Convolution Neural Network with Dual Tree Complex Wavelet Transform Preprocessor for Universal Image Steganalysis

Hanaa Mohsin Ahmed[a] , Halah Hasan Mahmoud[b]

[a] *Department of Computer Science , University of Technology-Iraq , Iraq.Email: 110113@uotechnology.edu.iq*

[b] *Computer Center, University of Baghdad , Iraq.Email:: halah@uob.edu.iq*

A R T I C L E   I N F O

A B S T R A C T

Recently, deep learning models based on convolutional neural networks (CNN) have been used in image steganalysis problems. In this paper, we present different architecture of CNN with dual tree complex wavelet transform for preprocessing before input images put into system. The main task of this transform is for exploiting the difference between cover and stego images through shift variance property. The net consists of five successive convolutions layers. Each one following by normalization and pooling layers ends with fully connected layer. The performance of system is evaluated through accuracy, precision, recall and f-score measures. The results show effectiveness of it with more than 0.9 precision values. HUGO, WOW and UNIWARD algorithms selected for implementation.

1 . Introduction

Revolution of the internet will increase the demand use of data which have available with different type like text, video, image. According to it information hiding has become essential for secure communication [1] has also increased. The steganography can be outlined the process of hiding information in an appropriate cover image and then the output after processed or changing in cover is known as stego-image [2]. The primary key is conserve the distortion in stego image as low as possible during perform embedding of secret image. One of the significant critertion must be considered embedding is the payload capacity. It can be dined as the number of bits that can be hidden in the carrier image and is generally represented in terms of Bits Per Pixel (BPP).

Depending on nature of embedding process the steganography methods can be considered divided into two categories ([1], [2]): Spatial and Transform (or frequency) domain. In First one, Least Significant Bit (LSB) is the

most common one [1]. The bits which constitute the information have directly hidden over the cover image pixel values in this technique. In Second, embedding of message is done after applying a suitable transformation of the image, allowing more bits to be embedded without altering the spatial domain pixel values of the stego-image. Main advantage of this method is that hidden data resides in more robust areas providing a superior resistance over statistical attacks and is unlikely to be decrypted by unintended recipients [2]. Comparing to spatial type, this process is more complicated.

Both areas of image processing [3] and data hiding [2] offers several types of methods for transform. There are numbers of transform which can be considered as the popular like the Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT) [4], Complex Wavelet Transform (CWT) [5] and Dual-Tree Complex Wavelet Transform (DT-CWT) [6]. Although, Wavelet transform are efficient computational algorithm and sparse representation they have suffer from four main problems: Shift Variance, Lack of Directionality, Absence of Phase Information, Oscillation, and Aliasing [7].

- **Oscillation**

  Main purpose of band-pass function is for computation of real valued wavelet transform, coefficients of wavelet have been oscillating in both positive and negative direction around singularity. Thus it is difficult in singularity extraction and signal modeling.

- **Shift Variance:**

  Any small changing in the input signal this may effect a substantive change in the energy distribution of wavelet coefficients on number level of decomposition. Fixed wavelet transform results shift invariance by removing the down-sampling step. On the other side this increases computational complexity and redundancy.

- **Lack of Directionality**

  DWT is not suitable for 2D Natural images. Estimating of lines and edges in it has been fail because it supports only horizontal, vertical and diagonal direction.

- **Aliasing**

  While computation values of wavelet, aliasing result through iterated down-sampling which causes artifacts in the reconstructed image.

2. Related Works

The goal of image Steganalysis is to discover whether an image (cover) contains (or not) a hidden message which produces stego images. This work present design to steganalyzer using machine learning tool called Convolution Neural Network (CNN). Two main important phases constitutes CNN-based steganalyzer the first one is feature extraction and the second is the classification. Today, there are many studies adopted for deep learning especially convolution neural network. Main motivation of this paper is to review different approaches of CNN architectures and present new architecture.   Increasing accuracy and performance and reduce complexity are the main metrics for evaluating works.

After that new model have been presented by [8] to replace two steps of traditional approach. Also adaptive steganography like HUGO, WOW, and S-UNIWARD algorithms are used for testing which result another detection accuracy percentage of 3% to 4% less than to Ensemble Classifier and SRM features. 5 convolutional layers constitutes model with results to 256 features ends with three fully connected layers for classification image. Activation function of type "ReLU" is used in two hidden layers. Images of BOSSbase which are used in implementation are downsized from 512×512 to 256×256. It is drawback still inferior to the SRM.

Later [9] examine framework of Qian's. In scenario of "reusing the same embedding key for different images", performance of detection has been improved more than 16% with S-UNIWARD at 0.4 bpp payload. To get

diversity in this net 64 and 16 filters in first and second layer respectively. C++ are used for implementation, which can be easy downloaded from the Lab of DDE which belong to Binghamton web site.

A new different net is presented by [10]. In this architecture of network implements Truncated Linear Unit (TLU) which is another type of activation function that combined the selection channel inside network. Ten layers and two consequent fully-connected layer end network with label class equal to two constitute model. Toolbox called "Caffe" with some changes with BOSSbase and BOWS2 datasets have been involved in this implementation.

Many Larger filters in convolution layer is a new CNN-based steganalyzer designed by [11], which is suitable when size of image is large lower in value of payloads. Size of input image is it take to system is 512×512 and was beginning by filtered of size 3×3, which trails by a layer composed of 64 filters with zero padding. Through experiments it concludes that his has been defeated many methods of state-of-the-arts especially for the "same embedding key".

On the other hand, final efforts in the field of image steganalysis is multi-columns framework instead of single one get high performance and better precisely regarding to state-of-art- model is presented by [12] Payload "0.1 bpp" and "0.4" bpp" are used for embedding message has been chosen with two adaptive algorithms HUGO, and S-UNIWARD. Detection accuracy is increased by "3%" compared with both model of Pibre et al.

3. Dual-Tree Complex Wavelet Transform (DT-CWT)

The Dual-Tree Complex wavelet Transform is first introduced by Kingsbury in 1998. It is one of the effective approaches for implementing an analytic wavelet transforms [13], [14] and an enhancement to the Discrete Wavelet Transform (DWT) to handle the four issues previously mentioned above. The principle of dual-tree approach is quite simple. It is similar to the idea of positive/negative post-filtering of real sub-band signals.

The dual tree CWT mainly consists of two real DWTs; the first DWT gives the real part of the transform while the imaginary part gives by the second DWTt. Filter Banks (FBs) for both analysis and synthesis that need to implement the dual-tree CWT and its inverse are illustrated in Figures 1 and 2. Perfect Reconstruction (PR) conditions must be satisfy when use two different sets of filters, finally these are jointly designed so that the overall transform is approximately analytic [15].
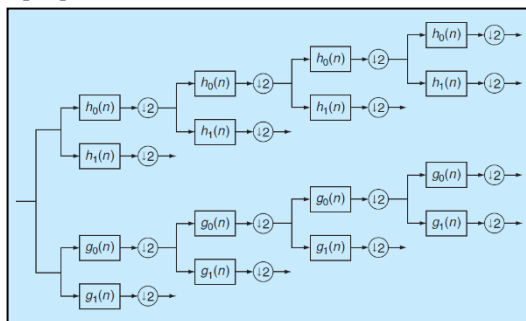

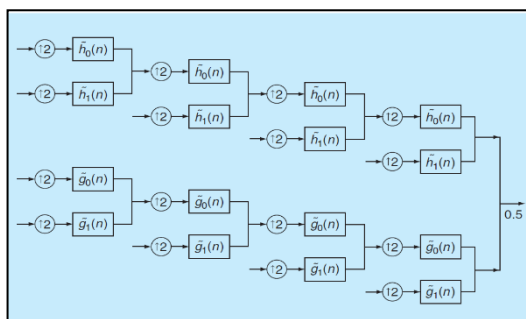Figure (1) Analysis FB for the dual-tree discrete CWT [15].

Figure (2) Synthesis FB for the dual-tree CWT [16].

The h0(n), h1(n) represent the low and high-pass/filter pair for the upper FB
The g0(n), g1(n) state the low and high-pass filter pair for the lower FB.

The two real wavelets associated with it transform denote by h(t) and g(t)
The filters that are designed must satisfy the conditions named PR thus the complex wavelet
$\psi(t) := \psi_h(t) + \psi_g(t)$ is approximately analytic.

The inverse of the dual-tree CWT is as simple as the forward transform. To perform it both the real and the imaginary parts are inverted. The results of inverse are used to obtain two real signals. The final output yield after averaged these signal. If the problem occur the original signal x(n) can be recovered from either the real part or the imaginary part alone;

4. Proposed work

The architecture of Convolution Neural Network has been consists of a set of layers. A number of neurons formed each layer; current and successive neurons are connected together through vector of weights. Each neuron has value represent weighted sum of the neurons from the previous layer. Also, Neurons can compute one type of activation function to introduce non-linearity. During classification phase the following operation performed: take a set the values of the input layer. After that, its values are distributed through number layers of the CNN until reach the final layer (output layer), which outputs the predicted class. A certain level of abstraction is produced at each layer. The "low-level features" have been captured by first layer, like, edges, noise, etc. while the most relevant one also named "high-level features" are captured by deeper layer like shapes related to only one class. Increasing number of layers will get higher in learning capacity of more complex and generic patterns [13]. There are four types of layer in CNN [14]:

- **Input**: This feed image into net.
- **Convolutional**: patch is the area where each neuron is connected within it, in the previous layer. The weights of neurons are shared among all neurons of the same layer, this cause reduce in the size of search space of the learning process.
- **Pooling**: its success the convolutional. Like convolution, each neuron is connected to previous one, and finally calculates the maximum or average of those values.
- **Fully connected**: this is the final layer where each neuron is connected by weights to all neurons of the previous layer.

In this paper present new framework of steganalyzer which consists of three main phases: Image Preprocessing, Feature Extraction and Image classification as shown in Figure (3). Figure illustrates Types of Layers and their parameter sizes (choices) are displayed inside boxes.
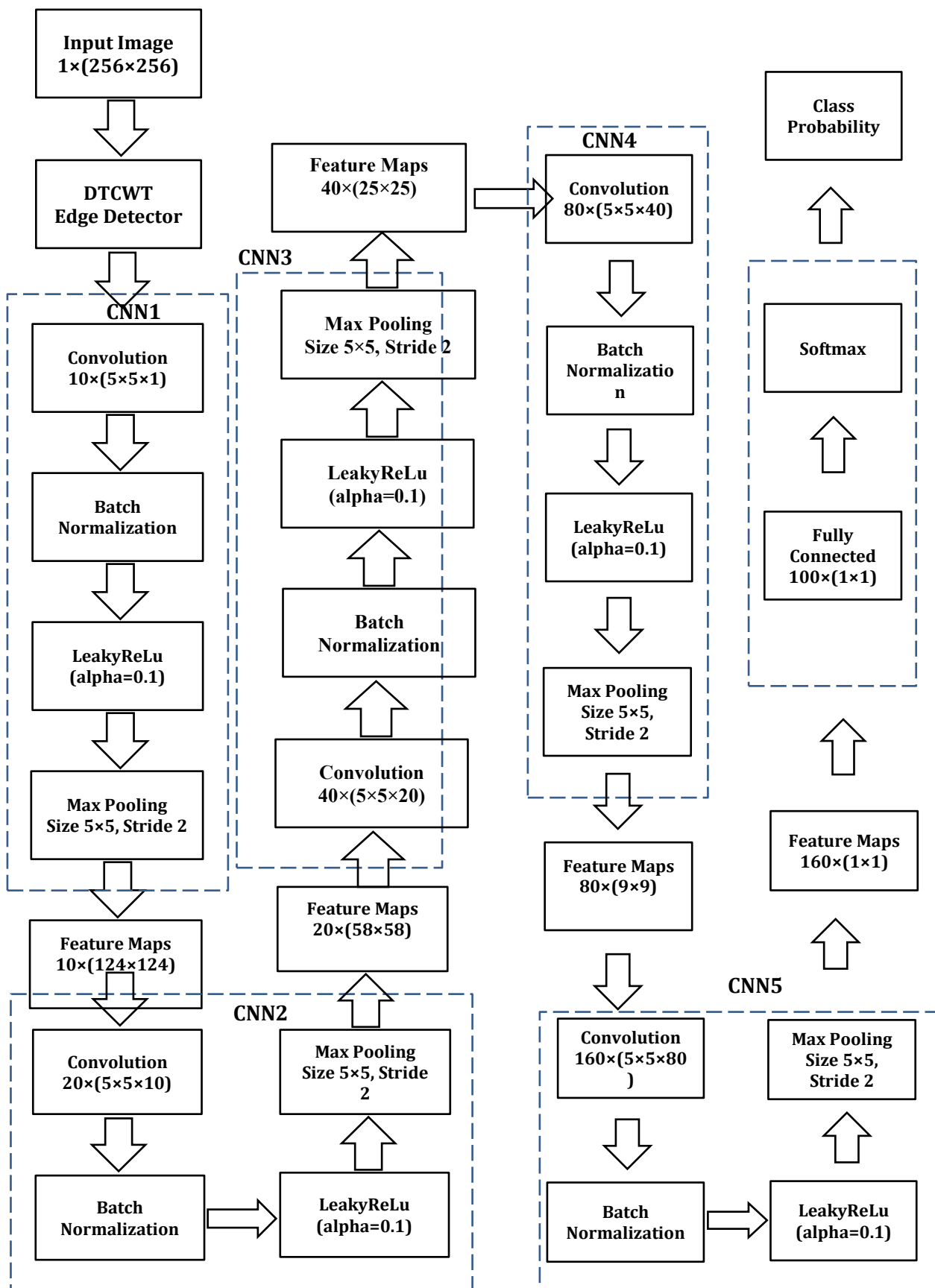
Input Image
1×(256×256)

DTCWT
Edge Detector

CNN1

Convolution
10×(5×5×1)

Batch
Normalization

LeakyReLu
(alpha=0.1)

Max Pooling
Size 5×5, Stride 2

Feature Maps
10×(124×124)

CNN2

Convolution
20×(5×5×10)

Batch
Normalization

LeakyReLu
(alpha=0.1)

Max Pooling
Size 5×5, Stride
2

Feature Maps
20×(58×58)

CNN3

Convolution
40×(5×5×20)

Batch
Normalization

LeakyReLu
(alpha=0.1)

Max Pooling
Size 5×5, Stride 2

Feature Maps
40×(25×25)

CNN4

Convolution
80×(5×5×40)

Batch
Normalizatio
n

LeakyReLu
(alpha=0.1)

Max Pooling
Size 5×5,
Stride 2

Feature Maps
80×(9×9)

CNN5

Convolution
160×(5×5×80
)

Batch
Normalization

LeakyReLu
(alpha=0.1)

Max Pooling
Size 5×5,
Stride 2

Feature Maps
160×(1×1)

Fully
Connected
100×(1×1)

Softmax

Class
Probability

Figure (3): Framework of proposal work

First Phase: Image Preprocessing

The vital part of any automatic learning process is the Data preprocessing [17]. It is used for several cases. Learning process may be damage due to omissions, noise and outliers. In this case, it is emphases on correcting the deficiencies of this damage [18]. Another cases, when simplification and optimization the training of learning system, it focuses on adapting the data. In classical classification techniques, there is minimum need for manually preparing the input due to the high abstraction capacity of CNNs which allows them to work on the original high dimensional space. One of the most critical issues is who select a suitable preprocessing to give quality of results [19]. The Dual-Tree Complex wavelet Transform is one of the most used preprocessing techniques with CNNs.

In preprocessing steps, we aim at maximizing the difference between the stego image and cover image. Generally, the high frequency region includes the most information to recognize stego from cover image and they are expressed as edges of image. Dual-Tree Complex wavelet Transform is used for emphasizing the edge elements of images because it has the property of less shift variance and more directional selectivity. The preprocessing steps include:

1. Resize images into 256 * 256.
2. Applied dual tree complex transform of 5 levels.
3. Remove the coefficients for low frequency and reconstruct images with high pass wavelet transform.
4. Normalize image in order that the elements of its have values in the interval [0, 1] to obtain better learning.

Second Phase: Feature Extraction

Feature learning's module comprised of five blocks of CNN named (CNN1, CNN2, CNN3, CNN4, and CNN5) as shown in Figure (3). Each one start with Convolutional layer for generating feature maps followed by Batch-Normalization layer with Leaky Rectified Linear Unit (ReLU) layer and ends with max-pooling layer except last one.

In each Batch-Normalization layer the operation is perform by subtracting the mean from each pixel of image and then dividing the result by the standard deviation. Then shifts the input by a "learnable offset β" and scales it by a "learnable scale factor γ". Normalization of image is an important step during training phase which ensures that each pixel input into net has a similar data distribution. For image inputs value of pixel numbers must be positive, so we might choose to scale the normalized data in the range [0, 1] or [0, 255]. A key significant is to put these layers between convolutional layers and nonlinearities layers. This makes convergence faster while training the network [20].

This ensure that the gradient back-propagation to put into poor local minima A leaky ReLU layer performs a threshold operation, to check if any input value less than zero is multiplied by a fixed scalar as shown in equation (1).

$$f(x) = \begin{cases} x, & x \geq 0 \\ scale * x, & x < 0 \end{cases}$$

Four main dimensions: (n, k, i, j) which denote to [19] are used to represent data

nimg: nimg-th input image in the mini-batch during training ($1 \leq nimg \leq Nimg$),

kfs: kfs-th feature map ($1 \leq kfs \leq Kfs$)

i_img and *j_img* ($1 \leq i\_img \leq H$, $1 \leq j\_img \leq W$): height and width indices in the feature maps.

As you shown in figure (3), for convolution and feature maps of all CNN sizes of convolution kernel and data are displayed insides boxes. As we can see in the diagram, when passing layers, the size of matrix gets smaller, while the number of channels gets increased.

In convolution layer, **Stride** is an important parameter during convolution operation. We convolve the kernel with the input at every possible spatial location, in this case stride s = 1. If value of stride more than 1 "s > 1", this mean every movement of the kernel skip s-1 pixel locations (note that convolution is performed once every s pixels in both direction i.e. horizontally and vertically). **Padding** is another important concept preferred done. The main purpose of it is to keeping the size of input and output of convolution operations constant and to handle the

borders. In case of pad equal to zero it is named the zero-padding which represent the default. ***Max and average pooling*** are two types of pooling operators which are widely used. For max, the pooling operator maps a sub-region to its maximum value. In case of average, pooling maps a sub-region to its average value [21].

In our model, we set (Stride =1 pad=0) and (stride=2, pad=0), for convolutions and pooling layers respectively for five CNN in net. Max pooling operator is select to get best result through all pooling layers.

Third Phase: Image Classification

In learning system, the classification must have the ability of extracting knowledge from a set of labeled data which can be predict the true label for any given instance. According to it, a there is a capability of good classifier to generalize on new data, i.e., "classify new instances that do not belong to the training set" [21]. Final phase of proposal work includes a fully connect layer which combines flattened elements to calculate the probability of classification using softmax function.

In field of mathematics, the softmax function, or normalized exponential function [22] is a function that takes an input vector of K numbers, and normalizes it into a probability distribution consisting of K probabilities. Before applying softmax, some values of vector may be negative or greater than one; and might not sum to 1. To solve this issue applying softmax, to make sure each component will be in the interval [0, 1]. Often, Softmax is used in convolution neural network especially in fully connected layer to map the non-normalized output of a network to a probability distribution over predicted output classes (cover or stego in this paper). The final numbers of neurons in our work are 100.

5. Evaluation and Testing

Our experiments have been made by selecting three main spatial domain content-adaptive steganographic algorithms: HUGO [23], WOW [24], and S-UNIWARD [25], with embedding rates of 0.1 bpp and 0.4 bpp.

There are well known dataset name BOSSBase used for testing. It consists of no of datasets (cover_jpg.tgz, jpg and cover pgm extension), (stego_hugo_0.1.tgz, stego_hugo_0.4.tgz, pgm format), (stego_uniward_0.1.tgz and stego_uniward_0.4.tgz, type of jpg) and (stego_wow_0.1.tgz and stego_wow_0.4.tgz in pgm format). Each one of previous dataset contains 8,156 images of size 512×512. All three types of content-adaptive method used two payloads 0.1bpp and 0.4 bpp for embedding images. To download these images, use the link (http://info.iut-bm.univ-fcomte.fr/staff/couturie/Code/DL/images/) followed by dataset name.

For each experiment, select 800 pairs of images, 400 pairs were set aside for testing to verify the performance. During training phase not all 500 testing pairs touched in it. We have five optimized CNN models from the 400 training pairs. Each of the CNN was split 300 pairs on trained and 100 pairs on validated. Finally in the testing phase, the 800 testing images (800 testing pairs) went through all the five trained CNNs one by one.

All of the experiments using the CNN were performed using MATLAB 2017b, Deep learning, Statistical and Machine learning, and Parallel Computing toolboxes are chosen for implementing code.

The performance of any classifier can be expressed by scalar values like accuracy, sensitivity, and specificity. Selecting best one depending on the. On other time it is possible to yields different problems like sensitivity to data in case of no balance. Many measures can be easily derived from the confusion matrix for evaluating a diagnostic test is reported in [26]. In this work four measures were choose for assessment net precision, Recall, F-score, and finally Receiver operating characteristics (ROC) which gives good and clear explanation for both basics and generating of the ROC curve. In addition to comprehensive discussions, and metric of Area under curve (AUC) give better explanations [27]. Our testing result shows that the proposed method provides stable high detection accuracy.

5.1 Confusion Matrix (CM).

Performance of model can be evaluated using Confusion matrix. The goal of the classifier is to distinguish each example into its corresponding class which may be one of two cases either a true (malignant) or may be false

(benign) class. The number of possible classifications equal to four rises for each instance: these are as its consequences True Positive (TP), True Negative (TN), False Positive (FP), and finally the False Negative (FN). The confusion matrix forms the basis for several performance measures like accuracy, Recall, F-Score.

5.2. Accuracy

In the field of image steganalysis, accuracy is a more important evaluation method to assess whether a model performs well. Higher accuracy means that system performs best. Equation (5-1) shows how to calculate this measure.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad \ldots\ldots\ldots(5-1)$$

5.3. Precision, Recall and F-Score

In addition to accuracy, the precision, recall and f1-score will also be used for more evaluation of the model. These metrics get better understanding of the results. The number of truly classified positive example divided by the number of its labeled by the system as positive can be defined as the Precision metric. On the other hand the number of correctly classified positive example divided by the number of positive example in the data outline recall. Finally, when perform a combination of two this output F-score. Equations (5-2), (5-3) and (5-4) show how to calculate these metrics.

$$Precision = \frac{TP}{TP + FP} \quad \ldots\ldots\ldots(5-2)$$

$$Recall = \frac{TP}{TP + FN} \quad \ldots\ldots\ldots(5-3)$$

$$F - Score = 2\frac{Precision * Recall}{Precision + Recall} \quad \ldots\ldots\ldots(5-4)$$

5.4. Receiver operating characteristics (ROC)

The primary assessment graph named receiver operating characteristics (ROC) curve can be outline as a two-dimensional in which both the True Positive Rate and False Positive Rate are represent the x-axis and the y – axis respectively. Main types of systems like diagnostic, and machine learning used ROC for evaluation. It presents many (i.e., true positives, and costs, i.e., false positives) [28]. Discrete and continuous are two main outputs from the classification and predict a class decision according to it. [27].

6. Implementation

This program is developed with MATLAB 2017. When we run "main.m", the following interface will appear as shown in Figure (4).
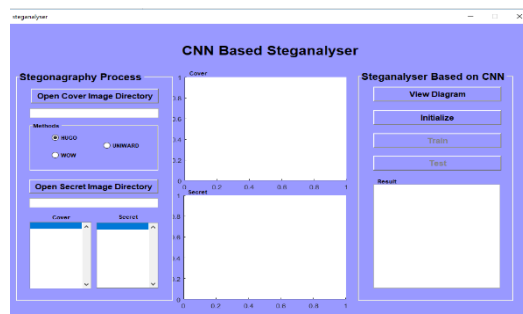
Figure (4) Main Program

For selecting cover image directory, you can press "**Open Cover Image Directory**" and for stego image directory, you can press "**Open Secret Image Directory**". Before choosing stego image directory, you can select

stego method. From the cover and stego directory, it collects images for training and testing. It makes some array with image matrices and label vector for expressing stego and non stego images. See Figure (5-8).
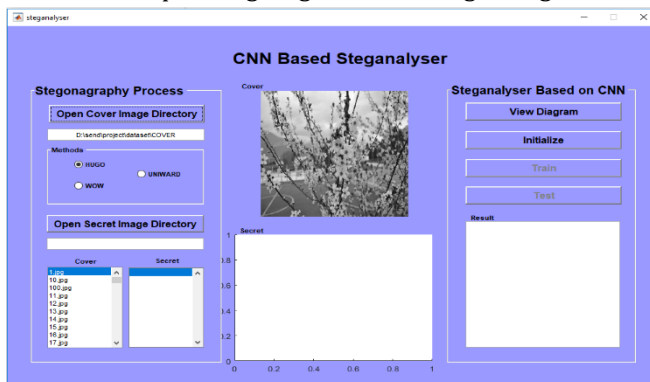

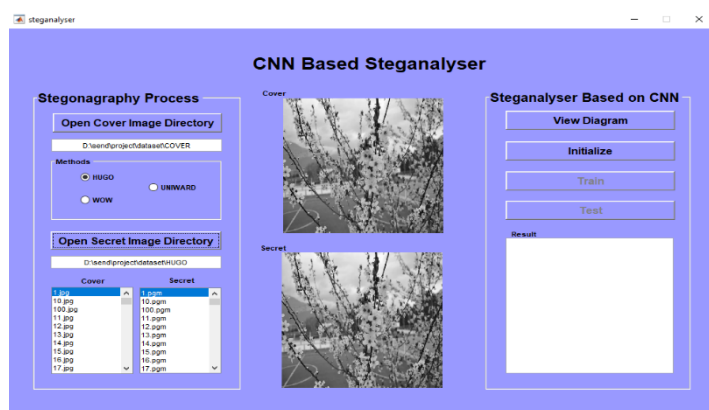Figure (5) Select Cover images
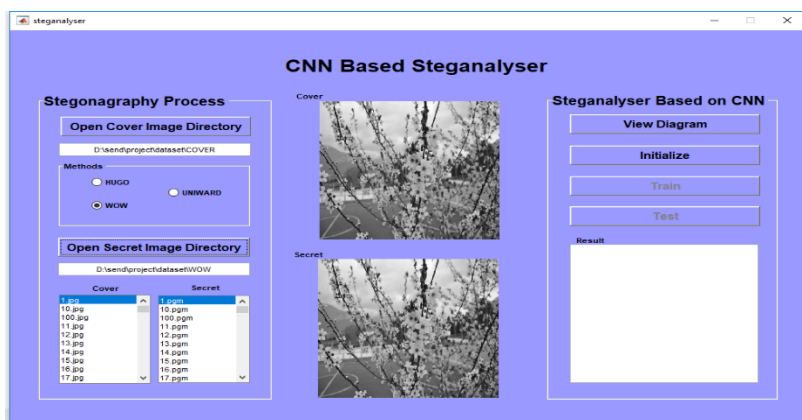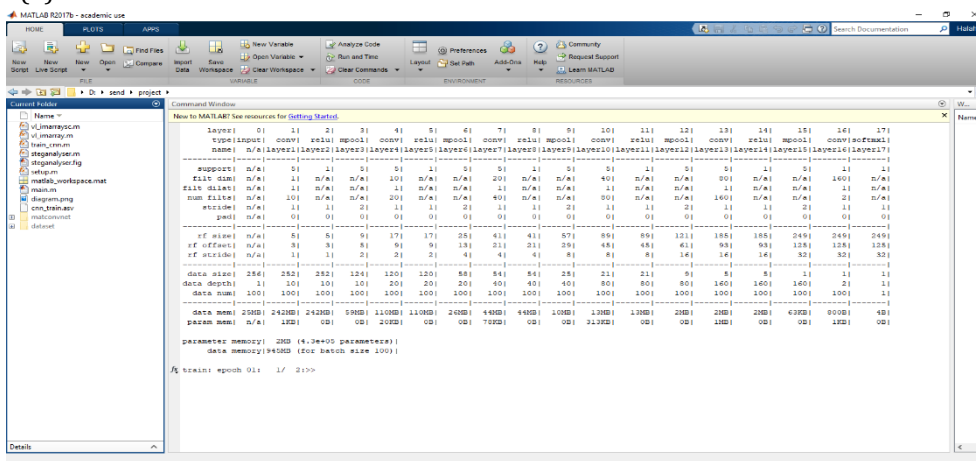

Figure (6) Select HUGO images

Figure (7) Select WOW images

Figure (8) Select S-UNIWARD images

For viewing diagram, you can press "**View Diagram**" button.

"**Initialize**" button is for initializing the CNN structure by adding layers and split data into training and testing as shown in figure (9).

"**Train**" button is for training analyzer and shows the progress with graph. Testing result for trained CNN will appear after pressing "**Test**" button. Table (1) reports precision, recall and f-score for three adaptive algorithms HUGO, WOW, and UNIWARD with two different payloads 0.1 and 0.4.

Table (1): Precision, Recall and F-score Results

	Payload	Precision	Recall	F-Score
HUGO	0.1 bpp	0.98387	0.9833	0.9836
HUGO	0.4 bpp	0.9799	0.97833	0.9736
WOW	0.1 bpp	0.96875	0.96667	0.96771
WOW	0.4 bpp	0.9666	0.96332	0.96441
UNIWARD	0.1 bpp	0.98387	0.98333	0.9836
UNIWARD	0.4 bpp	0.97277	0.97233	0.9711

To evaluate the system select important assessment measurement like accuracy, the receiver operating characteristic (ROC) curves, Confusion Matrix and loss. The Tables number (2, 3, 4, 5), shows these results. Its demonstrate effectiveness of proposal model. In particular, adding Dual-Tree Complex wavelet Transform in pre-processing phase is very important to detect edge element because its high frequency region to split stego from cover image

Table (2): Accuracy

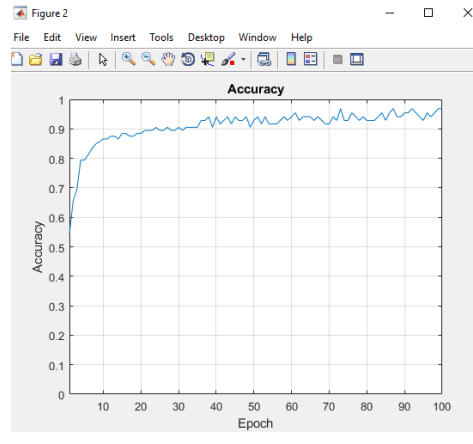a, b : HUGO (0.1 , 0.4 bpp) ,   c, d: WOW (0.1 , 0.4) bpp ,   e, f: UNIWARD (0.1 , 0.4 bpp)
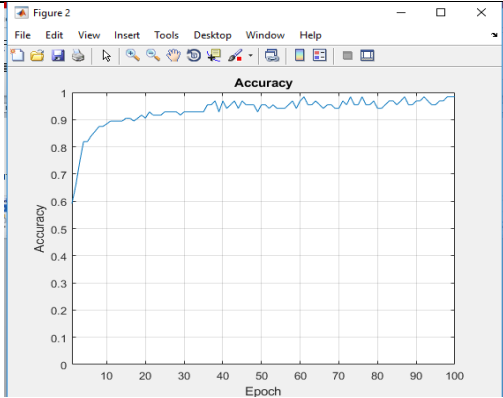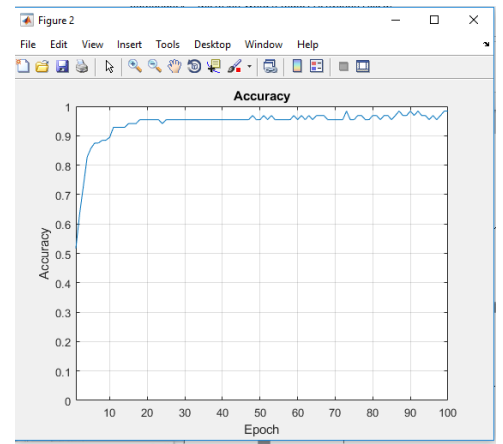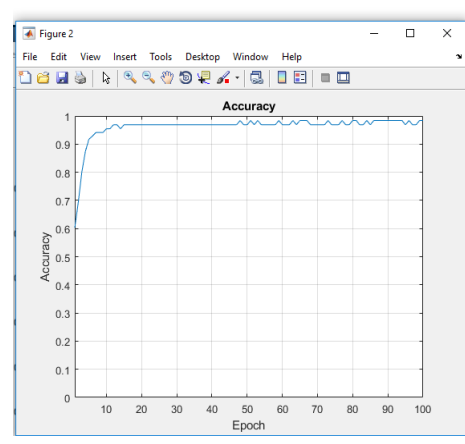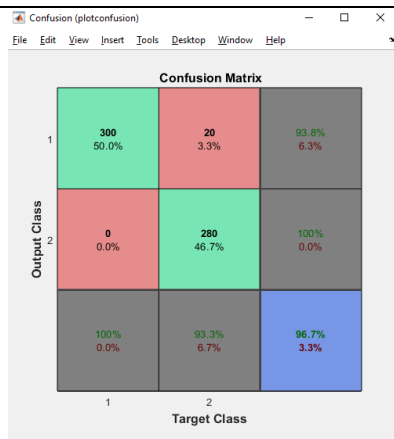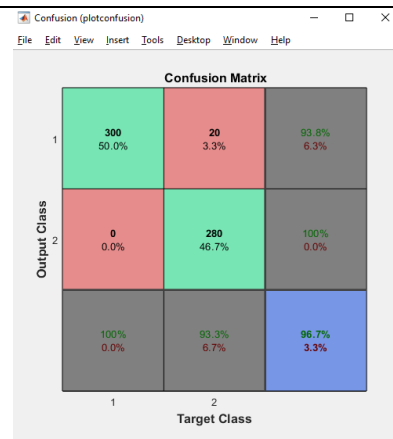

(a)
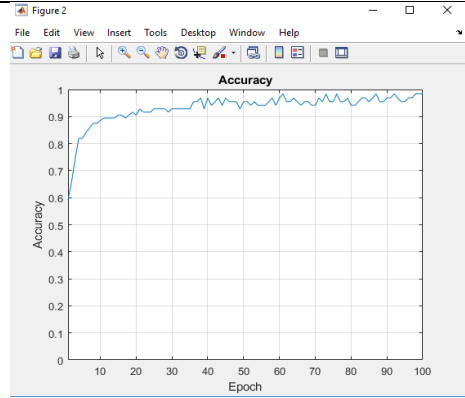
(b)
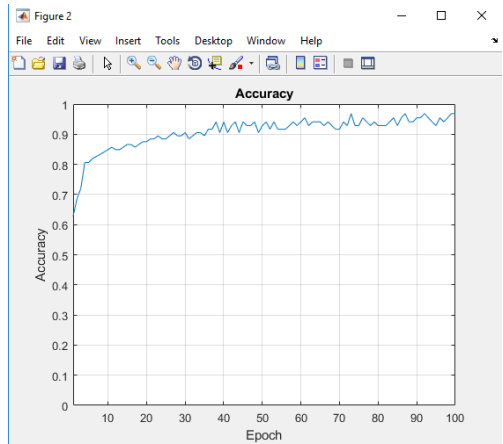
(c)

(d)

(e)

(f)

Table (3): Confusion Matrix

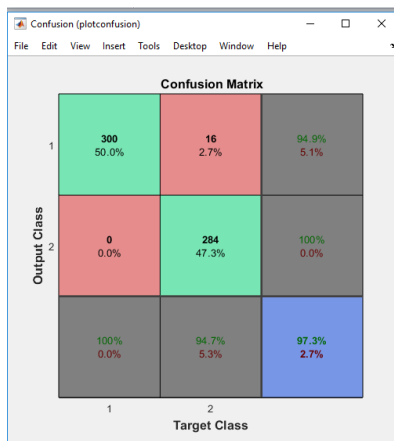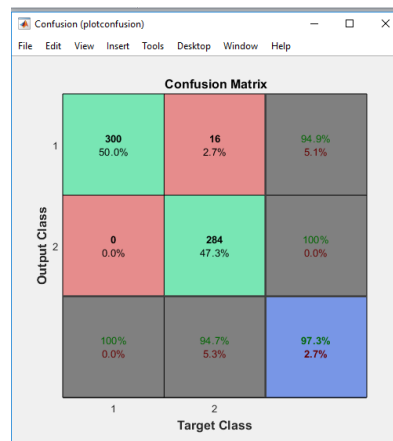a, b : HUGO (0.1 , 0.4 bpp) ,   c, d: WOW (0.1 , 0.4) bpp ,   e, f: UNIWARD (0.1 , 0.4 bpp)
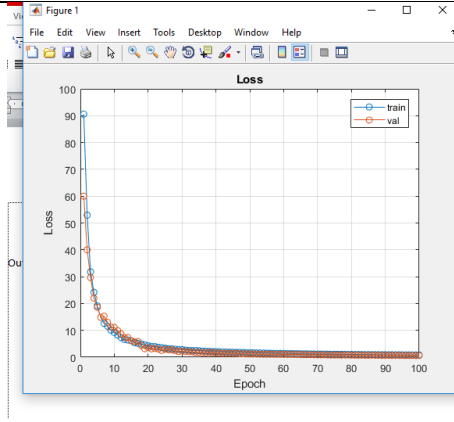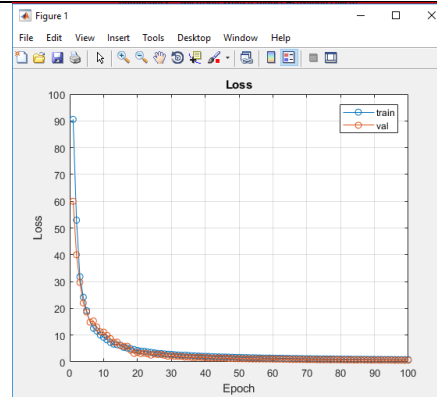

(a)

(b)

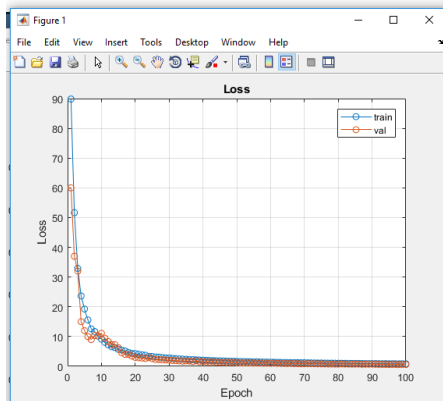(c)

(d)

(e)

(f)

Table (4): Loss

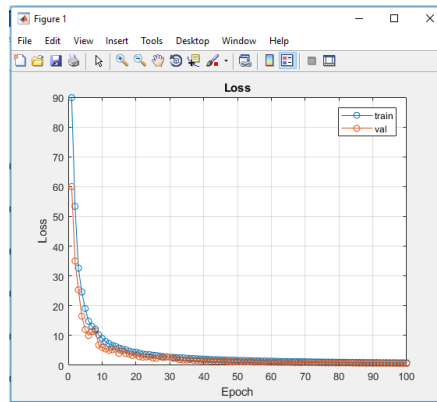a, b : HUGO (0.1 , 0.4 bpp) ,   c, d: WOW (0.1 , 0.4) bpp ,   e, f: UNIWARD (0.1 , 0.4 bpp)
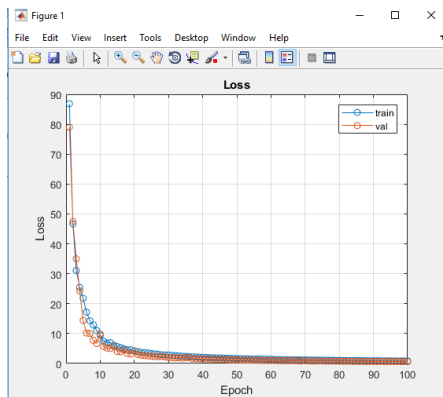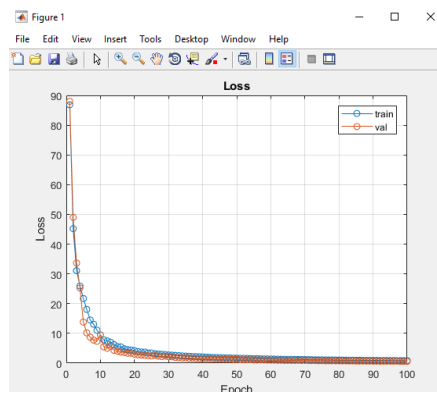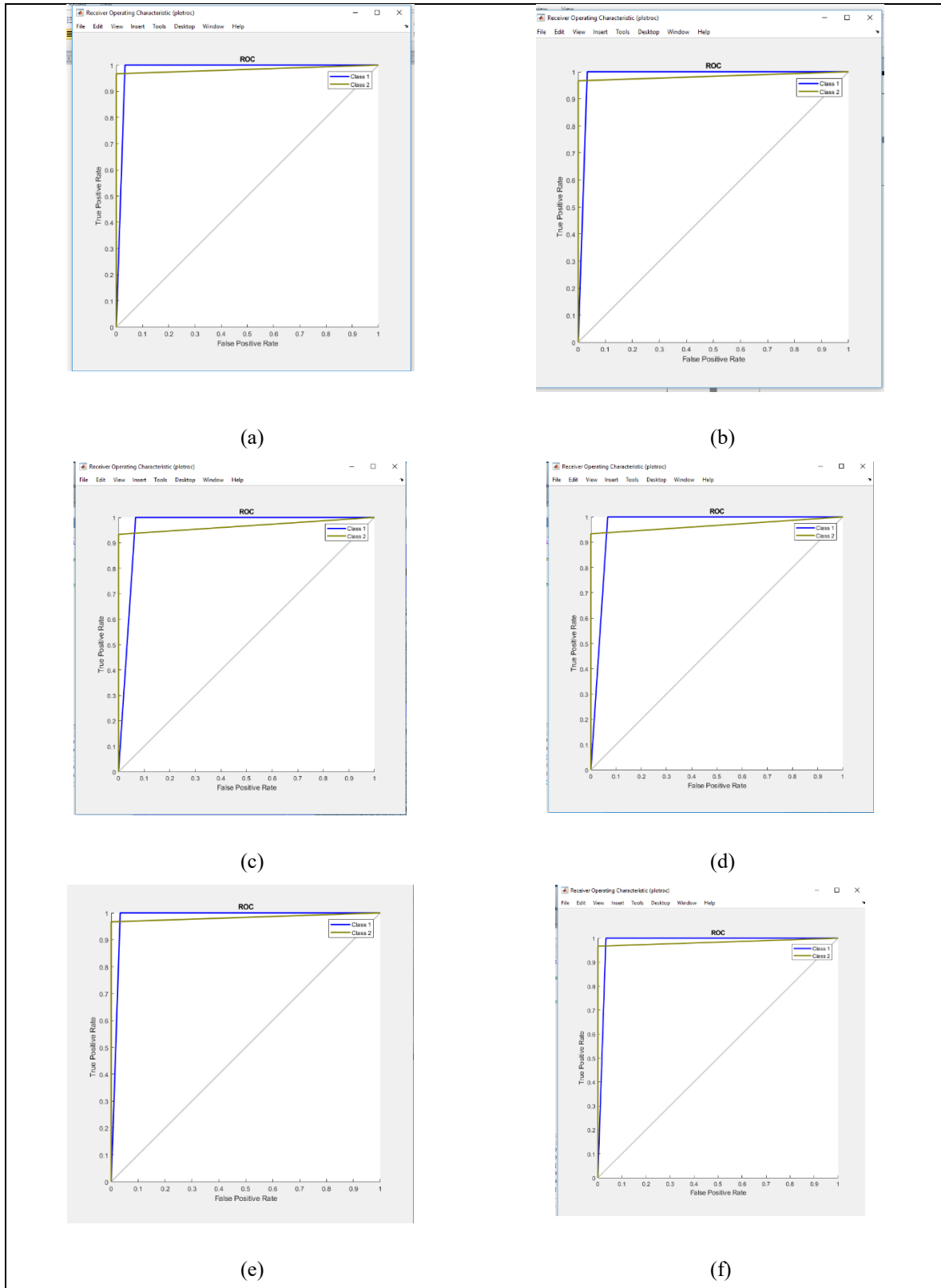

(a)

(b)

(c)

(d)

(e)

(f)

Table (5): Receiver Operating Characteristic (ROC) curves

a, b : HUGO (0.1 , 0.4 bpp) ,   c, d: WOW (0.1 , 0.4) bpp ,   e, f: UNIWARD (0.1 , 0.4 bpp)

(a)

(b)

(c)

(d)

(e)

(f)

7. Conclusion

Today, Main challenge faced field of image streganalysis is detecting steganographic content in its. Deep learning approaches play important tools for classification because of penetration in results. The proposal model successes to de defeat steganagoraphy in JPEG domain in png file format extension. It has able to provide high accuracy for detection with suitable precision more than 0.9 for three adaptive steganography algorithms HUGO, WOW, and UNIWARD with two different payload 0.1 and 0.4 bpp.

References

[1] S. Islam, M.R Modi,. Gupta, P, "Edge-based image steganography". EURASIP J. Inf. Security. vol 1, 8, 2014

[2] S.Singh, T.J Siddiqui, "Transform domain techniques for image steganography". In: Kayem, A., Meinel, C. (eds.) *Information Security in Diverse Computing Environments*, (2014) pp. 245– 259. Hershey, PA, IGI Global, https://doi.org/10.4018/978-1-4666-6158-5.ch013

[3] W. Shen, D. Li, S. Zhang, J. Ou, "Analysis of wave motion in one-dimensional structures through fast-Fourier-transform-based wavelet finite element method", 400, 369–386 (2017)

[4] N. Muhammad, N, Bibi, Z. Mahmood, T. Akram, S.R, Naqvi, "Reversible integer wavelet transform for blind image hiding method". *PLoS ONE* 12(5), (2017).

[5] S.Singh, T.J Siddiqui, "Robust image steganography using complex wavelet transform". In: *International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT),* IEEE, (2013), pp. 56–60.

[6] N.G , Kingsbury., "The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters," in Proc. 8th IEEE DSP *Workshop, Utah,* Aug. 9–12, paper no. 86, (1998).

[7] W. Ivan , R.G, Selesnick, N.G, Kingsbury, "The Dual-Tree Complex Wavelet Transform, A coherent framework for multiscale signal and image processing", *IEEE SIGNAL PROCESSING MAGAZINE* [123], (2005).

[8] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Proceedings of the IS&T/SPIE Electronic Imaging*, p. 94090J, (2015), San Francisco, Calif, USA.

[9] L. Pibre, P. Jérôme, D. Ienco, M. Chaumont , "Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch, *Media Watermarking, Security, and Forensics, IS&T Int. Symp. on Electronic Imaging*, SF, California, USA, (2016), 14-18.

[10] J. Ye, J. Ni, and Y. Yi, "Deep learning hierarchical representations for image steganalysis," I*EEE Transactions on Information Forensics and Security, TIFS*, vol. 12, no. 11, (2017), pp. 2545–2557.

[11] M.Salomon ,R.Couturier, C.Guyeux, J.F.Couchot, J.M.Bahi, "Steganalysis via a convolutional neural network using large convolution filters for embedding process with same stego key", vol. 6, Issue 2, (2017), pp: 79-92.

[12] Q. Ke, L. Dongming, "Image Steganalysis via Multi-Column Convolutional Neural Network", *Guangzhou University*, Feb (2018).

[13] A. Michael A. Nielsen, " Neural Networks and Deep Learning", *Determination Press*, (2015).

[14] L. Yann , B.Leon , B.Yoshua , "Gradient-based learning applied to document recognition". *Proc. IEEE*, 86(11), (1998), 2278– 2324

[15] N.G Kingsbury, "Image processing with complex wavelets," *Philos. Trans. R. Soc. London A, Math. Phys. Sci*., vol. 357, no. 1760, (1999), pp. 2543–2560,

[16] N.G, Kingsbury ,"Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May (2001).

[17] G. Salvador, L. Julian H. Francisco," Data preprocessing in data mining". *Springer*, (2015).

[18] Salvador Et al, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining". *Knowledge-Based Systems*, (2016), 98:1–29.

[19] C.Dan M.Ueli , M. Jonathan , M. Jurgen , "Multi-column deep neural network for traffic sign classification". *Neural Networks*, 32:, (2012), 333– 338.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift, (2015)

[21] H.A. Hamed ,E.J Heravi, "Guide to Convolutional Neural Networks, Practical Application to Traffic-Sign Detection and Classification", *springer*, (2017).

[22] M.B, Christopher , "Pattern Recognition and Machine Learning, Information Science and Statistics", *Springer*, (2006).

[23] T. Pevn´y, T. Filler, and P. Bas, "Using High-Dimensional Image Models to Perform Highly Undetectable Steganography," in *Proceedings of the 12th International Conference on Information Hiding*, IH'2010, Calgary, Alberta, Canada, , vol. 6387 of Lecture Notes in Computer Science, (2010), pp. 161–177, Springer.

[24] V. Holub and J. Fridrich, "Designing Steganographic Distortion Using Directional Filters," *in Proceedings of the IEEE International Workshop on Information Forensics and Security, WIFS'*, Tenerife, Spain,(2012), pp. 234–239.

[25] V. Holub, J. Fridrich, and T. Denemark, "Universal Distortion Function for Steganography in an Arbitrary Domain," *EURASIP Journal on Information Security,* JIS, vol. no. 1, (2014).

[26] S. Shaikh,, "Measures derived from a 2 x 2 table for an accuracy of a diagnostic test" 2 , 128, (2011).

[27] T.Fawcett , "An introduction to roc analysis," *Pattern Recognition. Lett.* 27 (8) , (2006), 861–874

[28] K.H. Zou, "Receiver operating characteristic (roc) literature research, *Online bibliography* available from: http://splweb.bwh.harvard.edu 8000, (2002).