International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I11P127

Volume 74 Issue 11, 425-442, November 2026
© 2026 Seventh Sense Research Group®

Original Article

Deep Learning based Multimodal Technique for Steganalysis

Sumit Kumar[1], Anita Ganpati[2]

[1,2]*Department of Computer Science, Himachal Pradesh University, Summerhill, Shimla.*

[1]*Corresponding Author: sumitchandel12@gmail.com*

Received: 30 April 2026 Revised: 18 September 2026 Accepted: 30 September 2026 Published: 25 November 2026

Abstract - *The high pace of electronic communication development has led to increased use of steganography to cover vital information in electronic photographs. Steganographic embedding causes very minuscule distortions that cannot be easily detected by traditional means; hence, steganographic embedding is not easily detected. The article suggests AMFG-StegNet, a steganalysis based on deep learning, to improve the revealing of hidden information within digital pictures. To capture the spatial, spectral, and relational properties of steganographic artifacts, the suggested architecture integrates a Graph Neural Network (GNN), attention mechanisms in the frequency domain, as well as Convolutional Neural Networks (CNNs) for feature removal. The tests use two open-source steganalysis datasets, StegoAppDB and ALASKA2, which include several cover and stego photos generated using different embedding techniques. We compare the proposed model against a number of deep learning models, containing CNN-LSTM, ZhuNet, along with a GNN-based model, and we assess it using the average classification metrics recall, precision, and F1-score. Experiment results reveal that the suggested framework outperforms the baseline models in terms of detection presentation, with an accuracy of 0.984. Additional indications of the strength and high discriminative ability of the architecture are observable in confusion analysis and ROC analysis. Another study, ablation study, also shows that the combination of the graph-based interaction of features and frequency-domain attention is quite effective in image detection. These findings show that AMFG-StegNet could be a viable model that can identify hidden information on digital images that can be used to allow a security of information and digital forensics tool.*

Keywords - *Deep Learning, Digital forensics, Convolutional Neural Networks, Frequency-Domain attention, Graph Neural Networks.*

1. Introduction

The global community is growing exponentially at a pace that requires secure communication as a necessary requirement in guarding data against any form of intrusion [1]. This has led to numerous technical solutions being suggested to address this issue. Cryptography and steganography [2, 3] have been commonly used, and they both utilize different working principles, although the two methods aim at securing the data during the transmission process. Steganography, unlike cryptography, not only makes the transmission unintelligible, making it impossible to intercept, but it also protects the message's substance [4]. When compared to cryptography, steganography has a distinct benefit [5, 6]. Stego media, which contains hidden data, looks identical to the original cover media after data embedding [7].

Common uses of steganography are reported in [8, 9], it is mostly utilized to avoid the unauthorized reception of the hidden message when passing through common networks [10]. Although steganography can be applied as a means of attaining other legitimate reasons such as safeguarding information and the use of secretive communication, steganography can also be misused to serve other illicit activities such as cybercrime, information leakage, and secret communication. This has led to the increased significance of steganalysis as a research expertise in computer forensics and data security.

Steganography is the process of deciphering hidden meanings inside digital information, e.g. images. Steganalysis involves two fundamental procedures, namely, feature extraction and binary classification [11]. Focusing on achieving a balance between accuracy and fluidity, it systematically utilizes methods of statistics, Machine Learning (ML) model, and analytical methodology to identify and discover hidden information within digital content. The stego and rebuilt image's visual quality has been greatly enhanced by these procedures, its disadvantages that include high complexity as well as low payload capacity, have not been clearly understood [12]. With the development of Graphics Processing Unit (GPU) and DL [13] Through steganalysis investigations, it is now feasible to decipher the

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

hidden steganographic payload in digital pictures. Given what is already known from the literature [14, 15], DL has been vital in steganalysis and also in several advanced image analysis tasks. Steganographic systems have been developed based on CNNs, GANs, deep learning models such as autoencoders are not susceptible to steganalysis. [16, 17]. However, most of these models still find it difficult to be resilient in the face of real-life digital security situations [18-20].

Recent studies have been related to multimodal steganalysis, where numerous types of media, including text, audio, as well as images, are analyzed together to improve the detection of concealed data. Multimodal DL models can help in the detection of minor embedding artefacts by using complementing representations with respect to modalities. Some recent works have demonstrated that multimodal frameworks are more successful than single-modality techniques when it comes to detection performance [21-23]. Even the current methods are not able to characterize frequency-domain artefacts and produce connections between suspicious regions, so there is a need to have stronger structures to do reliable steganalysis in the multimedia complicated situations.

2. Research Gap

A close reading of the existing literature reveals three specific and recurring shortcomings that motivate the present work. First, CNN-based detectors such as Xu-Net, Yedroudj-Net, SRNet, and ZhuNet learn spatial embedding traces effectively but do not explicitly model frequency-domain artefacts, so they lose sensitivity once the payload is concentrated in high-frequency sub-bands or the stego image is recompressed. Second, attention-based and hybrid CNN designs improve local feature weighting but treat each image region independently, so they cannot capture the distributed, correlated pattern that steganographic embedding leaves across spatially separated regions of the cover. Third, GNN-based approaches (e.g., GraphSAGE-based speech steganalysis) demonstrate that relational modelling of suspicious regions improves detection, but they have so far been applied within a single modality and do not combine relational learning with an explicit, adaptive frequency-domain representation.

Consequently, no existing framework in the reviewed literature jointly learns spatial, spectral, and relational cues within a single trainable pipeline, which leaves models exposed to modality-specific artefacts they were not designed to see. AMFG-StegNet is built specifically to close this gap: unlike prior single-branch or single-modality designs, it couples an adaptive frequency decomposition module with parallel spatial-CNN and frequency-attention branches and fuses their outputs through a graph neural interaction layer, so that spatial anomalies, spectral distortions, and inter-region dependencies are all captured and combined before classification. This combination, and its empirical validation through the ablation study reported in Section 5, is the principal novelty of the proposed work relative to the CNN-only, attention-only, and GNN-only baselines discussed above.

In order to overcome these difficulties, this study introduces the Adaptive Multimodal Frequency-Graph Steganalysis Network (AMFG-StegNet) to retrieve concealed information in the various forms of media. The structure takes input data that can be viewed as multisensory, such as images, audio spectrogram, and coded text features. First, a frequency decomposition module takes out all high and low-frequency components, and it is in this frequency component where the steganographic artifacts are most likely to exist. These characteristics are then fed to two parallel streams, one of them being a spatial CNN to detect local pixel-level anomalies and the other being a frequency attention module to identify spectral distortions that may have been added in data embedding. The features extracted are further refined with a graph neural interaction layer, which models the relationship between suspicious regions in order to bring to focus any hidden patterns. The fused multimodal structures are then transmitted through fully associated layers, whereby they are classified to provide accurate information in regard to presence of hidden information.

2.1. Contributions of the Study

The contributions of the research were as follows:

- Suggestion of a new AMFG-StegNet model of multimodal steganalysis.
- As part of the adaptive frequency decomposition module that should be introduced to remove frequency-domain artifacts in steganography.
- Combination of spatial CNN and frequency attention modules in order to identify small embedding distortions.
- The addition of a graph neural interaction-based approach to feature relationship modeling and enhanced identification of hidden patterns.
- Evidence of improved accuracy and strength over the traditional single-modality steganalysis techniques.

Organization of the Paper:

Here is the configuration of the rest of the paper: Section 2 will examine relevant work and identify areas where research is lacking. Refer to Section 3 for background information on the proposed models. The strategy is laid forth in Section 4. In Section 5, we provide and discuss the findings, comparisons, and experiments. The inquiry is concluded, and the following steps are outlined in Section 6.

3. Literature Review

In the fields of information security as well as digital forensics, steganalysis plays a vital role in the discovery of hidden data inside digital images. Statistical and perceptual variables like Image Quality Measures (IQMs) [24], along

with image moments [24] classic ML classifiers [12]. were the mainstays of old steganalysis methods. Although these hand-crafted methods outperformed a few steganographic strategies, they couldn't generalize well across different embedding schemes, picture origins, or payload dimensions.

The introduction of DL, and specifically CNNs, revolutionized steganalysis by removing the requirement for manual feature extraction and allowing for the end-to-end learning of hierarchical spatial connections [25, 26]. By hierarchically netting complicated residual connections and minor implanting errors in spatial and transform domains, approaches based on Convolutional Neural Networks (CNNs) considerably enhanced detection performance.

Limited generalizability is a result of Xu-Net's problems with altered images, despite the fact that it provides an early CNN-based steganalysis model for high-frequency artefact analysis. [27]. Although Yedroudj-Net's feature learning capabilities are enhanced by its pre-processing layers, the system is vulnerable to attacks that include resizing or compression [28]. In order to make steganographic detection more difficult, two complicated architectural models have been developed: ResNet and SRNet. SRNet excels in multi-scale spatial studies, while ResNet uses residual connections to keep deep feature representations [29-31].

Diverse activation functions [25], depthwise separable convolutions [26], selection-channel wakefulness [32], skip associates [33], hybrid as well as attention-based network designs [34, 35], and advanced data augmentation methods [36]. Convolutional Neural Network (CNN) models continue to face challenges in terms of efficiency and real-time application because to their increasing architectural depth and high processing needs. This is particularly true for high-resolution image dataset.

Table 1. Comparison with existing study

Ref	Method	Dataset	Key Results	Limitations
[37]	Deep-CNN framework for steganography detection	10,000 images (WOW, S-UNIWARD, HUGO)	Improved binary classification using fine-tuned CNN architecture	Limited capability for universal steganalysis
[38]	Hybrid CNN-NLP model for LSB steganography detection	5,000-12,500 grayscale images	Achieved 99.96% accuracy with semantic filtering to reduce false positives	Mainly focused on LSB-based steganography
[39]	CNN, RNN, and GAN comparative deep learning models	ALASK 2 dataset	CNN achieved 99.52% detection accuracy, GAN improved adversarial detection	High computational complexity
[40]	Graph Neural Network (GraphSAGE) for speech steganalysis	VoIP speech datasets	Achieved >98% accuracy with efficient detection time	Applicable mainly to audio steganalysis
[41]	CNN architecture with mixed pooling and separable convolution	Image steganography datasets	Improved detection accuracy by 10.2% over existing methods	Limited detection for very low payloads
[42]	Hybrid deep network (ZhuNet-ATT-BiLSTM) with attention	BOSSBase, COCO	Improved feature utilization and classification accuracy	High model complexity
[43]	Robustness evaluation of DL steganalysis models	BOSSBase dataset	EfficientNet showed better robustness against transformations	Significant performance degradation under noise and compression
[44]	CNN-LSTM fused architecture for medical image steganalysis	BOSSBase, BOWS, ALASKA2	Improved feature correlation learning and generalization in sensitive medical datasets	Increased computational overhead

While the detection of hidden information has been greatly enhanced by deep learning-based steganalysis techniques like CNNs, attention mechanisms, LSTM, and GNNs, the majority of these techniques concentrate on single media types (audio or images) and face difficulties like low payloads, transformations, compression, and subtle embedding patterns. Current models are limited in their application in real-time and multimodal settings because to their high computing requirements or lack of generalization across various steganography techniques. Spatial, spectral, and relational information have never been combined in a single paradigm for multimodal steganalysis. As summarised in the comparison table above, deep-CNN and hybrid CNN-NLP models [37, 38] report strong accuracy but remain confined to LSB or single-embedding-scheme detection; CNN/RNN/GAN comparative models [39] and ZhuNet-attention-BiLSTM designs [42] improve feature utilisation at the cost of high computational complexity; the GraphSAGE-

based model [40] confirms the value of relational learning but only for audio steganalysis; and CNN-LSTM fusion models [44] improve generalisation within a single domain without addressing frequency-domain artefacts. None of these works combine adaptive frequency decomposition, spatial CNN analysis, and graph-based relational modelling within one trainable architecture, which is precisely the gap this study targets. To seal these loopholes and enhance detection accuracy and resilience relative to the methods above, this paper introduces the AMFG-StegNet, which is effective in capturing the subtle patterns on text characteristics, audio spectrograms, and images.

4. Background

An outline of the suggested framework's deep learning models along with architectures is given in this subsection, including CNN, attention mechanism, and Graph Neural Networks (GNNs). These models combine to extract spatial, spectral and relationship properties of multimodal digital content, ensuring that small variations and latent patterns that can be induced by steganographic embedding are easier to locate. The framework can easily and successfully detect concealed information in pictures, audio, and text data using such methods.

4.1. CNN

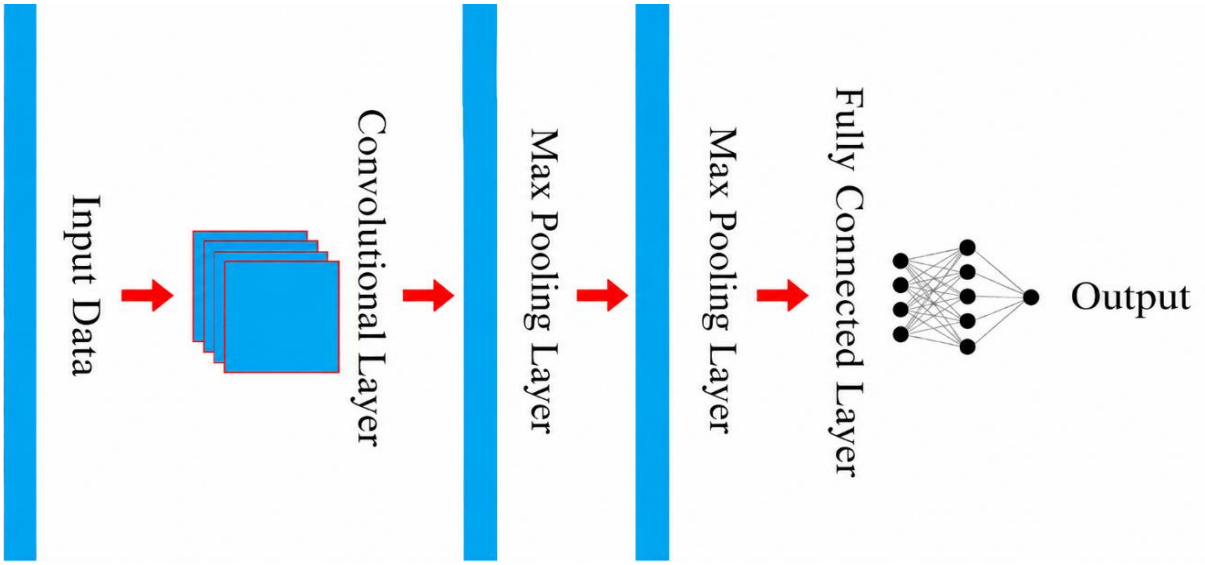

Fig. 1 Typical CNN model

A typical Convolutional Neural Network (CNN) has four layers: convolutional, pooling, flattening, and fully connected, each layer is intended to fulfil a particular function in feature extraction and classification, as explained in [45] Shown in Figure 1.

The fundamental characteristic is the convolutional layer, which makes use of kernels that are learnable and moving windows to extract local features without having to perform it with much computing complexity by sharing weights. Each kernel calculates certain patterns on the input data, e.g. edges or textures pertaining to target recognition.

The pooling layer lowers the computational cost and preserves the important characteristics while decreasing the spatial dimensions of the feature maps. It achieves this without overfitting. Max pooling as well as average combining are two popular methods of combining data, both of which contain details about nearby neighbourhoods. The flattening layer connects the pooled feature maps to the fully linked layer by mapping them to a one-dimensional vector. These characteristics get mapped on the output space using the fully connected layer and a complicated association that is learnt using the weights and the biases.

Equations:
Convolutional Layer:

$$F = ReLU(I * K + b_1) \quad (1)$$

Where X is the input, K represents the set of kernels, b_1 represents the bias term, and $*$ represents convolution.

Pooling Layer:

$$P = maxpool(F, n \times n, s) \quad (2)$$

Where n × n is the kernel size of pooling and s is the stride.

Flattening Layer:

$$v = flatten(P) \quad (3)$$

Where P represents the pooled feature maps.

Fully Connected Layer:

$$h = Wv + b_2 \quad (4)$$

Where W was the weight matrix, b was the bias vector, and h was the resulting hidden representation.

Output Layer:

A softmax function generates class probabilities:

$$\hat{y}_i = \frac{e^{h_i}}{\sum_j e^{h_j}} \quad (5)$$

Where $\hat{y}_i$ was the predicted probability for class i.

In this study, CNNs are selected because they have demonstrated success in detecting small spatial anomalies that exist in digital media.

The hierarchical feature extraction capability of CNNs enables them to detect pixel-level distortions which occur during data embedding. Secondly, CNNs have the capability to patterns of frequencies, such as high- and low-frequency responses, a feature that assists in detecting concealed distortions produced by steganographic embedding. This feature qualifies CNNs to be a practical component of the proposed multimodal steganalysis model.

4.2. Attention Mechanism

The attention mechanism is based on the human mental ability to concentrate on the most useful information using minimal effort [46]. In order to make the final judgment based on the feature's significance, the Self-Attention module smears common biases and weights to the retrieved feature vectors. The features are aggregated by a weighted summation strategy, which allows deep-level feature mining and focusing on the most informative parts.

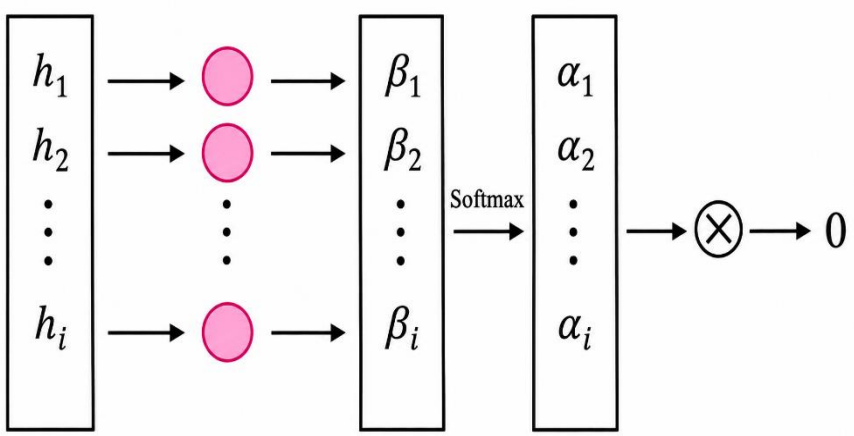

Fig. 2 Attention network structure diagram

For the purpose of calculation, the formulas that is as follows is utilized:

$$\beta_i = \sigma(W_i F_i + b_i) \quad (6)$$

$$a_i = \text{softmax}(\beta_i) = \frac{\exp(\beta_i)}{\sum_j \exp(\beta_j)} \quad (7)$$

$$O = \sum_i a_i \odot F_i \quad (8)$$

The extracted feature vector from the CNN or frequency-attention module is represented by F_i while β_i functions as the activation value and a_i acts as the softmax normalized attention weight. The weighted sum of features is represented by O while W_i and β_i and σ function as the weight matrix, as well as bias vector and activation function correspondingly. The suggested architecture applies features retrieved through the Convolutional Neural Network (CNN) and the frequency-attention module to dynamic weight coefficients generated by an attention mechanism. The weighting system enables the model to identify essential patterns which reveal concealed steganographic evidence, which leads to better prediction results.

4.3. Graph Neural Networks (GNNs)

A specific kind of neural networks called GNNs was created to learn from graph-structured data, in which nodes stand in for entities and edges indicate connections or interactions between them. A GNN's fundamental propagation rule is as follows:

$$h_v^k = F\left(\sum_{u \in N(v)} w_{uv}^{k-1} h_u^{k-1}, h_v^{k-1}\right) \quad (9)$$

In this case, the hidden states of node v at layers k and $k-1$ vare represented by h_v^k vand h_v^{k-1}, respectively. The neighbourhood of node v is represented by N(v), and w_{uv}^{k-1} is the weight of the edge in the $k-1$ th layer that connects nodes v_v and v.

This function F represents an activation function that is not linear. This update rule improves the hidden representation of the current node by combining data from nearby nodes.

Gated Graph Neural Networks (GGNNs), Graph Convolutional Networks (GCNs), Graph Sample and Aggregation (GraphSAGE), as well as others, comprise a popular family of GNN models. To capture more intricate interactions inside the graph, it is common for these designs to include tweaks to the basic propagation rule.

In steganalysis, GNNs can be applied to simulate relationships between suspicious regions or features derived out of digital media, and help detect hidden nuances that could be the signs of steganographic embedding. The schematic diagram of the GNN structure is represented by Figure 3.

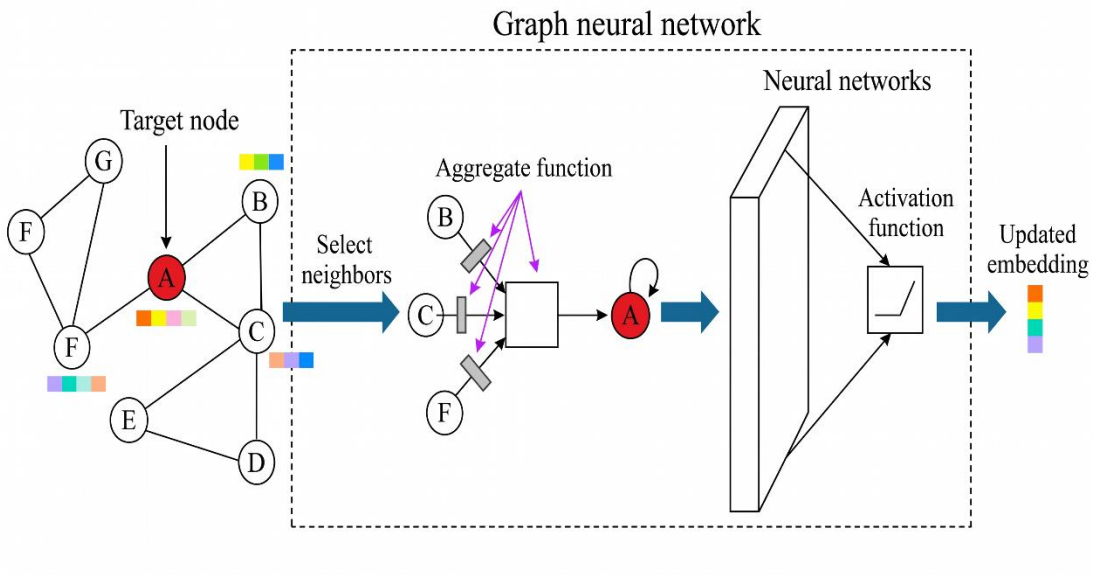

Fig. 3 GNN architecture

5. Methodology

Finding hidden data in digital media is the goal of this research's technique, which is structured like a systematic procedure. Gathering datasets, cleaning them up, using AFD, extracting multimodal features, graph-based relational learning, feature fusion, and lastly classifying them are all steps in the system. The aim is to detect fine anomalies that occur in steganographic embedding through the joint analysis of spatial, spectral and relational features of the input media. The study suggests the AMFG-StegNet in order to do so. The framework consists of three essential components, which include a CNN that detects spatial distortions and a frequency attention module that highlights spectral artifacts, and a GNN that shows how suspicious feature areas interact with each other. The system enables complete detection of concealed embedding techniques through its integrated design.

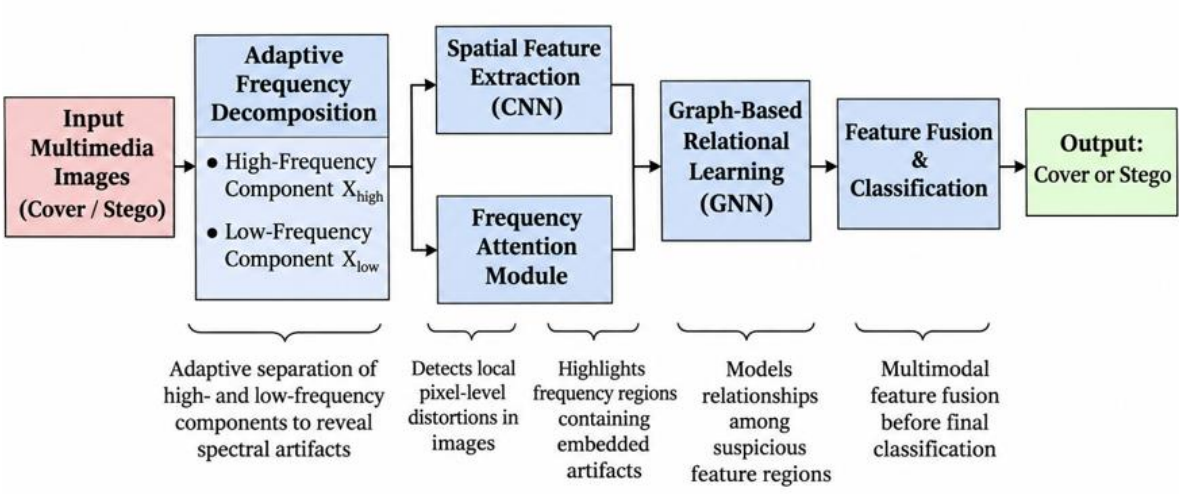

Fig. 1 Overview of the proposed framework

The process starts with the gathering and processing of multimedia data, which includes both cover samples and stego samples. The system uses frequency component breakdown to identify possible hidden data extraction points within the input material. The dual branches of the system utilize parallel CNN and frequency attention techniques to extract both spatial and spectral information from the decomposed dataset.

A graph neural interaction layer enhances the features by capturing interdependencies between suspicious areas. The system combines extracted features and processes them through fully connected layers to execute binary classification for hidden data detection.

Figure 4 demonstrates the architectural architecture of the proposed framework, whereas Algorithm 1 shows its whole process.

5.1. Dataset Collection

To assess the efficacy of the proposed steganalysis technique, two open-source datasets from Kaggle have been utilized. These datasets can teach the model the peculiarities of hidden information embedding because they can offer a variety of cover and stego images.

The initial dataset that is utilized in this work is the ALASKA2 steganalysis dataset that has become a standard in steganography research:
https://www.kaggle.com/datasets/meaninglesslives/alaska2trainvalsplit/data

The ALASKA2 dataset can contain a substantial number of high-resolution pictures that are manipulated using different steganographic methods. The dataset consists of about 80,000 images comprising of cover and stego images. Each picture is connected with metadata which is stored in CSV files. Labels indicating whether the copy is a cover image and a stego image, as well as information on the kind of embedding, are included in this metadata. The administration of databases is improved by the provision of structured metadata to facilitate simpler supervised learning and ease of administration.

Additionally, this study makes use of the Stego Images Dataset, which may be accessed at:
https://www.kaggle.com/datasets/marcozuppelli/stegoimagesdataset/data

This collection contains images that have hidden information hidden by the most common steganographic techniques. Approximately twenty thousand photos, containing both the original cover art as well as stego images, make up this dataset. It supplies both the original cover photos and the stego pictures so that the classic can tell the variance between the changed media and the natural image statistics.

The suggested methodology is more generalizable across various steganographic distortions due to the dataset's diverse texture, resolutions, and visual patterns. A more robust and reliable steganalysis method is the result of training the proposed model on a variety of multimedia examples drawn from many datasets.

5.2. Data Preprocessing

In order to keep data consistent as well as improve the suggested system's performance, many preprocessing activities are carried out prior to model training. To preserve data quality, faulty images, missing records, and duplicate samples are first eliminated from the collection. All images are downsized to a consistent dimension so that the neural network can analyse them effectively, since images from various sources may have varied resolutions. Pixel values are normalized to a constant numerical range in order to stabilize the learning process.

The definition of the normalization process is

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

Where x is the original pixel value, along with x_{min} and x_{max} indicate the lowest and highest intensities of the pixels. This transformation increases numerical stability during model training by scaling the input values between 0 and 1, as seen in Equation (10).

Using the proper encoding methods, information from CSV files is transformed into numerical representations for the ALASKA2 dataset. After preprocessing, the dataset is separated into testing, validation, as well as training subsets to confirm an objective evaluation of the proposed model.

5.3. Adaptive Frequency Decomposition

The host medium is usually altered by steganographic embedding processes in a subtle way that may not be readily apparent in the spatial domain. However, these modifications typically introduce noticeable artifacts in the frequency domain. To isolate frequency components that may contain hidden embedding patterns, the proposed approach incorporates an adaptive frequency decomposition module.

The input media X is transformed into its frequency representation as follows:

$$X_f = F(X) \tag{11}$$

Where $F(\,)$ denotes a frequency transformation operator, such as the Discrete Cosine Transform (DCT) or Discrete Fourier Transform (DFT). This transformation converts the spatial signal into its spectral representation, as shown in Equation (11).

The input signal is alienated into its low-frequency along with high-frequency mechanisms after the transformation as:

$$X = X_{\text{low}} + X_{\text{high}} \tag{12}$$

Where X_{high} represents the high-frequency component containing fine details such as edges and textures, and X_{low} shows the low-frequency component containing global structural information.

Separating these components allows the system to analyze subtle variations that are often introduced during steganographic embedding, as expressed in Equation (12).

5.4. Spatial Feature Extraction using CNN

After frequency decomposition, Convolutional Neural Networks (CNNs) examine the input media's spatial representations. CNNs automatically learn hierarchical feature representations that capture local spatial patterns, making them highly effective in image analysis tasks. In the context of steganalysis, hidden information embedding often modifies pixel values slightly, which may result in subtle spatial anomalies such as abnormal textures or noise fluctuations. Convolutional operations efficiently capture these distortions.

The convolution process in the CNN, which generates feature maps, is expressed as:

$$F^{(l)} = \sigma(W^{(l)} * X + b^{(l)}) \tag{13}$$

where $W^{(l)}$ represents the convolution kernel at layer $l, b^{(l)}$, represents the bias term, X represents the input feature map, and σ is the nonlinear activation function.

As described in Equation (13), the convolution operation applies learnable filters to the input feature map to extract spatial patterns related to hidden data embedding.

The produced feature maps undergo a pooling process to cut down on computational complexity while keeping just the most important characteristics. What this means is that:

$$P = \text{maxpool}(F) \tag{14}$$

The max-pooling operation selects the maximum activation value within a local region, thereby preserving the most discriminative spatial features while reducing the dimensionality of the representation, as shown in Equation (14).

Through these operations, the CNN branch captures pixel-level anomalies, structural inconsistencies, and noise variations introduced during the steganographic embedding process.

5.5. Frequency Attention Mechanism

Certain steganographic artifacts are more prominent in the frequency domain, even though CNNs are effective at capturing spatial patterns. To highlight these spectral features, the proposed system incorporates a frequency attention mechanism that assigns different importance weights to frequency components.

For each frequency feature representation, the attention score is computed as:

$$\beta_i = \sigma(W_i F_i + b_i) \tag{15}$$

Where W_i denotes the learnable weight parameter, b_i denotes the bias term, and F_i denotes the ith frequency feature vector. The attention score β_i represents the relevance of each frequency feature, as indicated in Equation (15).

The attention scores are normalized by the softmax function to obtain the corresponding attention weights:

$$a_i = \frac{e^{\beta_i}}{\sum_j e^{\beta_j}} \tag{16}$$

For all frequency characteristics, the attention weights should follow a probability distribution, and the softmax function makes sure of that Equation (16).

Finally, the frequency features are combined according to their attention weights to produce the weighted spectral representation:

$$O = \sum_i a_i F_i \tag{17}$$

As described in Equation (17), this weighted aggregation allows the network to emphasize frequency regions that contain stronger embedding artifacts while suppressing irrelevant components.

5.6. Graph Neural Interaction Layer

Steganographic artifacts often appear as correlated patterns distributed across multiple regions of the media, even though spatial and spectral features individually provide valuable information. To capture these dependencies, the proposed framework incorporates a GNN interaction layer. In this graph representation, each suspicious feature region identified by the CNN and frequency modules is treated as a node, where edges denote the connections between these groups of areas. By use of a message-passing system, the GNN is able to update node representations by combining data from nearby nodes. The propagation rule for updating node embeddings is defined as:

$$h_v^{(k)} = F(\sum_{u \in N(v)} w_{uv}^{(k-1)} h_u^{(k-1)}, h_v^{(k-1)}) \tag{18}$$

Where $h_v^{(k)}$ denotes the node embedding at iteration $k, N(v)$ denotes the set of neighboring nodes, as well as w^{uv} denotes the edge weight between nodes u and v. As shown in Equation (18), the GNN integrates information from neighboring nodes with the current node representation, enabling the model to capture contextual relationships among suspicious regions. This graph-based learning mechanism permits the system to detect distributed embedding designs that might not be noticeable when regions are analyzed independently.

5.7. Feature Fusion and Classification

The proposed framework will follow the extraction of spatial, spectral, and relational features by incorporating them into a single multimodal feature vector. The combination of these complementary features permits the model to be able to examine the patterns of hidden information in a more holistic way.

The fusion process of features is stated as

$$F_{fusion} = F_{cnn} \oplus F_{freq} \oplus F_{graph} \quad (19)$$

Where F_{cnn}, F_{freq} and F_{graph} are the graph, frequency and spatial feature vectors, respectively. As used in our equations, the operator $\oplus$ gives the concatenation of feature vectors as determined in our equations (Equation (19)).

The fused representation is then subjected to fully connected layers in order to do the classification. The final prediction is made using the softmax classifier.

$$y = \text{softmax}(WF_{fusion} + b) \quad (20)$$

W denotes the weight matrix and b denotes the bias term. The following equation (Equation (20)) shows how the softmax function converts the network output into prospects of the different classes, as the name implies. According to such probabilities, the system decides the input media is cover class (no hidden information) or stego class (has concealed information).

5.8. Training Procedure

Using supervised learning with labeled dataset samples, the suggested model is trained. When training a network, its parameters are fine-tuned according to the Binary Cross-Entropy (BCE) loss function, which associates the actual class labels to the anticipated probabilities. Here is one way to describe the BCE loss.

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (21)$$

Where N represents the total number of samples, y_i denotes the true label, and $\hat{y}_i$ denotes the forecast probability for the ith sample. By minimizing the loss function, the model can obtain the best parameters to differentiate a cover and stego media (Aslan et al. 2019).

Adam optimizer will be used to update the model parameters because it is adaptive in learning and it converges faster. Adam optimizer can be represented as a rule of parameter update, as provided in Equation (22),

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (22)$$

Where θ_t represents the model parameters, η denotes the learning rate, and $\hat{m}_t$ and $\hat{v}_t$ are the first in addition to second moment estimates of the gradients correspondingly. Also, early termination and validation checking are implemented in the course of training so as to avert overfitting and provide consistent model convergence.

5.9. Model Evaluation

Using a battery of industry-standard metrics, we test how well the proposed AMFG-StegNet performs. The model's capacity to uncover hidden information in multimedia data is evaluated by these metrics. Since the suggested framework employs binary classification on cover and stego media, the confusion matrix-a summary of the outcomes anticipated by the classification model-will serve as the basis for this evaluation.

A confusion matrix is comprised of four primary elements: True Positives (TP), True Negatives (TN), False Positives (FP), as well as False Negatives (FN). The number of cover samples that will be successfully recognized as cover is denoted by TN, whereas the number of stego samples that will be correctly identified as stego is shown by TP. FP stands for cover samples and FN for stego samples that were mistakenly identified as such. These factors form the basis of the determination of different performance measures.

One of the key metrics, which are used in the research is accuracy, because it is the percentage of properly recognized samples relative to the total number of predictions produced by the model. Accuracy is a general indicator of the effectiveness of the model to distinguish between cover and stego media. Accuracy is defined as follows.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (23)$$

Accuracy was used to determine the number of accurate predictions divided by the number of predictions, as shown in Equation (23).

Another measure used is precision, which evaluates the reliability of positive forecasts. Precision in steganalysis is the relationship between the percentage of samples which are matched as stego.

Precision is calculated using Equation (24).

$$Precision = \frac{TP}{TP+FP} \quad (24)$$

A greater precision value shows that the model makes few false positive detections.

Recall is also called sensitivity; It's the model's capacity to accurately detect stego samples from the given data. It is an expression of the model detection ability of the hidden information in multimedia files. Recall is expressed as

$$Recall = \frac{TP}{TP+FN} \quad (25)$$

As indicated in the equations below, recall is the percentage of the stego samples correctly identified out of all the real stego samples (Equation (25)).

In order to weigh off between accuracy and recall, F1-score is employed as an integrated measure. When evaluating classification models, the F1-score really shines when both false positives as well as false negatives can be included. This term is

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (26)$$

The F1-score represents the harmonic mean of recall and precision, as shown in the formula below (Equation (26)).

The ROC curve, which represents the proposed model's categorization at various decision thresholds, is also used for evaluation.

Equation (27) defines the TPR, which serves as a measure of the proportion of true stego samples that were successfully detected. Equation (28) also defines FPR, which shows the fraction of cover samples that are incorrectly labeled as stego.

$$TPR = \frac{TP}{TP+FN} \quad (27)$$

$$FPR = \frac{FP}{FP+TN} \quad (28)$$

One scalar measure, the Area Under the ROC Curve (AUC), represents the model's total discriminating power. A higher Area Under the Curve (AUC) designates that the model performed better at distinguishing between stego and cover samples.

The suggested model's presentation may be thoroughly evaluated using these evaluation measures, which assess the classification accuracy, detection capability, and prediction reliability.

Algorithm 1. AMFG-StegNet Steganalysis Framework

Require: Multimedia dataset Dcontaining cover and stego samples.
Ensure: Trained AMFG-StegNet model for hidden information detection
1: Load multimedia dataset $D = \{X_i, y_i\}$from ALASKA2 and Stego Images
2: Remove corrupted, duplicate, and incomplete samples
3: Resize all images to a uniform dimension
4: Normalize pixel values using Equation (10)
5: Convert metadata labels into numerical form
6: Divide the dataset into groups for testing, validation, and training.
7: Input multimedia image X
8: Transform the image into frequency domain using Equation (11)
9: Decompose the signal into low-frequency along with high-frequency components using Equation (12)
10: Forward decomposed features to spatial and frequency branches
11: Initialize CNN layers with convolution filters
12: Extract spatial feature maps using the convolution operation in Equation (13)
13: Apply activation function to enhance nonlinear patterns
14: Reduce feature dimensionality using max-pooling Equation (14)
15: Generate spatial feature representation F_{cnn}
16: Extract spectral feature vectors from decomposed signal
17: Compute attention scores for each frequency component using Equation (15)
18: Normalize attention weights using softmax Equation (16)
19: Generate weighted spectral representation using Equation (17)
20: Obtain frequency feature vector F_{freq}
21: Identify suspicious regions as graph nodes
22: Construct edges representing relationships among nodes
23: Perform message passing and update node embeddings using Equation (18)
24: Generate relational feature representation F_{graph}
25: Concatenate spatial, spectral, and graph features using Equation (19)
26: Pass fused features to fully connected layers
27: Compute classification probabilities using softmax Equation (20)
28: Predict final class label (Cover / Stego)
29: Compute Binary Cross-Entropy loss using Equation (21)
30: Optimize network parameters using Adam optimizer Equation (22)
31: Apply validation monitoring and early stopping until convergence
32: Compute evaluation metrics using Equations (23)-(28)
33: Return trained model and performance metrics.
34. End

6. Result and Discussion

In this part, we offer the experimental evaluation of the proposed AMFG-StegNet and compare its results to those of many existing DL methods, including ZhuNet, CNN-LSTM, and a GNN-based model. We evaluate the model's ability to differentiate between cover as well as stego pictures by standard classification metrics containing recall, precision, and F1-score. The overall presentation of the examined models is compared in Table 1.

Among all the approaches tested, the results show that AMFG-StegNet was the most efficient offer. Specifically, AMFG-StegNet has an accuracy of 0.984, which is significantly better than the results of ZhuNet (0.958), CNN-LSTM (0.932), and GNN (0.967). Due to the combination of spatial, spectral and relational learning of features, the proposed method is now more capable of detecting small steganographic distortions. Similarly, AMFG-StegNet has a score of 0.978 on the Precision rating, which means that the model makes fewer false positive predictions compared to the baseline models. This is very important in steganalysis jobs because false identification of cover images as stego images may lead to unnecessary security alarms. With a recall of 0.980, the suggested framework successfully identifies most of the stego pictures in the dataset. Compared to ZhuNet (0.949) along with CNN-LSTM (0.931), the proposed technique has a higher recall value, indicating that it is better at discovering hidden information in digital material.

Moreover, it is ensured that the proposed model is robust, bearing in mind that the F1-score of this model is 0.975, which demonstrates an excellent accuracy-recall balance. The comparative performances show that the combination of CNN-based spatial analysis, frequency-attention mechanisms, and graph-based relational learning is beneficial in a single framework. The AMFG-StegNet outperforms the other models in all performance parameters, as the graphical display of the assessment parameters in Figure 5 shows.

Table 2. Performance metrics

Model	Accuracy	Precision	Recall	F1-Score
AMFG-StegNet	0.984	0.978	0.98	0.975
ZhuNet	0.958	0.957	0.949	0.95
CNN-LSTM	0.932	0.927	0.931	0.92
GNN	0.967	0.963	0.959	0.962

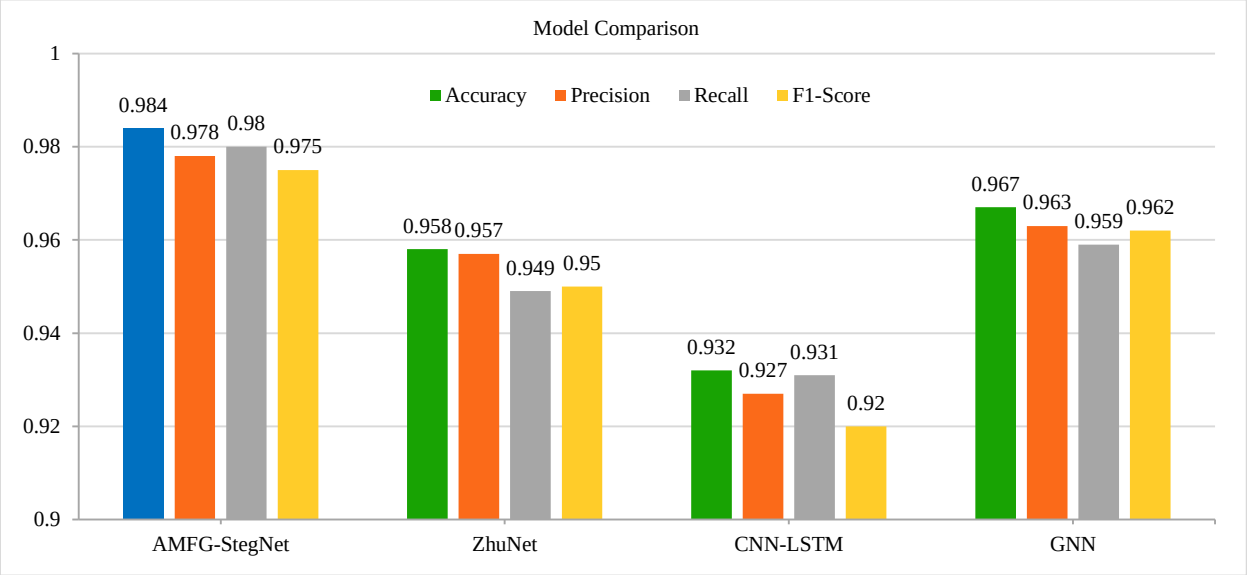

Fig. 5 Comparison of model performance

A confusion matrix analysis was also conducted so as to further explore the classification behaviour of the models being assessed. By breaking down the number of samples into those that were accurately and mistakenly classified, the confusion matrix offers a detailed view of the prediction findings. This analysis enables the ability to test the reliability of steganalysis systems, as it is possible to understand how each model can separate cover and stego images.

Figure 6 indicates that the proposed AMFG-StegNet has a high classification accuracy and minimal misclassification as depicted in the confusion matrix. Misclassifications occur in only 496 samples across all categories, with 30,504 cover photos mistakenly identified as creative, as well as 30,504 stego images as such. These outcomes show that the proposed framework is capable of extracting useful information from digital photos. Using spatial feature extraction, the model is

able to identify tiny steganographic artifacts that are often hard to detect by conventional approaches, as seen by the insignificance of the false positives as well as false negatives, frequency-domain attention, and graph-based relational learning.

Comparatively, ZhuNet has more classification errors in the confusion matrix, as can be observed in Figure 7. Specifically, 1,300 stego images are false projected as cover images, and 1,300 cover images are falsely defined as stego images. ZhuNet can learn spatial characteristics via convolutional processes, however, it is not able to grasp relational relationships between suspicious areas and frequency-domain aberrations. Because of this, the model may not be able to identify certain embedding patterns that the proposed framework is better at seeing.

Among the evaluated methods, the CNN-LSTM model's confusion matrix, shown in Figure 8, has the greatest number of misclassifications. In this instance, 2,100 samples from both classes are improperly categorised, suggesting that the model has trouble differentiating between stego and cover images. The model may be able to establish sequential relationships in the representations of features since it incorporates LSTM layers; however, this approach may not be the most suitable in extracting geographically spread steganographic artefacts of static image data. Consequently, CNN-LSTM model does not do very well in classification.

Figure 9 illustrates the confusion matrix of the model based on GNN and reflects a better result in comparison with ZhuNet and CNN-LSTM. A comparatively reduced error rate is shown by the model's inaccurate classification of 1,000 samples in each category. GNNs' capacity to identify connections between characteristics taken from various areas of the image is responsible for the enhanced performance. Its performance is still below that of the proposed model due to the lack of a specific frequency-domain feature extraction technique, which restricts its capacity to identify minute embedding distortions. The confusion matrix analysis clearly shows that AMFG-StegNet is the best model when compared to the others in terms of accuracy and rate of misclassification. These results prove that steganalysis works better in spatial spectral, and relational feature learning are combined to produce a more complete representation of hidden embedding patterns.

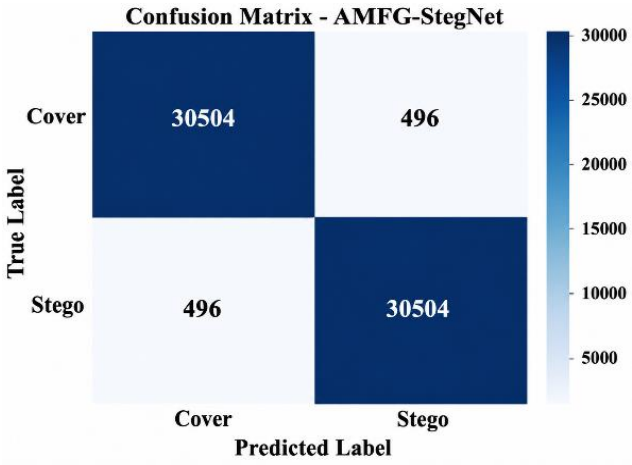

Fig. 6 Confusion matrix of AMFG-StegNet

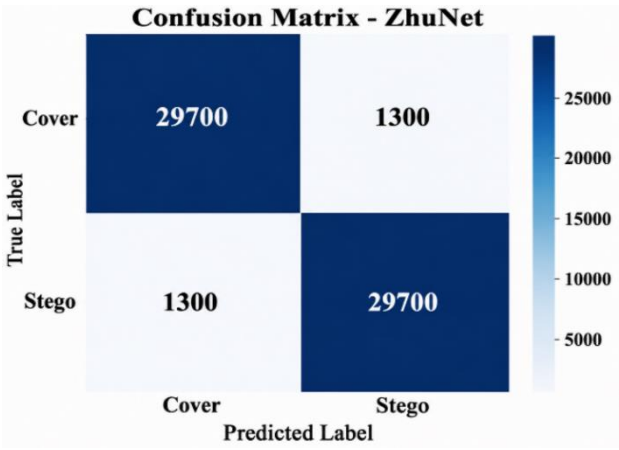

Fig. 7 Confusion matrix of ZhuNet

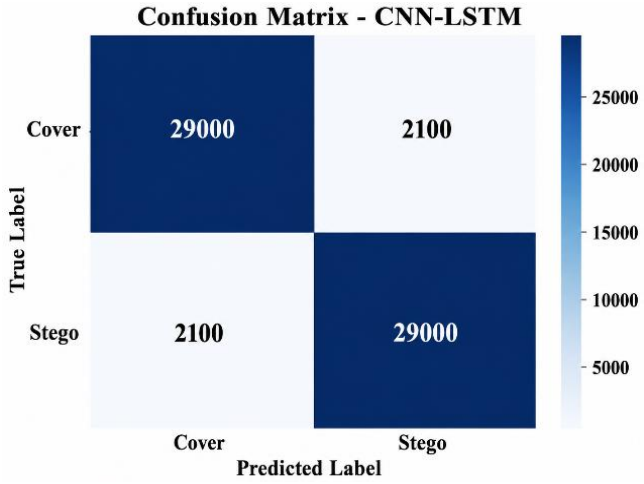

Fig. 8 Confusion matrix of CNN-LSTM

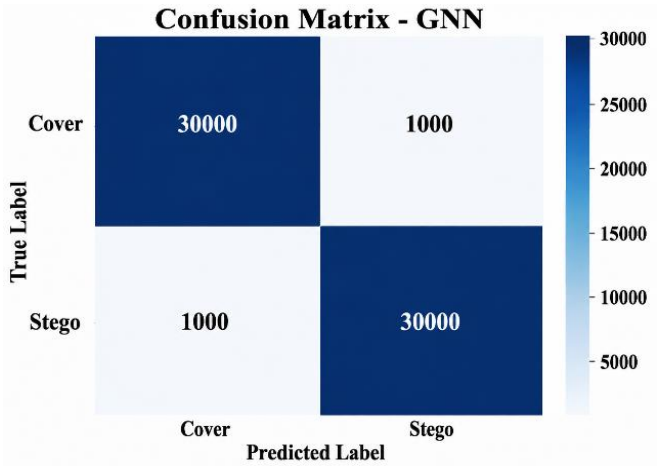

Fig. 9 Confusion matrix of GNN

ROC curves are examined for each evaluated model in order to assess the proposed framework's discriminative power in more detail. Excellent classification performance is shown by the proposed AMFG-StegNet's ROC curve, located quite close to the top left of the ROC area, as seen in Figure 10. With an AUC of 0.984, the model clearly excels at differentiating between stego and cover pictures across a range of decision criteria. Reliable steganalysis requires a high true positive rate along with a modest false positive rate, both of which the suggested model has, as seen by the high AUC value.

In comparison, the ROC curve of ZhuNet (Figure 11) achieves a value of 0.958 in the AUC. As compared to AMFG-StegNet, the curve is slightly less close to the ideal upper-left region, although this is a good sign of good classification performance. This is an indication that ZhuNet can detect delicate steganographic artefacts with relatively better false positive results.

The CNN-LSTM model's lowest reported AUC is 0.9325, as seen in Figure 12, which represents the ROC curve. The curve has a little lower discriminative capacity, as it is more distant to the ideal border. This suggests that complex spatial and spectral distortions brought about by steganographic insertion cannot be effectively modelled in sequential terms only.

Figure 13 shows that the GNN-based model, with an AUC value of 0.967, outperforms ZhuNet and CNN-LSTM. The graph learning technique also allows the model to capture links of suspicious areas, which it can subsequently identify as suspicious. Even the efficacy of the proposed AMFG-StegNet framework is somewhat higher than that of its own.

Overall, AMFG-StegNet has the highest classification performance at different decision levels, and this result is demonstrated by the ROC analysis.

The suggested framework manages to detect many steganographic artefacts and improves steganalysis accuracy more than the existing deep learning models, integrating CNN-based spatial feature extraction, frequency-domain focus approaches, and graph neural interaction layers.

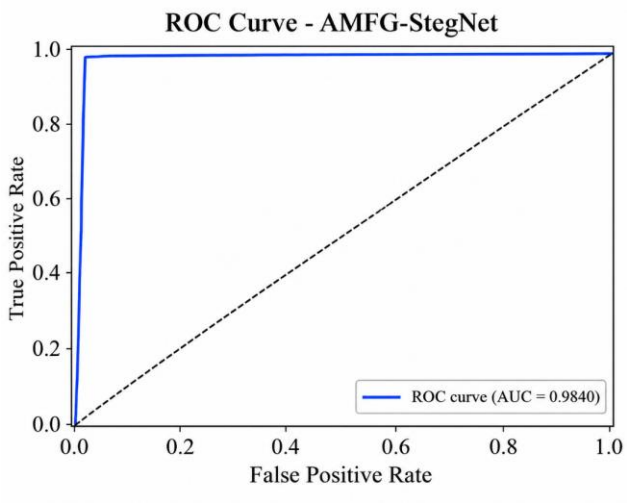

Fig. 10 ROC curve of AMFG-StegNet

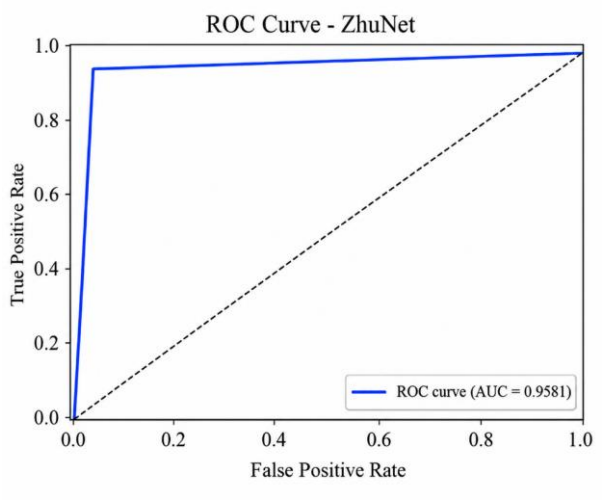

Fig. 11 ROC curve of ZhuNet

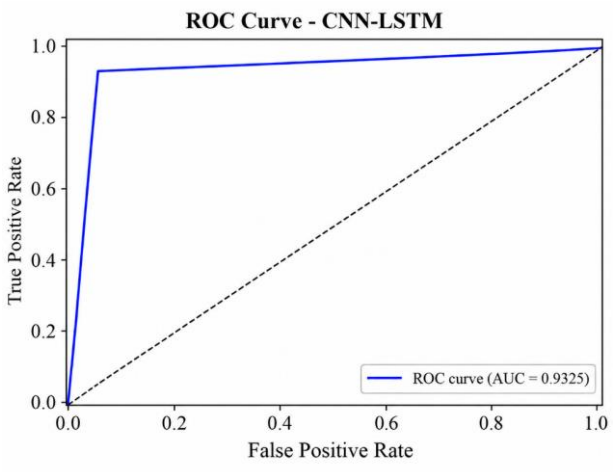

Fig. 12 ROC curve of CNN-LSTM

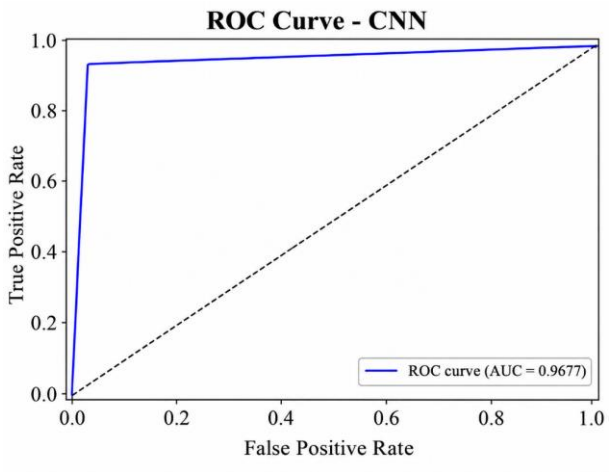

Fig. 13 ROC curve of GNN

6.1. Ablation Study

In this research, the value of the basic components of the suggested AMFG-StegNet architecture is evaluated using an ablation analysis. This study aims to demonstrate the improvements in the collective steganalysis performance by the incorporation of each of these modules and experimentally justify the importance of each of the modules in the framework. The ablation study narrows down to the input of (1) graph neural interaction layers, (2) frequency-domain attention processes and (3) CNN-based spatial feature extraction. The effectiveness of each architectural enhancement can be measured by the gradual introduction of these factors.

Table 3. Ablation study on model architecture components

Model Configuration	Accuracy	Precision	Recall	F1-Score
Baseline CNN (Spatial Feature Extraction)	0.946	0.941	0.944	0.939
CNN + Frequency Attention Module	0.962	0.959	0.957	0.956
CNN + Graph Neural Layer	0.968	0.964	0.961	0.962
Proposed AMFG-StegNet (CNN + Frequency Attention + GNN)	0.984	0.978	0.98	0.975

This ablation study evaluates the enhancement in performance of the individual architectural elements of the proposed framework as exposed in Table 3. The basic Convolutional Neural Network (CNN) model employs convolutional layers to extract geographical characteristics; it achieves a 0.946 precision and a 0.939 F1-Score. Because many steganographic artefacts manifest as minor frequency-domain distortions that cannot be completely captured by spatial characteristics alone, this model's performance is restricted even if it can detect certain spatial abnormalities. A

frequency-domain attention module is included into the CNN backbone to overcome this constraint. The model presentation is improved to a precision of 0.962 and an F1-Score of 0.956 as a result of this improvement, suggesting that spectral feature representation aids in highlighting hidden embedding artefacts created during the steganographic process.

The addition of a GNN interaction layer, which allows the model to record relationships between various image areas, results in further improvement. This setup shows improved detection capabilities via relational feature learning, achieving a precision of 0.968 and an F1-Score of 0.962. With a precision of 0.984, precision of 0.978, recall of 0.980, along with an F1-Score of 0.975, the whole AMFG-StegNet architecture-which integrates CNN spatial features, frequency-domain attention, and graph neural interaction layers-achieves the best results.

These results confirm that the combination of spatial, spectral, as well as relational feature knowledge provides a more comprehensive representation of steganographic artefacts. All in all, the ablation findings demonstrate the mutual additive value of all modules, as the proposed AMFG-StegNet model is more effective because of the use of graph-based interactions and frequency-domain attention, which significantly enhances the performance of the CNN backbone.

6.2. Discussion

Results from experimental evaluations show that the suggested AMFG-StegNet framework outpaces the state-of-the-art deep learning models when it comes to steganalysis. Consistently rising F1-score, precision, as well as recall levels show that the model is quite good at distinguishing between stego and cover photos. A variety of steganographic artifacts may be better captured by combining spatial feature extraction, frequency-domain attention, along with graph-based relational learning, which leads to improvement.

Examination of the confusion matrix, which shows a much lower number of misclassifications associated to the baseline models, further proves the strength of the suggested framework. This shows that the design can help discover unknown information, which is necessary for trustworthy steganalysis systems, and can also decrease false positives and negatives. At various classification thresholds, the suggested model obtains a larger area beneath the curve, as seen in the ROC analysis. This results in a higher level of discrimination. This result demonstrates that the framework can control the false alarm rate while keeping the discovery rate high.

In addition, the ablation research is validated at every architectural piece. Spatial, spectral and relational feature learning are very useful to enhance the detecting power of CNN backbone, which can be observed by the gradual improvement after adding frequency-domain attention and graph neural interaction layers. The results indicate that the suggested AMFG-StegNet system provides a comprehensive and effective approach to finding concealed knowledge in digital images in general.

6.2.1. Comparative Discussion with State-of-the-Art Techniques

The accuracy, precision, recall, confusion-matrix, and ROC results reported above are consistently higher for AMFG-StegNet than for ZhuNet, CNN-LSTM, and the GNN baseline, and the reasons for this gap follow directly from architectural differences rather than from dataset advantage, since all models were trained and evaluated on the same ALASKA2 and Stego Images Dataset splits. First, ZhuNet and comparable CNN-only detectors rely solely on spatial convolution, so embedding traces that are concentrated in high-frequency sub-bands are only partially captured; the adaptive frequency decomposition and frequency-attention branch in AMFG-StegNet is explicitly designed to isolate and weight these spectral artefacts, which explains the reduction in false negatives seen in Figure 6 relative to Figure 7.

Second, CNN-LSTM treats spatial feature sequences as if they were temporally ordered, which is a weak inductive bias for static images and accounts for its comparatively higher misclassification count in Figure 8; AMFG-StegNet instead uses a graph neural interaction layer that models' suspicious regions as nodes connected by learned edges, which is a more natural fit for spatially distributed, non-sequential embedding patterns. Third, the GNN-only baseline captures these relational dependencies but lacks a dedicated frequency-domain pathway, so it still misses spectral distortions that the proposed frequency-attention branch recovers; this is reflected in its intermediate performance, better than ZhuNet and CNN-LSTM but below AMFG-StegNet, in both Table 1 and the ROC comparison in Figure 13.

The ablation study in Table 2 substantiates this explanation quantitatively: accuracy rises from 0.946 for the CNN-only backbone to 0.962 after adding frequency attention and to 0.968 after adding the graph interaction layer in isolation, and reaches 0.984 only when all three components operate jointly. This monotonic, additive improvement indicates that the performance gain over state-of-the-art techniques is not attributable to any single module but to the complementary way in which spatial, spectral, and relational cues compensate for each other's blind spots, which is the central architectural contribution of this work.

7. Conclusion

This paper introduced AMFG-StegNet, a deep learning model that has enhanced the process of recognizing concealed content in digital photos. To effectively detect a range of steganographic artefacts, the proposed architecture will integrate CNN-based spatial feature extraction, frequency-domain attention mechanisms and graph neural interaction layers. By combining the learning of spatial, spectral and

relational features into one framework, the model is able to recognize small embedding abnormalities which are often hard to identify using traditional means. According to the experimental analysis, the suggested framework demonstrates higher performance in a range of performance criteria compared to the existing deep learning frameworks, including ZhuNet, CNN-LMN and GNN-based methods. The findings demonstrate that AMFG-StegNet achieves a superior precision-recall balance and a greater classification accuracy, demonstrating its tremendous capacity to differentiate between cover and stego pictures. Additionally, the ROC and confusion investigations show that the suggested model is strong and selective in finding concealed data at various decision thresholds. Also, the ablation test shows that frequency-domain attention with feature interactions via the graph significantly enhances the CNN backbone in terms of detection, which substantiates the role of each structural aspect in the architecture. The suggested paradigm presents a deeper representation of steganographic patterns, which is evident in these findings. For modern image steganalysis, the proposed AMFG-StegNet provides a dependable and efficient solution. In addition to investigating lightweight architectures for effective deployment in real-world security and forensic applications, future study may concentrate on expanding the framework to cross-domain datasets, adaptive steganography approaches, and real-world multimedia environments.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The author(s) received no specific funding for this research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to express their sincere gratitude to all those who provided valuable support, guidance, and encouragement during the preparation of this manuscript. The authors also acknowledge the providers of the StegoAppDB and ALASKA2 datasets, which were used in this study.

References

[1] Yousef Sanjalawe et al., "A Deep Learning-Driven Multi- Layered Steganographic Approach for Enhanced data Security," *Scientific Reports*, vol. 15, no. 1, pp. 1-30, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mehdi Hussain et al., "Image Steganography in Spatial Domain: A Survey," *Signal Processing: Image Communication*, vol. 65, pp. 46-66, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] Adifa Widyadhani Chanda D'Layla et al., "Steganography in Grayscale Images: Improving the Quality of a Stego Image," *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, Bangalore, India, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Songtao Wu, Shenghua Zhong, and Yan Liu, "Deep Residual Learning for Image Steganalysis," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10437-10453, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Qiang Liu et al., "Coverless Steganography based on Image Retrieval of DenseNet Features and DWT Sequence Mapping," *Knowledge-based Systems*, vol. 192, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Qiang Liu et al., "A Robust Coverless Steganography Scheme using Camouflage Image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 4038-4051, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Mukesh Dalal, and Mamta Juneja, "Steganography and Steganalysis (in Digital Forensics): A Cybersecurity Guide," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5723-5771, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Alfa Fakhrur Rizal Zaini, Ntivuguruzwa Jean De La Croix, and Tohari Ahmad, "A Steganographic Approach based on Pixel Blocks Differencing to Enhance the Quality of the Stego Image," *2024 Conference on Information Communications Technology and Society (ICTAS)*, Durban, South Africa, pp. 63-68, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Moh. Muzayyin Amrulloh, Tohari Ahmad, and Ntivuguruzwa Jean De La Croix, "Analysis of the Smoothing and Payload Distribution Method on Reversible Audio Steganography," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, pp. 1-5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ntivuguruzwa Jean De La Croix et al., "Towards a High-Capacity Data Concealment for Spatial Domain Image-Steganography," *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, India, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ntivuguruzwa Jean De La Croix, Tohari Ahmad, and Fengling Han, "Enhancing Secret Data Detection using Convolutional Neural Networks with Fuzzy Edge Detection," *IEEE Access*, vol. 11, pp. 131001-131016, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Weifeng Sun et al., "An SVM based Secural Image Steganography Algorithm for IoT," *International Symposium on Cyberspace Safety and Security*, Haikou, China, vol. 11982, pp. 357-371, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ivan Castillo Camacho, and Kai Wang, "Convolutional Neural Network Initialization Approaches for Image Manipulation Detection," *Digital Signal Process*, vol. 122, pp. 1-40, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Kai Wang et al., "Efficient Multi-Object Detection for Complexity Spatio-Temporal Scenes," *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, pp. 186-200, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Wei Zhao et al., "TelecomNet: Tag-based Weakly-Supervised Modally Cooperative Hashing Network for Image Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7940-7954, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Nandhini Subramanian et al., "Image Steganography: A Review of the Recent Advances," *IEEE access*, vol. 9, pp. 23409-23423, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Cong Yu et al., "An Improved Steganography Without Embedding based on Attention GAN," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1446-1457, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ntivuguruzwa Jean De La Croix, Tohari Ahmad, and Fengling Han, "Comprehensive Survey on Image Steganalysis using deep Learning," *Array*, vol. 22, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Ying Zou, Ge Zhang, and Leian Liu, "Research on Image Steganography Analysis based on deep Learning," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 266-275, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Yuchen Li et al., "A Deep Learning Driven Feature based Steganalysis Approach," *Intelligent Automation and Soft Computing*, vol. 37, no. 2, pp. 2213-2225, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Hamza Kheddar et al., "Deep Learning for Steganalysis of Diverse data Types: A Review of Methods, Taxonomy, Challenges and Future Directions," *Neurocomputing*, vol. 581, pp. 1-39, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Mohamed A. Elshafey, Ahmed Saleh Amein, and Khaled S. Badran, "Universal Image Steganography Detection using Multimodal deep Learning Framework," *Journal of Information Hiding and Multimedia Signal Process*, vol. 12, no. 3, pp. 152-161, 2021. [Google Scholar]

[23] Jingyuan Jiang, Zichi Wang, and Xinpeng Zhang, "Image-to-Image Steganography based on Multimodal Generative Model," *Signal Processing*, vol. 238, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[24] Miroslav Goljan, Jessica Fridrich, and Taras Holotyak, "New Blind Steganalysis and its Implications," *Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, pp. 1-13, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[25] Hai Su et al., "Steganalysis of Image with Adaptively Parametric Activation," *arXiv preprint arXiv*, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ru Zhang et al., "Depth-Wise Separable Convolutions and Multi-Level Pooling for an Efficient Spatial CNN-based Steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1138-1150, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Mohd Arif Wani, and Bisma Sultan, "Deep Learning based Image Steganography: A Review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, no. 3, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] Anuragh Vijjapu et al., "Steganalysis using Convolutional Neural Networks-Yedroudj Net," *2023 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Donghui Hu et al., "A Novel Image Steganography Method via deep Convolutional Generative Adversarial Networks," *IEEE Access*, vol. 6, pp. 38303-38314, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[30] Eugenia Kuchumova, Sergio Mauricio Martínez-Monterrubio, and Juan A. Recio-Garcia, "STEG-XAI: Explainable Steganalysis in Images using Neural Networks," *Multimedia Tools and Applications*, vol. 83, no. 17, pp. 50601-50618, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Arivazhagan Selvaraj et al., "Digital Image Steganalysis: A Survey on Paradigm Shift from Machine Learning to deep Learning based Techniques," *IET Image Processing*, vol. 15, no. 2, pp. 504-522, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] Tomas Denemark et al., "Selection-Channel-Aware Rich Model for Steganalysis of Digital Images," *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, Atlanta, GA, USA, pp. 48-53, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[33] Guanshuo Xu, Han-Zhou Wu, and Yun-Qing Shi, "Structural Design of Convolutional Neural Networks for Steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708-712, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[34] Ge Luo et al., "Image Steganalysis with Convolutional Vision Transformer," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, Singapore, pp. 3089-3093, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[35] Feng Liu et al., "Image Steganalysis via Diverse Filters and Squeeze-and-Excitation Convolutional Neural Network," *Mathematics*, vol. 9, no. 2, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[36] Tomer Itzhaki, Yassine Yousfi, and Jessica Fridrich, "Data Augmentation for JPEG Steganalysis," *2021 IEEE International Workshop on Information Forensics and Security (WIFS)*, Montpellier, France, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[37] Awab Qasim Karamanji, Asia S. Ahmed, and Ali F. Fadhil, "Comparative Deep Learning Models in Applications of Steganography Detection," *Journal of Image and Graphics*, vol. 12, no. 3, pp. 312-319, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[38] Karen Angulo et al., "Hybrid CNN-NLP Model for Detecting LSB Steganography in Digital Images," *Applied System Innovation*, vol. 8, no. 4, pp. 1-20, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[39] Sachin Mishra, and Dhanashri Patil, "Implementation and Comparative Analysis of CNN, RNN, and GAN Models for Steganalysis in Color Images," *International Journal of Environmental Sciences*, vol. 11, pp. 3312-3336, 2025. [Google Scholar]

[40] Mustapha Hemis et al., "Hierarchical Graph Neural Network for Compressed Speech Steganalysis," *Array*, vol. 28, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[41] Bayu Aditya Triwibowo et al., "Convolutional Neural Network to Detect the Secret Data in the Spatial Domain Images," *International Journal of Safety and Security Engineering*, vol. 14, no. 6, pp. 1707-1718, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[42] Shichen Yang et al., "A Novel Hybrid Network Model for Image Steganalysis," *Journal of Visual Communication and Image Representation*, vol. 103, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[43] Othman A. Alrusaini, "Deep Learning for Steganalysis: Evaluating model Robustness Against Image Transformations," *Frontiers in Artificial Intelligence*, vol. 8, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[44] Doaa Shehab, and Mohmmed Alhaddad, "Image Steganalysis using LSTM Fused Convolutional Neural Networks for Secure Telemedicine," *Frontiers in Medicine*, vol. 12, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[45] Vishnu Suresh et al., "Forecasting Solar PV Output using Convolutional Neural Networks with a Sliding Window Algorithm," *Energies*, vol. 13, no. 3, pp. 1-15, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[46] Karishma Mohiuddin et al., "Retention is All You Need," *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 4752-4758, 2023. [CrossRef] [Google Scholar] [Publisher Link]