International Journal of Safety and Security Engineering
Vol. 15, No. 10, October, 2025, pp. 2135-2142
Journal homepage: http://iieta.org/journals/ijsse

Integrating Wavelet Feature Decomposition and 3D CNNs for Accurate Blind Steganalysis

Natiq M. Abdali[1*], Salah Al-Obaidi[2], Hiba Al-Khafaji[3], Hawraa Al-Janabi[4]

[1] College of Arts, University of Babylon, Babel 51001, Iraq
[2] Department of Computer Science, College of Science for Women, University of Babylon, Babel 51001, Iraq
[3] Department of Software, College of Information Technology, University of Babylon, Babel 51001, Iraq
[4] College of Medicine, University of Warith Alanbiyaa, Karbala 56001, Iraq

Corresponding Author Email: art.natiq.mutashar@uobabylon.edu.iq

https://doi.org/10.18280/ijsse.151016

Received: 17 September 2025
Revised: 18 October 2025
Accepted: 28 October 2025
Available online: 31 October 2025

Keywords:
steganalysis, steganography, convolutional neural network, deep learning, accuracy, dataset

ABSTRACT

Blind steganalysis aims to determine whether a piece of media possesses hidden information without prior knowledge of the embedding algorithm. This task has become increasingly challenging as steganographic techniques continue to evolve rapidly. In this paper, we present a novel approach that integrates wavelet-based feature representations with a three-dimensional deep convolutional neural network (3D CNN) for robust blind image steganography. The discrete wavelet transform (DWT) is employed to capture spatial–frequency characteristics across subbands, enabling the preservation of subtle embedding distortions that conventional spatial-domain approaches often overlook. These wavelet-based feature volumes serve as inputs to the 3D CNN, which jointly models inter-band, spatial, and frequency-domain dependencies through volumetric convolution. To rectify class imbalance and increase classification robustness, we introduce a custom weighted classification layer. We conducted extensive experiments on the BOWS2 and BOSSBase v1.01 datasets, and the results demonstrate that the proposed method outperforms baseline models using 2D CNN architectures in terms of accuracy, precision, recall, and F1-score across all embedding schemes. Our results demonstrate the potential of combining wavelet-domain methods with volumetric deep learning (DL) to improve blind steganalysis in practical digital forensics and cybersecurity applications.

1. INTRODUCTION

The rise of digital communication has made images, audio, and video the most prominent examples of data transmitted via public networks. Digital communication allows for global connections through possibility, but it also creates significant risks for security. Since information could be obscured within digital media by means of steganography, this aspect raises key policy concerns [1-3]. Steganography can create malicious abuse cases, clandestine communications, and violate privacy by embedding hidden data into otherwise benign content. Steganalysis is a counter-strategy to steganography that seeks to discover secret information regardless of which steganographic process has been employed to conceal the information [3-6]. Steganalysis targets to discover underlying patterns and extract hidden information. Perhaps most interestingly, secret analysis is able to do its job without requiring prior knowledge of which specific embedding algorithm was used. Thus, secret analysis is a means for examining information to discover hidden data, regardless of how it was previously structured or encoded. Efforts to uncover hidden information have grown more difficult as advanced embedding algorithms feature subtle, high-capacity, and adaptive methods [7, 8].

Classic approaches to steganalysis based on statistical features such as pixel correlations or histogram derivatives typically do not account for these subtle changes [9, 10]. Recently, deep learning (DL)-based methods that integrated the feature extraction and classification stages into one framework [11, 12]. CNNs have performed particularly well in modeling spatial artifacts introduced during the embedding process [13-16]. However, CNNs have limitations, which include the following:

- Spatially constrained features: 2D CNNs effectively operate in the spatial domain only and ignore characteristics from the frequency domain, which is where embedding artifacts are often more readily identifiable [15, 17, 18].
- Expensive to compute: deep networks have large parameter spaces and require a larger amount of data to train on, as well as more processing resources to run, which may not be practical in crowded, real-world scenarios [12].
- Limited robustness: The models generalize weakly to other payload sizes and embedding algorithms, hence not as trustworthy for forensic or security uses [19-21].

To tackle such limitations, a wavelet-domain three-dimensional deep convolutional neural network (3D CNN) architecture was proposed for blind image steganalysis. To overcome limitations of conventional 2D CNNs, we utilized multi-resolution discrete wavelet transform (DWT) features that co-harness spatial and frequency domain information. By

taking wavelet subbands volumetric inputs into account, the proposed 3D CNN can learn between-band relationships and, therefore, is more capable of revealing hidden steganographic signals. In addition, we have used a weighted classification layer specifically customized for dealing with class imbalance and enhancing detection performance. The key contributions of this paper are:

1. Using wavelet domain priors in DL: We suggest implementing multi-resolution DWT subbands, which serve as structured inputs to identify slight embedding artifacts that spatial CNNs may not achieve.

2. New class of 3D CNN for cross-domain feature learning: Unlike existing 2D CNNs, using volumetric convolutions enables us to capture spatial patterns and correlations between frequency domains, enhancing sensitivity to intricate steganography distortions.

3. Enhanced robustness and accuracy: Through extensive experiments on benchmark datasets, our method consistently outperforms state-of-the-art CNN baselines in terms of accuracy, precision, recall, and F1-score, particularly across diverse embedding algorithms and payloads.

4. Generalizable blind steganalysis framework: The proposed method is independent of the embedding algorithm knowledge, making it adaptable for practical forensic and cybersecurity applications where the embedding strategy is unknown.

The remainder of this paper is organized as follows. Section 2 reviews related work in conventional and DL-based steganalysis. Section 3 describes the proposed methodology in detail. Section 4 presents and discusses the experimental results and performance comparisons. Finally, conclusions based on the reality revealed in this work are drawn in Section 5.

2. RELATED WORK

Due to advances in applying artificial intelligence techniques to address steganalysis in digital images, the performance outcomes have increased the general interest in this topic. Consequently, two categories —conventional learning and DL — have been recognized. In addition, a combination of machine learning and DL has been presented [22]. Conventionally, there are spatial and transform methods based on feature extraction and machine learning. One of these approaches is given by Chhikara and Bansal [23] that utilizes gray-level co-occurrence matrix (GLCM) features extracted from the images, which are then fed into a classifier to differentiate between cover and stego images. Similarly, Gui et al. [24] introduced another hand-crafted approach to compute local binary pattern (LBP) features in conjunction with a linear support vector machine (SVM) to detect hidden data. The histogram correlation derivatives have exhibited far more significant fluctuations. The first derivative of the histogram correlation can be used to identify LSB-steganography, as indicated in the study by Abdali and Hussain [25]. Abdali and Hussain [26] proposed a differential histogram-correlation approach for spatially identifying secret information in images. Differential histogram-correlation analyzes color and grayscale images with varying derivative orders. In some analyses, it is discovered that the first and second derivatives are insufficient; thus, when the ratio of stego to cover images is tiny, the third derivative is required to uncover the concealed information. This technique allows a little secret message to get past the system. Shankar and Azhakath [27] designed a model to extract features from the spatial and frequency domains to distinguish between stego and cover images. Their method concatenates discrete cosine transform (DCT) features with Markov features to create a comprehensive feature vector to train the classifier. They utilized both standard SVM and an enhanced version incorporating particle swarm optimization to SVM (SVM+PSO) to improve detection accuracy. In another emphasis of research, Akram et al. [28] proposed a binary classification-based SVM to determine if an image was either a stego or a cover source image by extracting curvelet histogram features. This proposed model was evaluated on the low- and high-payload steganography images without shifting features as expected, and this model performed decently at identifying stego images, as the additional embedding value changes the dependence on each pixel, but these pixel correlation values would validate the image value changes and thus not determine if a tampering had occurred through a straightforward correlative analysis of a stego image.

DL has become a strong method in steganalysis due to the weaknesses in manual feature engineering [22]. An unsupervised technique [29] with stacked convolutional autoencoders first introduced DL to steganalysis. This early attempt was ultimately not successful, obtaining only 48% accuracy. Supervised learning emerged with Qian et al. [13] proposing a five-layer CNN that had a Gaussian activation function. This type of work was an improvement, but thresholds are still low compared to the other known steganalysis work. Qian et al. [14] improved upon their model from the previous work [13] by using transfer learning - rather than a specially made CNN—indicating that transfer learning is effective and that it can extract and learn better features for image steganalysis, particularly at low payload sizes. Following that work, Xu et al. [15] proposed a CNN architecture that utilized a number of batch normalization layers. Overall, this architecture was much better at distinguishing cover and stego images; thus, it became a critical and formative method for CNN-based models in the previous studies [16, 30, 31].

However, adding many batch layers raised concerns about the model's reduced stability and generalization. To counter this problem, Wu et al. [19] proposed a different architecture for a CNN with the aim of increasing the accuracy of various steganography techniques by utilizing shared normalization instead of batch normalization. The layer utilized a shared set of statistics to normalize input data to solve the issue of network learning instability, which resulted in better detection accuracy and generalization. Boroumand et al. [17] presented a DL model of an end-to-end steganalysis detection network. The detection model has a deep residual architecture that minimizes heuristics and external constraints, thus achieving state-of-the-art detection accuracy through both spatial-domain and JPEG steganography tasks. Following in this pattern, Zhang et al. [32] proposed a compact CNN framework intended for spatial-domain steganography. The framework of this model utilized a low-dimensional kernel as a method of reducing the number of parameters while still preserving local feature representation. In addition, the model implemented separable convolutions and spatial pyramid pooling to receive a better detection performance. Yedroudj-Net [21] builds on the core ideas behind Xu-Net and Ye-Net while improving the architecture of Yedroudj-Net by implementing SRM, TLU activation, and batch normalization. It also integrates adaptive

filter banks and uses data augmentation, both of which provide substantial improvements in steganography detection performance. Overall, despite some advancement with Yedroudj-Net, there is still complexity that could cause overfitting, especially when little data is available for training. Therefore, from a generality perspective, one would need to be careful about hyperparameter tuning. ZhuNet employs two separable convolutional layers, which are also inspired by precedent designs noted in recent work [32]. Later, in 2021, GBRAS-Net [33], a significant advancement, followed Zhu-Net as inspiration for a new direction, which combined depthwise separable convolutional layers with "skip connections" and leveraged SRM filter bank preprocessing in nontrainable mode. GBRAS-Net successfully built upon earlier methods, outperforming the model's performance on benchmark datasets BOWS and BOSSBase v1.01. Lin et al. [18] devised a multifrequency residual convolutional neural network (MRF-CNN) aimed at detecting color image steganography by leveraging multiscale analysis technology to extract steganographic noise from distinct frequency components, resulting in a more efficient and lightweight multifrequency model architecture. Attention mechanisms were added for additional architectural improvements, enhancing the ability of the CNN to focus on slight marks of concealed data to enhance the model's detection accuracy and recognition efficiency [34]. Recently, a new architecture called HSDetect-Net [20] presented a fuzzy logic layer as part of the CNN structure, which enhanced the network's ability to accurately detect concealed information. In conjunction with small convolutional kernels, HSDetect-Net was lightweight. Still, the presented architecture does not always provide the best accuracy for every steganography situation.

3. METHODOLOGY

The suggested framework for blind image steganalysis is made up of four main parts: preprocessing, multi-wavelet feature extraction, 3D CNN feature learning, and classification, as illustrated in Figure 1.

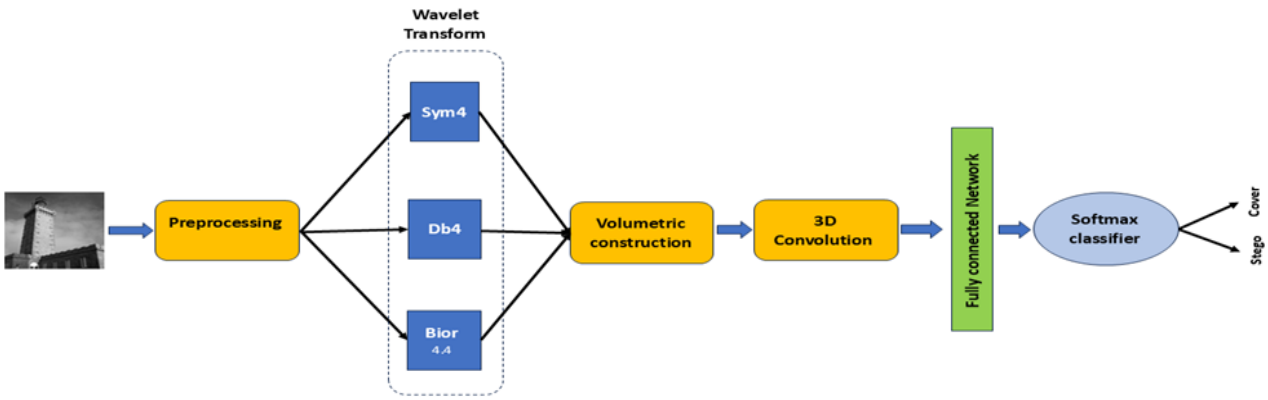

Figure 1. The block diagram of the proposed model

3.1 Preprocessing

The images that are part of the dataset, either cover or stego images, are initially standardized. The images are resized and normalized for consistency and to minimize complicated calculations prior to feature extraction, which sets the data for wavelet decomposition.

3.2 Multi-wavelet feature extraction

The process of extracting wavelet features uses the DWT to capture the spatial and frequency-domain information. DWT tackles the problem of non-stationary signal decomposition by using wavelets—functions confined in both time and frequency—produced by scaling and translating a mother wavelet. DWT decomposes each image into multiple subbands: Both low-frequency approximation and high-frequency detail components.

Given an input image $I \in R^{H\times W}$, denote an input grayscale image, the DWT decomposes it into approximation and detail subbands at level l using a chosen wavelet basis ψ:

$$\{LL^{(l)}, LH^{(l)}, HL^{(l)}, HH^{(l)}\} = DWT^{(l)}(I, \psi), \quad (1)$$

where, $LL_{\psi}^{(l)}$ is the low-frequency approximation and $LH_{\psi}^{(l)}$, $HL_{\psi}^{(l)}$, and $HH_{\psi}^{(l)}$ are the high-frequency detail components.

To capture complementary information, we apply DWT using multiple wavelet families $\Psi = \{\psi_1, \psi_2, \ldots, \psi_F\}$ across L decomposition levels. The resulting volumetric representation V is obtained by stacking only low-frequency subbands across a specific level, l, and wavelet families:

$$V = stack(\{LL_{\psi}^{(l)}\}\psi \in \Psi, l = 1, \ldots, L), \quad (2)$$

Depending on only low-frequency to form the volumetric representation since this band carries global statistical structure and salient image content that machine-learning models use for classification [35, 36]. In addition, CNN can extract stable and discriminative features more effectively from low-frequency coefficients for detecting hidden data [37].

The stacked volumetric representation using Eq. (2) serves as input to the 3D CNN, enabling volumetric convolutions to jointly model both spatial and frequency-domain dependencies. Three distinct wavelet families are used — Symlets, Daubechies, and Biorthogonal — each providing complementary representations of the same image. Incorporating multiple wavelets improves robustness across diverse embedding algorithms by capturing subtle directional edges or noise residues missed by a single wavelet. Furthermore, employing three wavelets enhances the model's ability to generalize across different steganographic techniques.

The resulting subband coefficients are stacked into a volumetric feature construction to form a three-dimensional representation, which preserves multi-resolution dependencies. This volumetric input preserves crucial inter-band and intra-band dependencies, enabling the detection of subtle embedding traces that purely spatial-domain models may overlook. In addition, this structured input is particularly suited for 3D CNN processing, as it allows the network to jointly analyze correlations across both spatial and frequency domains.

3.3 3D convolutional neural network

In this paper, we propose a volumetric-based 3D CNN model for blind steganalysis. The volumetric DWT features are fed into a 3D CNN designed to exploit cross-band relationships. Each volume contains information to represent spatial and frequency information relating to the cover and stego images.

The first stage of this architecture is a three-dimensional convolutional layer, with a kernel of size 5×5×5, with 16 filters, followed by a Rectified Linear Unit (ReLU) non-linearity. This stage captures the information of low-level spatial-frequency features while maintaining small embedding features. Subsequently, 3D max pooling (with a stride of 2×2×2) reduces the size of the 3D feature maps, enhancing the model's robustness to noise. The next convolutional layer block is another 5×5×5 convolutional layer with 32 filters, followed again by a ReLU non-linearity, and a 3D pooling layer block. Chosen smaller kernels are inadequate to capture the correlation that occurs within the local neighbourhoods of the wavelet subband, whereas larger kernels to a create unnecessary complications for the computation and offer no improvements to performance. Therefore, selecting a medium-sized kernel such as 5×5×5 successfully identifies steganographic embedding artefacts without increasing costly computation. Filters were increased from 16 to 32 progressively through the different layers, also by following typical DL techniques that allow for hierarchical feature abstraction while addressing concerns about reducing overfitting on the limited size of the steganalysis datasets.

Once the feature extraction is done, the volumes are flattened into a 1D vector and passed into fully connected layers. These linked layers do the high-level reasoning by mixing the extracted features into a discriminative representation for classification. Dropout regularization can be used at this point to mitigate overfitting. Finally, the softmax classification layer provides the output probabilities that identify the cover or stego images. Through this design, the 3D CNN effectively learns to discriminate between cover and stego images by exploiting subtle embedding cues that span both spatial and frequency domains.

4. EXPERIMENT RESULTS AND DISCUSSION

To thoroughly assess performance, the system is examined across three widely used benchmark datasets with a variety of embedding algorithms and payload sizes. The typical metrics are used: accuracy, precision, recall (sensitivity), and F1-score. The accuracy rate is calculated as follows.

$$Accuaracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

In this context, TP and TN represent the correct classifications of the stego image and the cover image, respectively, while FP and FN denote incorrect classifications (misclassifications). By using these metrics, the evaluation will reveal not only the correctness of the system overall but also the system's robustness for detecting steganographic images of different embedding difficulties.

4.1 Datasets and experimental setup

The evaluation of the proposed methodology was conducted using two datasets: BOWS2 [38] and BOSSBase v1.01 [39]. Each of these datasets comprises 10,000 grayscale images. The dimensions of each image within these datasets are 512×512 pixels. Furthermore, the image datasets incorporate images that have undergone processing using specific steganographic algorithms. Specifically, the datasets contain images that have been manipulated using the S-UNIWARD, WOW, MiPOD, HUGO, and HILL steganographic algorithms.

The MATLAB R2020a version was employed to conduct all experiments on a laptop equipped with an Intel Core i7-9750H 2.6 GHz CPU and 16 GB of RAM.

4.2 Evaluation

Several experiments were conducted to evaluate the proposed method. The set of experiments focuses on proving the superiority of the proposed method to deal with various embedding payload and their outperformance compared to the existing work.

In the first experiment, the BOWS2 dataset was used. This dataset contains stego images manipulated using the S-UNIWARD and WOW steganography algorithms. For this medium-sized dataset, the grayscale images are randomly split into 80% training and 20% testing sets to balance sufficient data representation with reduced training time. The accuracy rate is calculated using Eq. (3). The results of accuracy, precision, recall, and F1-score of this experiment are shown in Table 1.

Table 1. The average of evaluation metrics for the proposed schema using the BOWS2 dataset

Model	Accuracy	Precision	Recall	F1-Score
Proposed model	92.1%	90.8%	92.0%	91.3%

The results displayed in Table 1 demonstrate that the proposed method achieves strong and balanced performance across all four evaluation metrics on the BOWS2 dataset. These findings underscore the ability of the model to accurately identify cover and stego images and maintain a reasonable trade-off of false positives and false negatives. The high precision shows that the model produces, on average, relatively few false alarms for stego images, while the high recall signifies that the model can accurately detect the collection of stego content, even when the artifacts of embedding are small. The F1-score, which reflects both precision and recall scoring, supports the credibility of the model evaluations across the various testing conditions. Compared to traditional handcrafted feature approaches, the proposed model has further benefits associated with its volumetric multi-wavelet representation, allowing the 3D CNN to extract richer spatial-frequency interactions. The data

overall suggests that the detection of steganography adheres to quality and robustness based on the framework's performance in real-life blind application, where reliability and minimization of error are of the utmost concern.

The second experiment used the BOSSBase v1.01 database to examine the performance of the proposed wavelet-based 3D CNN framework under multiple embedding ratios. The embedding algorithms used in this dataset are S-UNIWARD, WOW, MiPOD, HUGO, and HILL steganographic algorithms. The grayscale images are from this dataset, which was randomly divided into 80% for training and 20% for testing. The accuracy rates were analyzed using Eq. (3) to determine the overall accuracy rate. The efficiency rates for this experiment are shown in Table 2 and Figure 2, respectively.

Table 2. Accuracy of the various embedding ratios for the proposed technique with the BOSSBase v1.01 dataset

Embedding Ratio (%)	TP	FN	TN	FP	Accuracy
10	905	95	806	194	85.5%
20	912	88	831	169	87.2%
30	923	77	841	159	88.2%
40	936	64	856	144	89.6%
50	941	59	862	138	90.1%
100	952	48	876	134	91.4%

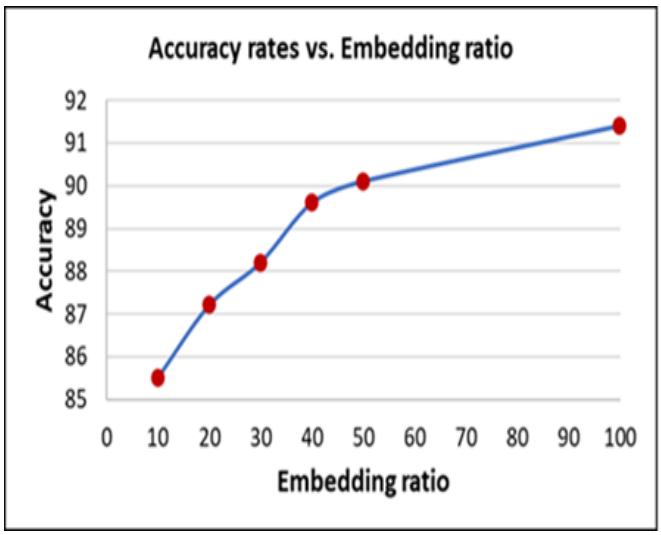

Figure 2. The accuracy values of various embedding ratios

The results in Table 2 and Figure 2 showcase the performance of the new wavelet-based 3D CNN framework at alternate embedding ratios. The results confirmed that the method clearly demonstrated effectiveness in detecting steganographic content even at lower payloads. At an embedding ratio of 10%, the model obtains an accuracy of 85.5%, demonstrating the ability to discern modest distortions, which on many occasions are difficult to determine. However, this accuracy is considered insufficient in information security applications, which shows the shortcomings of the proposed method; it introduces a limitation at very low payload. The performance improved incrementally with increased embedding ratios, as accuracy improved from 87.2% at 20% to 90.1% at 50%. The maximum accuracy rate of 91.4% was achieved at a full embedding ratio (100%), thereby validating the proposed framework under a more robust embedding ratio. This increase in accuracy reflected the model's performance in distinguishing the different strengths of the embedding artifact, since it remained consistent in the detection of each cover and stego class equally.

Furthermore, the low false negative rates of results presented in Table 2 established reliability in the model's capacity to mitigate the missed detection of undetected stego images to perform hidden communication for security purposes; that is of paramount importance for applications from a security perspective. If not, the true positive and true negative rates for each payload were stable during testing, which indicates that the model was not overfitting toward either class, which is typically a challenge of steganalysis tasks in uneven class distributions.

4.3 Comparing the outcomes with the existing techniques

The accuracy of five prominent models of steganalysis, Yedroudj-Net, Zhu-Net, GBRAS-Net, Bayu-Net, and the Proposed Model, is compared in this section, using the two steganography algorithms S-UNIWARD and WOW, as described in Table 3.

This comparative analysis examines the efficiency of these techniques in identifying stego images with varied steganography methods; it was compared to DL techniques. By analyzing the accuracy results, we can determine where the model and algorithm pairs are most effective in steganalysis. A comprehensive study was conducted in terms of five models of steganalysis, Yedroudj-Net, Zhu-Net, GBRAS-Net, Bayu-Net, and the proposed method, based on S-UNIWARD and WOW algorithms with a payload of 0.4 bpp, demonstrating substantial improvements in the evaluation performance measures, especially accuracy.

Table 3. Accuracy comparison for the 3DCNN model and existing techniques using the BOSSBase v1.01 dataset

Model	S-UNIWARD	WOW
Yedroudj et al.'s method [21]	77.2	84.1
Zhang et al.'s method [32]	80.1	84.4
Reinel et al.'s method [33]	81.4	85.9
Triwibowo et al.'s method [40]	83.7	86.4
Proposed 3D CNN	**85.1**	**89.2**

The results presented in Table 3 indicate that the wavelet-based 3D CNN model we introduced has a better detection accuracy than some leading CNN architectures on the BOSSBase v1.01 dataset using the S-UNIWARD and WOW embedding algorithms. Our accuracies of 85.1% (S-UNIWARD) and 89.2% (WOW) surpass those of existing work, and the consistent performance shows that leveraging multi-wavelet feature representations in volumetric 3D convolutions increases performance by capturing spatial–frequency dependencies in a stronger way than a conventional 2D CNN architecture. The fact that these gains are larger when the WOW embedding is used suggests that the proposed model is better able to detect adaptive steganography methods that introduce very localized and subtle distortions.

When comparing the suggested model to Yedroudj-Net, there is a marked improvement, as both WOW (5.1%) and S-UNIWARD methods (7.9%) accuracy increased from the proposed model. Results of the comparison to other models will be discussed below. As discussed, the results show the suggested model has increased detection accuracy and generalization across embedding algorithms and offers a strong approach for real-world applications of blind steganalysis.

Table 4. Comparing the accuracy of our proposed model with SA-CNN and Simulated Dual-CNN using the BOSSBase v1.01 dataset

Hiding Rate (bpp)	CNN Network	Accuracy (%)
0.1	Simulated Dual-CNN [41]	50
	SA-CNN [42]	60
	Proposed model	**85.5**
0.5	Simulated Dual-CNN [41]	66.67
	SA-CNN [42]	76.67
	Proposed model	**90.1**
1.0	Simulated Dual-CNN [41]	80
	SA-CNN [42]	90
	Proposed model	**91.4**

Table 5. Comparison of suggested models, ablation, and DL performance measures for image steganalysis on the BOWS-2 dataset

Model	Accuracy
LEA [43]	76.334%
MLR [44]	77.509%
RL-GAN [45]	79.601%
Proposed method	**92.1%**

Table 4 presents the accuracy for concealing rates of 0.1, 0.5, and 1.0 bpp for both SA-CNN and Simulated Dual-CNN. This table demonstrates that the proposed approach outperforms the SA-CNN and Simulated Dual-CNN. In fact, the suggested technique achieves an accuracy of 85.5%, 90.1%, and 91.4% for concealing rates of 0.1 bpp, 0.5 bpp, and 1.0 bpp, respectively. In contrast, Simulated Dual-CNN obtains 50%, 66.67%, and 80% accuracy for the same concealing rates.

The suggested model was rigorously compared to three other DL models throughout the evaluation phase: LEA [43], MLR [44], and RL-GAN [45]. To evaluate the effect of the accuracy components in our suggested model, we also carried out ablation experiments. Table 5 provides a summary of the outcomes of these studies for the BOWS-2 datasets. The significant performance difference comes from using a multi-wavelet three-dimensional representation of the direct model of high-frequency steganographic artifacts through the proposed method.

5. CONCLUSIONS

This paper has presented a novel framework for blind image steganalysis that integrates wavelet decomposition with 3D CNN. Whereas a 2D CNN approach is limited to the dependencies of spatially constituted pixels, our method leverages DWT to capture spatial–frequency characteristics to enhance the sensitivity to subtle embedding artifacts that conventional spatial-domain approaches often overlook. The ability to use volumetric convolution enables joint learning of inter-band correlations, thereby improving the discriminative power of the steganalysis process.

Experimental evaluation on the BOSSBase v1.01 and BOWS2 data sets demonstrates that the proposed method achieves better detection accuracy compared to the existing 2D CNN-based models. Overall, these findings illustrate the potential efficacy of wavelet-guided volumetric learning technology to distinguish between embedded values or messages in blind steganalysis, especially when the embedding process presents unavoidable perceptible distortions.

Moreover, the proposed framework is shown to contribute a general and data-driven research framework to blindly distinguish between embedding techniques without prior knowledge of the embedding algorithm. The dimension of offering a data-driven research structure ensures the proposed framework is useful in real-world digital forensic and cybersecurity applications where unknown, evolving steganographic techniques create challenges.

Future work will consider integrating hybrid spatial-frequency feature learning, specifically attention-based hybrid feature learning, to enhance feature extraction. Additionally, we will explore practical implementations of the proposed architecture with lightweight implementations. possible in the real-world digital forensics and cybersecurity settings.

REFERENCES

[1] Rana, K., Singh, G., Goyal, P. (2023). SNRCN2: Steganalysis noise residuals-based CNN for source social network identification of digital images. Pattern Recognition Letters, 171: 124-130. https://doi.org/10.1016/j.patrec.2023.05.019

[2] Yin, Z., She, X., Tang, J., Luo, B. (2021). Reversible data hiding in encrypted images based on pixel prediction and multi-MSB planes rearrangement. Signal Processing, 187: 108146. https://doi.org/10.1016/j.sigpro.2021.108146

[3] Su, W., Ni, J., Hu, X., Huang, F. (2022). Towards improving the security of image steganography via minimizing the spatial embedding impact. Digital Signal Processing, 131: 103758. https://doi.org/10.1016/j.dsp.2022.103758

[4] Mahmoud, M.M., Elshoush, H.T. (2022). Enhancing LSB using binary message size encoding for high capacity, transparent and secure audio steganography–An innovative approach. IEEE Access, 10: 29954-29971. https://doi.org/10.1109/ACCESS.2022.3155146

[5] Martini, M. (2023). A simple relationship between SSIM and PSNR for DCT-based compressed images and video: SSIM as content-aware PSNR. In 2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP), Poitiers, France, pp. 1-5. https://doi.org/10.1109/MMSP59012.2023.10337706

[6] Li, N., Qin, J., Xiang, X., Tan, Y. (2023). Robust coverless video steganography based on inter-frame keypoint matching. Journal of Information Security and Applications, 79: 103653. https://doi.org/10.1016/j.jisa.2023.103653

[7] Fu, Z., Chai, X., Tang, Z., He, X., Gan, Z., Cao, G. (2024). Adaptive embedding combining LBE and IBBE for high-capacity reversible data hiding in encrypted images. Signal Processing, 216: 109299. https://doi.org/10.1016/j.sigpro.2023.109299

[8] Al-Obaidi, S., Abdali, N.M., Al-Khafaji, H. (2025). Robust blind watermarking method for high capacity RGB image in wavelet domain. Ingenierie des Systemes d'Information, 30(5): 1209-1218. http://doi.org/10.18280/isi.300509

[9] Fu, T., Chen, L., Fu, Z., Yu, K., Wang, Y. (2022). CCNet: CNN model with channel attention and convolutional pooling mechanism for spatial image

steganalysis. Journal of Visual Communication and Image Representation, 88: 103633. https://doi.org/10.1016/j.jvcir.2022.103633

[10] Gupta, S., Mohan, N., Kaushal, P. (2022). Passive image forensics using universal techniques: A review. Artificial Intelligence Review, 55(3): 1629-1679. https://doi.org/10.1007/s10462-021-10046-8

[11] Tabares-Soto, R., Arteaga-Arteaga, H.B., Mora-Rubio, A., Bravo-Ortíz, M.A., et al. (2021). Strategy to improve the accuracy of convolutional neural network architectures applied to digital image steganalysis in the spatial domain. PeerJ Computer Science, 7: e451. https://doi.org/10.7717/peerj-cs.451

[12] Reinel, T.S., Raul, R.P., Gustavo, I. (2019). Deep learning applied to steganalysis of digital images: A systematic review. IEEE Access, 7: 68970-68990. https://doi.org/10.1109/ACCESS.2019.2918086

[13] Qian, Y., Dong, J., Wang, W., Tan, T. (2015). Deep learning for steganalysis via convolutional neural networks. Media Watermarking, Security, and Forensics 2015, pp. 171-180. https://doi.org/10.1117/12.2083479

[14] Qian, Y., Dong, J., Wang, W., Tan, T. (2016). Learning and transferring representations for image steganalysis using convolutional neural network. In 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp. 2752-2756. http://doi.org/10.1109/ICIP.2016.7532860

[15] Xu, G., Wu, H.Z., Shi, Y.Q. (2016). Structural design of convolutional neural networks for steganalysis. IEEE Signal Processing Letters, 23(5): 708-712. http://doi.org/10.1109/LSP.2016.2548421

[16] Zeng, J., Tan, S., Li, B., Huang, J. (2017). Large-scale JPEG image steganalysis using hybrid deep-learning framework. IEEE Transactions on Information Forensics and Security, 13(5): 1200-1214. http://doi.org/10.1109/TIFS.2017.2779446

[17] Boroumand, M., Chen, M., Fridrich, J. (2018). Deep residual network for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 14(5): 1181-1193. http://doi.org/10.1109/TIFS.2018.2871749

[18] Lin, J., Yang, Y. (2021). Multi-frequency residual convolutional neural network for steganalysis of color images. IEEE Access, 9: 141938-141950. http://doi.org/10.1109/ACCESS.2021.3119664

[19] Wu, S., Zhong, S.H., Liu, Y. (2019). A novel convolutional neural network for image steganalysis with shared normalization. IEEE Transactions on Multimedia, 22(1): 256-270. http://doi.org/10.1109/TMM.2019.2920605

[20] De La Croix, N.J., Ahmad, T., Han, F., Ijtihadie, R.M. (2025). HSDetect-Net: A fuzzy-based deep learning steganalysis framework to detect possible hidden data in digital images. IEEE Access, 13: 43013-43027. http://doi.org/10.1109/ACCESS.2025.3546510

[21] Yedroudj, M., Comby, F., Chaumont, M. (2018). Yedroudj-Net: An efficient CNN for spatial steganalysis. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, pp. 2092-2096. https://doi.org/10.1109/ICASSP.2018.8461438

[22] Ahmed, I.T., Hammad, B.T., Jamil, N. (2022). Image steganalysis based on pretrained convolutional neural networks. In 2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA), Selangor, Malaysia, pp. 283-286. https://doi.org/10.1109/CSPA55076.2022.9782061

[23] Chhikara, R.R., Bansal, D. (2014). GLCM based features for steganalysis. In 2014 5th International Conference-Confluence the Next Generation Information Technology Summit (Confluence), Noida, India, pp. 385-390. https://doi.org/10.1109/CONFLUENCE.2014.6949284

[24] Gui, X., Li, X., Yang, B. (2014). Steganalysis of LSB matching based on local binary patterns. In 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, Japan, pp. 475-480. http://doi.org/10.1109/IIH-MSP.2014.125

[25] Abdali, N.M., Hussain, Z.M. (2020). Reference-free detection of LSB steganography using histogram analysis. In 2020 30th International Telecommunication Networks and Applications Conference (ITNAC) Melbourne, VIC, Australia, pp. 1-7. http://doi.org/10.1109/ITNAC50341.2020.9315037

[26] Abdali, N.M., Hussain, Z.M. (2022). Reference-free differential histogram-correlative detection of steganography: Performance analysis. Indonesian Journal of Electrical Engineering and Computer Science, 25(1): 329-338. http://doi.org/10.11591/ijeecs.v25.i1.pp329-338

[27] Shankar, D.D., Azhakath, A.S. (2021). Minor blind feature based Steganalysis for calibrated JPEG images with cross validation and classification using SVM and SVM-PSO. Multimedia Tools and Applications, 80(3): 4073-4092. https://doi.org/10.1007/s11042-020-09820-7

[28] Akram, A., Khan, I., Rashid, J., Saddique, M., Idrees, M., Ghadi, Y.Y., Algarn, A. (2024). Enhanced Steganalysis for Color Images Using Curvelet Features and Support Vector Machine. Computers, Materials & Continua, 78(1): 1311-1328. http://doi.org/10.32604/cmc.2023.040512

[29] Tan, S., Li, B. (2014). Stacked convolutional auto-encoders for steganalysis of digital images. In Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific, Siem Reap, Cambodia, pp. 1-4. http://doi.org/10.1109/APSIPA.2014.7041565

[30] Tang, W., Tan, S., Li, B., Huang, J. (2017). Automatic steganographic distortion learning using a generative adversarial network. IEEE Signal Processing Letters, 24(10): 1547-1551. http://doi.org/10.1109/LSP.2017.2745572

[31] Chen, M., Sedighi, V., Boroumand, M., Fridrich, J. (2017). JPEG-phase-aware convolutional neural network for steganalysis of JPEG images. In Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, Philadelphia, Pennsylvania, USA, pp. 75-84. http://doi.org/10.1145/3082031.3083248

[32] Zhang, R., Zhu, F., Liu, J., Liu, G. (2019). Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. IEEE Transactions on Information Forensics and Security, 15: 1138-1150. http://doi.org/10.1109/TIFS.2019.2936913

[33] Reinel, T.S., Brayan, A.A.H., Alejandro, B.O.M., Alejandro, M.R., Daniel, A.G., Alejandro, A.G.J., Buenaventura, B.J.A., Simon, O.A., Gustavo, L., Raul, R.P. (2021). GBRAS-Net: A convolutional neural

network architecture for spatial image steganalysis. IEEE Access, 9: 14340-14350. https://doi.org/10.1109/ACCESS.2021.3052494

[34] Li, H., Dong, S. (2024). Image steganalysis algorithm based on deep learning and attention mechanism for computer communication. Journal of Electronic Imaging, 33(1): 013015. http://doi.org/10.1117/1.JEI.33.1.013015

[35] Li, Q., Shen, L., Guo, S., Lai, Z. (2020). Wavelet integrated CNNs for noise-robust image classification. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 7245-7254. http://doi.org/10.1109/CVPR42600.2020.00727

[36] Saini, M., Chhikara, R. (2015). DWT feature based blind image steganalysis using neural network classifier. International Journal of Engineering Research and Technology, 4(4): 776-782. http://doi.org/10.17577/IJERTV4IS040887

[37] Mo, C., Liu, F., Zhu, M., Yan, G., Qi, B., Yang, C. (2023). Image steganalysis based on deep content features clustering. Computers, Materials and Continua, 76(3): 2921-2936. http://doi.org/10.32604/cmc.2023.039540

[38] Bas, P, Furon, T. BOWS-2 Contest (Break Our Watermarking System). Organized between the 17th of July 2007 and the 17th of April 2008. https://data.mendeley.com/datasets/kb3ngxfmjw/1.

[39] Break Our Steganographic System Base Webpage (BOSSBase). http://agents.fel.cvut.cz/boss/.

[40] Triwibowo, B.A., Delenia, E., Putra, Y.H., Croix, N.J. D.L., Ahmad, T. (2024). Convolutional Neural Network to detect the secret data in the spatial domain images. International Journal of Safety & Security Engineering, 14(6): 1707-1718. http://doi.org/10.18280/ijsse.140606

[41] Kim, J., Park, H., Park, J.I. (2020). CNN-based image steganalysis using additional data embedding. Multimedia Tools and Applications, 79(1): 1355-1372. http://doi.org/10.1007/s11042-019-08251-3

[42] Ahmad, M.A., Al-Qhtani, E., Samak, A.H., Ibrahim, A., Elloumi, M., Ahmed, A. (2025). Deep learning-based steganalysis for detection and classification of possible hidden content in images. Fusion: Practice and Applications, 17(2): 377-393. https://doi.org/10.54216/FPA.170228

[43] Liu, S., Zhang, C., Wang, L., Yang, P., Hua, S., Zhang, T. (2023). Image steganalysis of low embedding rate based on the attention mechanism and transfer learning. Electronics, 12(4): 969. http://doi.org/10.3390/electronics12040969

[44] Sun, Y. (2024). Enhancing image steganalysis via integrated reinforcement learning and dilated convolution techniques. Signal, Image and Video Processing, 18(Suppl 1): 1-16. http://doi.org/10.1007/s11760-024-03113-4

[45] Al-Obaidi, S.A.R., Lighvan, M.Z., Asadpour, M. (2024). Enhanced image steganalysis through reinforcement learning and generative adversarial networks. Intelligent Decision Technologies, 18(2): 1077-1100. http://doi.org/10.3233/IDT-240075