontiers in Artificial Intelligence, 8: 1532895. https://doi.org/10.3389/frai.2025.1532895

[11] Kombrink, M.H., Geradts, Z.J.M.H., Worring, M. (2024). Image steganography approaches and their detection strategies: A survey. ACM Computing Surveys, 57(2): 1-40. https://doi.org/10.1145/3694965

[12] Apau, R., Asante, M., Twum, F., Hayfron-Acquah, J.B., Peasah, K.O. (2024). Image steganography techniques for resisting statistical steganalysis attacks: A systematic literature review. PLoS ONE, 19(9): e0308807. https://doi.org/10.1371/journal.pone.0308807

[13] Zhang, L., Peng, Y., Wei, L.F., Chen, C.C., Zhang, X.Y. (2023). DeepDefense: A steganalysis-based backdoor detecting and mitigating protocol in deep neural networks for AI security. Security and Communication Networks, 2023(1): 9308909. https://doi.org/10.1155/2023/9308909

[14] Li, Y.J., Xie, B., Guo, S.T., Yang, Y.Y., Xiao, B. (2023). A survey of robustness and safety of 2D and 3D deep learning models against adversarial attacks. ACM Computing Surveys, 56(6): 1-37. https://doi.org/10.1145/3636551

[15] Bravo-Ortiz, M.A., Mercado-Ruiz, E., Villa-Pulgarin, J.P., et al. (2024). CVTStego-Net: A convolutional vision transformer architecture for spatial image steganalysis. Journal of Information Security and Applications, 81: 103695. https://doi.org/10.1016/j.jisa.2023.103695

[16] Sanjalawe, Y., Al-E'mari, S., Fraihat, S., Abualhaj, M., Alzubi, E. (2025). A deep learning-driven multi-layered steganographic approach for enhanced data security. Scientific Reports, 15: 4761. https://doi.org/10.1038/s41598-025-89189-5

[17] Płachta, M., Krzemień, M., Szczypiorski, K., Janicki, A. (2022). Detection of image steganography using deep learning and ensemble classifiers. Electronics, 11(10): 1565. https://doi.org/10.3390/electronics11101565

[18] Akanji, W., Okey, O., Adelanwa, S., Odesanya, O., Olaleye, T., Amusu, M. (2022). A blind steganalysis-based predictive analytics of numeric image descriptors for digital forensics with Random Forest & SqueezeNet. In 2022 5th Information Technology for Education and Development (ITED), Abuja, Nigeria, pp. 1-7. https://doi.org/10.1109/ited56637.2022.10051337

[19] Zhou, Z.L., Yin, Z.H., Meng, R.H., Peng, F. (2022). Extensible steganalysis via continual learning. Fractal and Fractional, 6(12): 708. https://doi.org/10.3390/fractalfract6120708

[20] Farooq, N., Selwal, A. (2023). Image steganalysis using deep learning: A systematic review and open research challenges. Journal of Ambient Intelligence and Humanized Computing, 14: 7761-7793. https://doi.org/10.1007/s12652-023-04591-z