A Steganography Method Using Neural Networks

A. López-Hernández, R. Martínez-González, J. Hernández-Reyes, L. Palacios-Luengas, and R. Vázquez-Medina

***Abstract*—This work proposes a method of steganography without embedding that uses an artificial neural network (ANN) for subliminally communicating some sensitive information using digital images. The proposed method uses the scaled conjugate gradient (SCG) learning to configure an ANN and replicate a secret message from a cover image. The configuration values are sent to destination alongside the cover image, and the secret message is easily recovered. The method was proved using four ANN architectures changing the number of neurons and by using different cover images. This work presents a steganographic method that allows the exchange of sensitive information between two entities. This method has no capacity limitations, it does not produce perceptibility features of the sensitive information in the cover image and it includes an implicit mechanism of integrity/authentication of the cover image. To estimate the performance of the proposed method, the mean square error (MSE) and the peak signal to noise ratio (PSNR) between the recovered secret image and the original secret image, as well as the computation time in the training stage, for artificial neural network architectures with different number of neurons are calculated. Finally, it is shown that the proposed method has better performance when using digital images with large changes in hue. In these cases, the recovered secret image was equal to the original secret image. This opens the possibility of testing the proposed steganographic method to communicate data and not just digital images.**

***Index Terms*—Artificial Neural Networks, Scaled Conjugate Gradient, Steganography application, Steganography Without Embedding.**

Este trabajo fue soportado por el Consejo Nacional de Ciencia y Tecnología (A. A. López-Hernández, CVU-951769), el Tecnológico Nacional de México a través del proyecto 6622.18-P (R. F. Martínez-González), la Universidad Autónoma Metropolitana Unidad Iztapalapa (L. Palacios-Luengas, profesor visitante) y el Instituto Politécnico Nacional a través del proyecto SIP-20196692 (R. Vázquez-Medina).

A.A. Lopez-Hernandez is with Tecnológico Nacional de México, Instituto Tecnológico de Veracruz, M. A. Quevedo 2779, Col. Formando Hogar, 91897 Veracruz, México ((andres_ali_22@hotmail.com).

R.F. Martinez-Gonzalez is with Tecnológico Nacional de México, Instituto Tecnológico de Veracruz, M. A. Quevedo 2779, Col. Formando Hogar, 91897 Veracruz, México (corresponding autor, imag@tecnm.mx).

J.A. Hernandez-Reyes is with Tecnológico Nacional de México, Instituto Tecnológico de Veracruz, M. A. Quevedo 2779, Col. Formando Hogar, 91897 Veracruz, México (jantoniohr@gmail.com).

L. Palacios-Luengas is with Universidad Autónoma Metropolitana, Iztapalapa, Departamento Ingeniería Eléctrica, San Rafael Atlixco 186, 09340 Ciudad de México, Mexico (lpluengas@gmail.com).

R. Vazquez-Medina is with Instituto Politécnico Nacional, Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada, Cerro Blanco 141, Colinas del Cimatario, 76090 Querétaro, Mexico (ruvazquez@ipn.mx).

I. Introducción

CON el acelerado desarrollo de la tecnología de actual, el contenido multimedia como imágenes, audio y video se comparten fácil y rápidamente a través de la Internet. Sin embargo, la seguridad es una preocupación principal al compartir información confidencial a través de una red pública e insegura. Por lo tanto, es esencial disponer de mecanismos que sean robustos y confiables que ayuden a proteger la información confidencial mientras se comunica a través de una red pública [1]. En este sentido, se han propuesto diversas técnicas y estrategias para otorgar seguridad a información confidencial. Estas técnicas y estrategias se basan en tres áreas principales: i) criptografía para otorgar confidencialidad, integridad, autenticidad y no repudio cuando alguna información sensible se comunica entre dos instancias o entidades [2], ii) marcado digital para incluir información del propietario en un medio digital [3] y iii) esteganografía para comunicar información sensible de forma desapercibida [4, 5, 6].

La esteganografía digital se refiere a la comunicación inadvertida de información sensible en un medio portador. Aunque hay muchos medios portadores disponibles, el uso de imágenes digitales es de particular atención. Esto se debe a que la Internet permite que la creación, edición, borrado y distribución de imágenes digitales sea muy fácil y de bajo costo [7], y hace de la comunicación visual un medio de expresión común debido a que es universal, directo y de fácil interpretación. Así, la esteganografía se basa en estrategias para ocultar información sensible en un medio portador, de manera que pase inadvertida [8]; las imágenes digitales a color o en escala de grises son el medio portador más utilizado [9].

Las técnicas de esteganografía en imágenes digitales se clasifican en dos categorías: i) técnicas en el dominio espacial y ii) técnicas en el dominio de una transformada. Las técnicas de dominio espacial modifican directamente la intensidad de la imagen portadora incorporándole los datos sensibles o secretos [11]. Las principales ventajas de estas técnicas incluyen la facilidad de implementación y la velocidad de procesamiento. Sin embargo, estas técnicas son sensibles a ataques de procesamiento de imágenes. Por otro lado, las técnicas en el dominio de una transformada aplican alguna transformación a una imagen portadora para incrustar datos secretos en los coeficientes de esa versión transformada de la imagen portadora [7]. Luego, para obtener los datos secretos se aplica la transformación inversa a los coeficientes modificados de la imagen portadora. Sin embargo, estas técnicas dan como resultado una baja capacidad de ocultamiento y han demostrado

ser computacionalmente complejas [12].

En cualquiera de las dos categorías, las técnicas esteganográficas deben considerar tres características: i) fortaleza para recuperar la información oculta ante alguna modificación intencional o no del medio portador, ii) seguridad, para otorgar imperceptibilidad de la información oculta en el medio portador; y finalmente, iii) capacidad de incrustación de información sensible en un medio portador sin comprometer su estructura.

Por otra parte, las redes neuronales artificiales (ANN: Artificial Neural Nertworks) constituyen una herramienta útil en diversas áreas de la ciencia y la ingeniería. Las ANNs son herramientas de inteligencia artificial que pueden aprender y generalizarse a partir de ejemplos y experiencias para producir soluciones significativas a los problemas concretos [13]. En el aprendizaje automático, las ANNs son modelos computacionales que se pueden utilizar para aproximar funciones que pueden depender de una gran cantidad de datos y que, por lo general, se desconocen. Como modelos eficientes para el reconocimiento de patrones en imágenes [14, 15, 16, 17, 18], las ANNs se han implementado en diversas aplicaciones. Uno de los modelos más comunes y sencillos de utilizar es la ANN prealimentada (*feed-forward*) [19]. Aquí, se pretende usar las ANNs como herramienta en la construcción de una comunicación esteganográfica comunicando una versión aproximada de una imagen secreta obtenida a partir de una imagen portadora.

Así, este trabajo se encuentra organizado de la siguiente manera. En la Sección II se ofrece una descripción de los trabajos que reportan algoritmos basados en ANNs; se consideran dos enfoques, la esteganografía basada en ANNs profundas y la esteganografía sin incrustación. En la Sección III se presenta la principal aportación de este trabajo asumiendo un algoritmo de esteganografía sin incrustación. En la Sección IV se describe el método esteganográfico propuesto. En la Sección V se presentan los resultados obtenidos del método propuesto cuando se aplica a imágenes digitales usando ANNs con diferente cantidad de neuronas en la capa oculta. Finalmente, la Sección VI presenta las conclusiones.

II. Trabajos Relacionados

La mayoría de las técnicas en el dominio espacial se basan en la técnica del bit menos significativo (LSB: Least Significant Bit) [20, 21], la cual sustituye algunos bits de la portadora por los bits del mensaje secreto, utilizando para esto diferentes estrategias. En este sentido, existen técnicas LSB de inserción pseudoaleatoria, las cuales hacen la inserción de los bits del mensaje secreto siguiendo una secuencia producida por un generador de ruido pseudoaleatorio [22]. También, existen técnicas del dominio espacial que buscan mejorar la capacidad o la calidad del método esteganográfico haciendo un procesamiento previo sobre la imagen portadora antes de la inserción LSB. Por ejemplo, existen técnicas que aplican interpolación [23], o técnicas que rotan la imagen de cubierta en diferentes ángulos usando una clave secreta [24]. Finalmente, existen técnicas en el dominio espacial que usan estrategias de inserción bio-inspiradas [11]. Por otro lado, las técnicas en el dominio de alguna transformada emplean por ejemplo la transformada wavelet discreta [25], o la transformada de dispersión (ST: Spread Transform). Ambas técnicas pueden combinarse con métodos de cuantificación [26].

Ahora bien, para darle contexto a este trabajo se debe considerar que existen dos tendencias de diseño de sistemas de comunicación esteganográfica; estas se basan en el tipo de clave esteganográfica que usan. Igual que para el caso de los sistemas criptográficos, existen sistemas esteganográficos simétricos (clave simétrica) y sistemas asimétricos (clave pública). Similar a lo que ocurre en los sistemas criptográficos, el intercambio de claves está orientado a proporcionar servicios de confidencialidad, autenticación y no repudio para los involucrados en la comunicación esteganográfica. Estas claves, junto con el tipo y parámetros del algoritmo, se deben intercambiar entre las entidades involucradas antes de que comience el proceso de comunicación entre ellas; así, este intercambio ocurrirá en una etapa a la que aquí se denomina *Etapa de Negociación Inicial (ENI)*.

Por otro lado, al diseñar un sistema esteganográfico también se deben tener en cuenta tres condiciones: capacidad, seguridad/imperceptibilidad y robustez. Además, se debe establecer entre ellas un equilibrio que dependerá de donde se aplique el algoritmo. Al respecto, según Cox *et al.* [27], los sistemas esteganográficos se pueden clasificar en tres tipos:

1) El objeto portador o de transporte existe y la información secreta no lo modifica.
2) El objeto portador o de transporte existe y la información secreta si lo modifica.
3) El objeto portador o de transporte se crea incluyendo o a partir de la información secreta.

Los algoritmos esteganográficos comúnmente se desarrollan considerando las premisas de los incisos (2) y (3). Existen pocas alternativas esteganográficas desarrolladas bajo la premisa (1).

Con estos antecedentes, se presentan a continuación los principales trabajos que se han reportado considerando dos enfoques: la esteganografía basada en ANNs profundas y la esteganografía sin incrustación basada en ANNs.

A. Esteganografía Basada en ANNs Profundas

Las técnicas de aprendizaje profundo (Deep Learning) aplicadas a la esteganografía (Deep Steganography), comúnmente se enfocan primordialmente a la detección de esteganogramas, más que a su creación [28]. Sin embargo, existen algunas estrategias con criterio *Deep Steganography* que buscan superar a los métodos esteganográficos tradicionales en cuanto a capacidad de inserción, seguridad y robustez contra posibles ataques.

Generalmente, los métodos esteganográficos tradicionales están limitados, ya que los mensajes a ocultar deben ser de menor capacidad comparados con el objeto portador. Este objeto portador permite disponer, sin una degradación notoria, una capacidad límite en la incrustación de la información sensible o secreta. En el área de inteligencia artificial (AI: Artificial Intelligence), las ANNs se han utilizado para proponer soluciones que permitan ocultar información

fundamentalmente bajo la premisa (2) de Cox *et al.* [27], ya que son una herramienta que aprovecha la información en los mensajes secretos y realiza generalizaciones sobre esta información para ofrecer soluciones de problemas específicos [29, 30]. Así, las ANNs ofrecen una variedad de beneficios [31]; específicamente, la esteganografía basada en ANNs profundas tiene como ventaja que oculta información del mismo tamaño que el objeto portador. A diferencia de los métodos tradicionales, las ANNs permiten distribuir los valores del mensaje secreto en todos los bits disponibles del objeto portador [32, 33]. En este sentido, también se han reportado algoritmos basados en ANN de confrontación generativa (GAN: Generative Adversarial Networks), las cuales son arquitecturas de ANN que se componen de dos redes que se confrontan entre sí [34, 35] para encontrar una solución específica.

Considerando las ventajas de las ANN, se han propuesto algunas estrategias como las que se citan a continuación. Por ejemplo, Baluja [32] en 2017 propuso un algoritmo para ocultar imágenes secretas del mismo tamaño que la imagen portadora y, a partir del entrenamiento de ANN profundas, comprimen y distribuyen la representación de la imagen secreta en todos los bits disponibles de la imagen portadora. Wu *et al.* [33] en 2018 emplearon ANN de convolución profunda para ocultar imágenes secretas en imágenes del mismo tamaño con una tasa de codificación del 98.2% y, de acuerdo con los resultados reportados, esta propuesta es robusta ante ataques de análisis estadístico. Husien y Badi [36] propusieron en 2014 un método esteganográfico que emplea el algoritmo Levenberg-Marquardt (LM) para entrenar un neuro-identificador en la ANN; para esto, los usuarios deben proporcionar al sistema el mensaje secreto y el objeto portador (una imagen digital), con lo cual se obtiene una imagen (estego-imagen) resultante que contiene el texto embebido en su interior. Brandão y Jorge [37] describieron en 2016 una técnica para ocultar mensajes confidenciales; esta técnica se basó en la inserción LSB usando ANNs, las cuales además se usaron como clave esteganográfica del sistema para extraer la información previamente embebida. Note que el uso de ANNs como clave incrementa la seguridad del método esteganográfico, pues sin el patrón de la clave es imposible obtener el mensaje embebido. De manera adicional, Brandão y Jorge [37] mostraron que esta característica de funcionamiento puede usar una infinidad de claves debido a que se pueden utilizar diferentes estructuras y composiciones de ANNs. Por otro lado, Chawla y Muttoo [38] propusieron en 2017 una técnica para ocultar información basada en una ANN multicapa que se utiliza para comprimir una imagen, lo cual resulta ventajoso, ya que se pueden utilizar imágenes secretas de mayor tamaño. También en 2017, Alam *et al*. [39] propusieron un método para ocultar información en imágenes digitales en el borde de la imagen portadora utilizando ANNs e inserción pseudoaleatoria basada en transformaciones caóticas. En este caso, las ANNs se utilizaron para identificar los bordes en una imagen y la transformación caótica se utilizó para dispersar los bits del mensaje de manera pseudoaleatoria en los pixeles del borde. Los resultados mostraron un valor de relación señal a ruido pico (PSNR: Peak Signal to Noise Ratio) que indica que la diferencia entre la imagen original y el estego-imagen es muy pequeña [40, 41]. Zhu *et al*. [42], propusieron en 2018 un método esteganográfico para ocultar una imagen en una portadora del mismo tamaño usando ANNs profundas. Lo relevante de esta técnica es que se puede recuperar la imagen a pesar de la presencia de ruido, omisiones de pixeles, recorte, y compresión JPEG. Finalmente, Duan *et al*. [43] propusieron en 2019 una estructura U-Net que considera una forma de entrenamiento emparejado; este procedimiento emplea una ANN para inserción y otra para la extracción; los resultados muestran que el esquema propuesto comprime y distribuye la información de la imagen secreta en todos los bits disponibles de la imagen portadora.

En las estrategias antes mencionadas, los autores concluyeron que, los esquemas propuestos incluyen mecanismos de seguridad y ofrecen una alta capacidad para ocultar los datos sensibles. La mayoría de ellos se basan en técnicas de inserción, las cuales permiten una capacidad limitada que depende del tamaño y de las características de la imagen portadora, sin mencionar los problemas conocidos de seguridad. Por otra parte, las ANNs profundas consideran mejor capacidad de inserción que los algoritmos tradicionales, inclusive se puede ocultar una imagen del mismo tamaño de la imagen portadora.

B. Esteganografía sin Incrustación

Como puede notarse hasta el momento, los algoritmos esteganográficos basados en ANNs comúnmente se desarrollan bajo la premisa (2) de Cox *et al.* [27]. Sin embargo, existe otro tipo de algoritmos que se basa en las premisas (1) y (3) de Cox *et al.* [27]; esto es, son algoritmos que permiten una comunicación esteganográfica sin modificar un objeto de transporte con la información sensible; o bien, lo hacen generando un objeto de transporte incluyendo en el proceso la información sensible. Es importante destacar que este tipo de algoritmos se han desarrollado como una contramedida a dos hechos: i) los algoritmos basados en la premisa (2) de Cox *et al.* [27] dejan trazas de la incrustación de la información sensible en el objeto de transporte y ii) existen algoritmos de esteganálisis muy poderosos basados en ANNs que logran detectar un esteganograma en una comunicación esteganográfica. Así, existen trabajos que en los últimos tres años reportan algoritmos esteganográfico que no modifican la imagen portadora, y se han desarrollado bajo el concepto de esteganografía sin incrustación (SWE: Steganography Without Embedding).

En el caso de los algoritmos que se basan en la premisa (1) de Cox *et al.* [27] la información sensible no modifica el objeto de transporte. Un ejemplo de este tipo de algoritmos es la esteganografía lingüística. Sin embargo, contrariamente a las tendencias actuales, la seguridad de los algoritmos de esteganografía lingüística se basa en la obscuridad; esto es, la seguridad se basa en mantener en secreto el algoritmo de incrustación/extracción. Por ello, se considera que este tipo de algoritmos no son prácticos, además de que no permiten comunicar información sensible cuyo tamaño sea mayor que algunos cientos de bits.

Para el caso de los algoritmos esteganográficos que se basan en la premisa (3) de Cox *et al.* [27], el objeto de transporte se crea automáticamente incluyendo la información secreta. Como ejemplo, existen algoritmos en los que la información secreta se traduce a un vector de ruido, el cual se usa en un módulo llamado generador, el cual se basa en algún tipo de ANN que genera el objeto de transporte. De esta manera, este tipo de algoritmos no requieren realizar operaciones de modificación o incrustación durante el proceso de generación del objeto de transporte; luego, la información sensible se puede extraer a través de otra ANN, a la que comúnmente se le llama extractor.

Así, existen algoritmos del tipo SWE que generan objetos de transporte a partir de una versión procesada de la información sensible, usando ANN del tipo GAN. De acuerdo con Chaumont [28], quien ofreció en 2019 una panorámica de los algoritmos de esteganografía y esteganálisis reportados entre 2015 y 2018, los algoritmos del tipo GAN se basan en la Teoría del Juego. Bajo este concepto cada una de las tres entidades involucradas, en este caso las tres instancias de una comunicación subliminal, intentan encontrar una estrategia que maximice sus ganancias. La solución a este problema, si existe, se ha denominado Equilibrio de Nash. De acuerdo con Chaumont [28], la formalización matemática del problema de la esteganografía/esteganálisis por la Teoría de Juego es difícil y, a menudo, está lejos de la realidad práctica. Una manera menos complicada que se usa para determinar el equilibrio de Nash es simulando el juego. En esta simulación se considera que existen tres agentes, uno por cada una entidad involucrada en la comunicación esteganográfica, los cuales contienden hasta alcanzar el equilibrio de Nash. Una vez alcanzado tal equilibrio, una de las entidades detiene la simulación y conserva su agente como el algoritmo de incrustación y el otro agente como el algoritmo de extracción. Este esquema, que permite alcanzar el equilibrio, ocurre en la *ENI* de la comunicación esteganográfica, ya que es la etapa en la que se definen los algoritmos de incrustación y extracción que habrán de utilizarse. Luego de ello, es posible comenzar la comunicación esteganográfica entre las dos entidades.

Por otro lado, también está el trabajo que propusieron Hu *et al*. [44] en 2018; este trabajo consiste en un método del tipo SWE basado en ANNs profundas del tipo GAN convolucional, donde la información sensible se asigna a un vector de ruido. Este método SWE se usa para entrenar a una ANN para generar la imagen portadora. Este método no requiere operaciones de modificación o incrustación durante el proceso de generación de imágenes, y la información contenida puede extraerse con éxito. Los resultados experimentales de Hu *et al*. [44] muestran la capacidad que tiene su algoritmo para resistir la detección mediante algoritmos de esteganálisis actuales.

Como puede notarse, los algoritmos del tipo SWE deben construir una relación funcional entre la información secreta y la portadora sin realizar la incrustación del mensaje o imagen.

En este contexto, este trabajo presenta una alternativa esteganográfica del tipo SWE que usa imágenes digitales y la técnica de optimización conocida como el gradiente conjugado escalado (SCG: Scaled Conjugate Gradient) [45]. La técnica de optimización SCG se usa para configurar una ANN que busca generar la imagen secreta a partir de la imagen portadora con el menor error posible. A continuación, se describen las contribuciones realizadas.

III. Principal Aportación

La principal contribución de este trabajo es el desarrollo de un método esteganográfico del tipo SWE que permite comunicar imágenes digitales en escala de grises entre dos entidades, sin limitaciones de capacidad de inserción y sin rasgos de perceptibilidad de la información sensible en la imagen portadora. El método propuesto no tiene limitaciones de capacidad de inserción, ya que es del tipo SWE y, por tanto, no hace una incrustación de la información sensible en la imagen portadora, sino que entrena a una ANN para que a partir de la imagen portadora y de la imagen secreta se pueda estimar una versión de la imagen secreta (imagen secreta recuperada), la cual se habrá de comunicar esteganográficamente. También al ser del tipo SWE no tiene limitaciones de imperceptibilidad debido a que la imagen portadora se comunica en su forma original, sin alteraciones, y la información sensible se encuentra implícita en la configuración de la ANN. Así, la ANN toma como condición inicial la imagen portadora para recuperar la versión comunicada esteganográficamente de la imagen secreta.

Para lograr una alta efectividad del método propuesto se asume, como en cualquier algoritmo esteganográfico, que se requiere de una *ENI* de la comunicación. En este caso, esta etapa ocurre en un canal seguro de comunicaciones alterno al canal inseguro por donde ocurrirá la comunicación esteganográfica. En la *ENI* se debe intercambiar la información de operación del algoritmo y sus parámetros funcionales, además de la información requerida para lograr confidencialidad, autenticación y no repudio en la comunicación. Así en esta etapa, las entidades participantes deben intercambiar la clave de sesión (esteganografía simétrica) o las claves públicas (esteganografía asimétrica). Note que la *ENI* es inevitable en cualquier comunicación esteganográfica.

De esta forma, en el esquema propuesto se asume que existen dos canales de comunicación, uno por el que se comunica la imagen portadora, el cual puede ser seguro o no, y el otro, por el que se comunica, en la *ENI*, el tipo y la configuración de la ANN que permitirá obtener la información sensible a partir de la imagen portadora. Adicionalmente, el método propuesto ofrece condiciones de seguridad en la comunicación esteganográfica, dado que permite detectar cuando una imagen portadora se modifica o sustituye.

Así, la alta capacidad de carga, la imperceptibilidad de la información sensible en la imagen de la portadora (algoritmo del tipo SWE) y el mecanismo implícito de integridad/autenticación hacen del método propuesto una alternativa prometedora cuando la configuración de la ANN se comunica por un canal seguro criptográficamente.

Así, este trabajo presenta un método de comunicación esteganográfica del tipo SWE que permite el intercambio de información sensible entre dos entidades.

Con estas precisiones, el método que se propone en este trabajo tiene las siguientes características:

1) Puede utilizarse en estrategias simétricas o asimétricas, y el intercambio de la información de la ANN ocurre durante la Etapa de Negociación Inicial (*ENI*).
2) Es del tipo SWE y tiene en cuenta la premisa (1) de Cox *et al.* [27]; los métodos del tipo SWE, reportados hasta ahora, se basan en la premisa (3) de Cox *et al.* [27].
3) Tiene una funcionalidad que alcanza una eficacia con degradación cero del medio de transporte y perceptibilidad cero de la información secreta; se aprovecha la *ENI* de la comunicación esteganográfica para enviar, además de las claves del sistema, la información de la ANN.
4) Puede aprovechar para seguridad la información de la ANN, ya que puede ser considerada como una clave de sesión (esteganografía simétrica), o como una clave adicional de intervención (esteganografía asimétrica), para la extracción de la información secreta; sin esta información, la extracción de la información sensible no podrá ocurrir.
5) Es general, dado que las condiciones de la comunicación se establecen, como en todos los algoritmos, durante la *ENI* de comunicación esteganográfica.

IV. Descripción del Método Propuesto

En el empleo de métodos esteganográficos, para comunicar información sensible sin ser detectada, la invisibilidad y la cantidad de información que puede transmitirse es de gran importancia. Por tal motivo, se propone un método que supera las limitaciones inherentes de imperceptibilidad y capacidad de los actuales sistemas esteganográficos. Esto se logra a través del uso de ANNs, las cuales se emplean como unidad de procesamiento de la información sensible. Así, el método propuesto se puede aplicar a cualquier tamaño de imágenes, ya que a partir del uso de ANNs, en cada iteración del proceso de entrenamiento, y con base en el número de neuronas de la ANN utilizada, busca la mejor aproximación a la imagen secreta teniendo como condición inicial a la imagen portadora. Para ello, se toma como criterio el menor MSE y tiempo de procesamiento posible. La única consideración que se debe tener presente es que la imagen portadora debe ser del mismo tamaño que la imagen secreta. Así, entre más grande sean las imágenes, portadora y secreta, más grande será el tiempo de procesamiento requerido.

El método propuesto se compone por dos procesos: uno se encarga de la transmisión y el otro de la recepción. En el proceso de transmisión, se deben seleccionar dos imágenes digitales, una como portadora y otra como imagen secreta, ambas del mismo tamaño. Seguido a esto, la ANN debe entrenarse para que, a partir de la imagen portadora, se pueda generar una versión lo más parecida a la imagen secreta, ajustando los pesos de la ANN en cada paso para obtener el menor MSE posible. Al completarse la tarea de entrenamiento, se extrae de la ANN, la arquitectura y su configuración. De esta forma, se transmitirán dos elementos, cada uno por un canal diferente. Por un canal seguro o inseguro se comunica la imagen portadora que se utilizó en el entrenamiento. Por otro canal se comunica la configuración de la ANN extraída después de ser entrenada; es altamente recomendable que esta comunicación se realice a través de un canal criptográficamente seguro. En el proceso de recepción se emplea la imagen portadora y la configuración de la ANN para obtener la versión estimada de la imagen secreta (imagen secreta recuperada).

A. Normalización de Pixeles en las Imágenes

Una ANN acepta un conjunto de valores que deben escalarse y normalizarse dependiendo de las funciones de transferencia (funciones de activación) que posea. La ecuación (1) corresponde a la función de transferencia Sigmoide tangente hiperbólica que formará parte de la ANN que se usa en este caso.

$$f(x) = \frac{e^{g\,x} - e^{-g\,x}}{e^{g\,x} + e^{-g\,x}}, \tag{1}$$

donde g es el parámetro de la función que escala a la variable independiente.

Con la función descrita en la ecuación (1), los valores de entrada superiores a 0 producen un valor de salida positivo cercano a 1 y valores inferiores a 0 producen una salida negativa cercana a -1. Debido a esta característica en las funciones de transferencia de las ANNs, los pixeles de la imagen deberán escalarse y normalizarse para garantizar que los valores de entrada no se encuentren fuera de los límites de la función. Esto se soluciona con (2):

$$In_{pixel} = \frac{pixel - \frac{L}{2}}{\frac{L}{2}}, \tag{2}$$

donde In_{pixel} representa la matriz de entrada del entrenamiento a la ANN, *píxel* representa los valores que puede contener la matriz de $M \times N$ en una de las capas de color de la imagen digital, y L representa los niveles de cuantización en escala de grises en la imagen digital.

B. Implementación de ANNs

Para la implementación de la ANN necesaria para el método propuesto se utilizó MATLAB™ Neural Network Toolbox [48], ya que es una herramienta que proporciona algoritmos, modelos preprogramados y aplicaciones para crear, entrenar, visualizar y simular ANNs superficiales y profundas. Esta herramienta puede realizar la clasificación, la regresión, la agrupación, la reducción de dimensionalidad, el pronóstico de series de tiempo y el modelado y control de sistemas dinámicos. En esta herramienta, la instrucción `net` sin argumentos devuelve una nueva ANN sin entradas, capas intermedias y salidas. Sin embargo, la instrucción `net` puede tomar argumentos opcionales y se puede emplear de la siguiente forma: `net = network (numInputs, numLayers, biasConnect, inputConnect, layerConnect, outputConnect)`. Para ello, se toma como valores predeterminados los mostrados en la Tabla I.

Al generar el código en MATLAB™ [48], la ANN predeterminada tiene los siguientes elementos:

1) *Input* corresponde al número de entradas que la ANN puede procesar.

2) *Hidden* representa el número de capas ocultas que contendrá la ANN.
3) *Output* corresponde a la capa de salida y su número de capas es proporcional al número de salidas que la ANN pueda entregar.

TABLA I
VALORES PREDETERMINADOS EN LA INSTRUCCIÓN NET

Parámetro	Descripción
numInputs	Número de entradas, 0
numLayers	Número de capas, 0
biasConnect	Vector booleano de tendencia, ceros
inputConnect	Matrix booleana de conexión entrada-capa, ceros
layerConnect	Matriz booleana de conexión capa-capa, ceros
outputConnect	Vector booleano de salida, ceros

C. Neuronas en la ANN Feed-Forward

Debido a que una ANN del tipo *feed-forward*, con una capa oculta y suficientes neuronas en las capas ocultas, puede adaptarse a cualquier problema de asignación de entrada/salida finita, se propone usarla para este trabajo. Para ello, con base en la documentación proporcionada por MATLAB™ [48] para ANNs, se tiene que una ANN *feed-forward* se crea mediante la siguiente instrucción: `net=feedforwardnet([C])`, donde `C` corresponde al número de neuronas que la ANN almacenará en su capa oculta. El número de neuronas se escoge de forma arbitraria, en función de la complejidad en los patrones de aprendizaje.

D. Algoritmos de Entrenamiento

Se puede utilizar cualquier algoritmo de optimización numérica estándar para entrenar ANNs *feed-forward* multicapa, pero hay algunos métodos que han demostrado un mejor rendimiento. Estos métodos utilizan el gradiente de rendimiento de la ANN con respecto a las ponderaciones de la ANN, o el Jacobiano de los errores de la ANN con respecto a los ponderadores. El gradiente y el Jacobiano se calculan utilizando el algoritmo de retro propagación, que implica realizar cálculos hacia atrás a través de la ANN [46]. El cálculo de la propagación inversa se deriva utilizando la regla de la cadena del cálculo, esto de acuerdo con el algoritmo SCG que se describe a continuación. El algoritmo SCG denota, a través de la ecuación (3), la aproximación cuadrática al error E en una vecindad y de un punto w en el espacio de pesos, donde x^T es la matriz transpuesta de x.

$$E_{qw}(w+y) = E(w) + E'(w)^T y + \frac{1}{2} y^T E''(w) y \,. \quad (3)$$

Para determinar el mínimo de $E_{qw}(y)$ se deben encontrar sus puntos críticos, los cuales son la solución al sistema lineal definido por Moller [39] en la ecuación (4).

$$E'_{qw}(w+y) = E''(w)y + E'(w) = 0 \,. \quad (4)$$

De acuerdo con Orozco y Reyes [47], el algoritmo SCG pertenece a la clase de métodos gradiente conjugado, los cuales muestran convergencia superlineal en la mayoría de los problemas. Ellos afirman que, al usar un escalamiento del tamaño de paso, el algoritmo SCG evita una búsqueda con lo que se reduce el tiempo por iteración de aprendizaje y, por tanto, el algoritmo que proponen es más rápido que otros algoritmos de segundo orden. Ellos también afirman que obtienen mejores resultados que otros métodos de entrenamiento y ANNs probadas, como la ANN de propagación de retorno estándar y la ANN en cascada.

El algoritmo SCG es una buena opción para entrenar ANNs grandes y ANNs que se usan para reconocimiento de patrones. Esto se debe a que sus requisitos de memoria son relativamente pequeños y es mucho más rápido que los algoritmos estándar de descenso de gradiente. Su uso se basa en la siguiente declaración: `net.trainFcn='trainscg'`, donde `net.trainFcn` es la ANN *feed-forward* y `'trainscg'` es el algoritmo de entrenamiento SCG.

E. Entrenamiento, Validación y Prueba

Debido a que se cuenta con una cantidad limitada de elementos de entrenamiento supervisado para la ANN, se deben considerar tres aspectos relevantes: el entrenamiento, la validación y la prueba. Es importante decidir la cantidad de datos de entrenamiento que se utilizarán en cada caso. Para ello, se deben considerar lo siguiente:

(i) `net.divideParam.trainRatio` = *a*;
(ii) `net.divideParam.valRatio` = *b*;
(iii) `net.divideParam.testRatio` = *c*;

donde (i) define la proporción de datos de entrenamiento supervisado que se destina al entrenamiento, (ii) define la proporción de esos mismo datos destinada a validación y (iii) define la proporción destinada a las comprobaciones que realizará la ANN durante el entrenamiento; a su vez, las variables *a*, *b* y *c* pueden tomar valores comprendidos en el intervalo de 0 a 1.

Cabe aclarar que la suma de estas proporciones debe ser 1.0; esto es, el 100% de los datos destinados al entrenamiento supervisado.

Un aspecto importante a considerar es que se requiere que la ANN aprenda el 100% de los patrones de entrada y que calcule la salida con el menor error posible para la configuración de ANN seleccionada. La configuración descrita anteriormente es importante, ya que impacta de manera directa en la efectividad del entrenamiento de una ANN. Así, para entrenar una ANN es indispensable disponer de dos elementos: los datos de entrada que podrá procesar y los datos deseados de salida. En este caso, la imagen portadora y la imagen secreta. Así, en el método propuesto, tanto de la imagen portadora como de la imagen secreta, ambas en tamaño $M \times N$, se selecciona solo una de las capas de su estructura RGB. El método propuesto procesa matrices de tamaño $M \times N$ por cada par de imágenes (portadora y secreta). Si M es el número de filas y N es el número de columnas en las imágenes a procesar, entonces la ANN procesará N vectores columna como entradas, cada una de los cuales tiene M elementos. De esta forma, en cada paso de entrenamiento, la ANN producirá N vectores columna cada uno con M elementos. Esto muestra que el método propuesto requiere la imagen portadora tenga el mismo tamaño que la

imagen secreta.

En MATLAB™, esto se representa con la siguiente instrucción: `[nete,tr]=train (net, X, T, Xi, Ai, EW)`, que toma como argumentos los que se describen en la Tabla II y retorna las variables indicadas en la Tabla III.

TABLA II
ARGUMENTOS DE LA ETAPA DE ENTRENAMIENTO

Argumento	Descripción
net	Nombre de la red neuronal
X	Matriz de entrada
T	Matriz de muestras de salida (default: ceros)
Xi	Elemento de retardo de entrada (opcional)
Ai	Elemento de retardo de salida (opcional)
EW	Errores en los pesos (opcional)

TABLA III
VARIABLES DE SALIDA EN LA ETAPA DE ENTRENAMIENTO

Argumento	Descripción
nete	Red neuronal entrenada
tr	Registros del entrenamiento

En este caso, el argumento *X* es la matriz *M*×*N* normalizada de la imagen portadora y *T* la matriz *M*×*N* normalizada de la imagen secreta.

F. Desempeño

Es importante resaltar primero que el algoritmo propuesto, al ser del tipo SWE, tiene índices de modificación del objeto portador del 0% e índices de perceptibilidad también del 0%, ya que el portador existe y se comunica sin cambio entre las entidades que participan en la comunicación esteganográfica; esto es, la información sensible no se incrusta en el objeto portador. Como se ha comentado anteriormente, la información del tipo y configuración de la ANN se comunica entre las entidades participantes durante la *ENI* de la comunicación esteganográfica.

Ahora bien, en cuanto al desempeño del algoritmo propuesto, durante la etapa de entrenamiento es importante definir ciertas características que determinen hasta qué punto la ANN deberá seguir aprendiendo. Esto impacta en el tiempo de procesamiento y en el tiempo de uso del canal de comunicación seguro durante la *ENI* de la comunicación esteganográfica. Así, la técnica que se usa para determinar el estado de aprendizaje de una ANN es el error cuadrático medio (MSE: Mean Square Error) [40, 41]. Este error mide la diferencia global de dos conjuntos de datos de acuerdo a la ecuación (5).

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=1}^{n-1}[I_1(i,j) - I_2(i,j)]^2 . \quad (5)$$

En este caso, I_1 es la imagen secreta original e I_2 es la imagen secreta aproximada durante el entrenamiento.

De esta manera, en el entrenamiento de una ANN, el MSE mide la magnitud de las diferencias que existen entre los resultados reales obtenidos después del entrenamiento, contra los resultados esperados. Un MSE de cero significa que la ANN se encuentra completamente entrenada. Alcanzar un MSE de cero en escenarios donde los patrones de reconocimiento son muy complejos, equivale a un alto costo computacional. En este caso, un MSE del orden de $1x10^{-7}$ se considera suficiente.

G. Comunicación Esteganográfica

Note que para la comunicación esteganográfica, usando el método propuesto, se requiere de dos canales de comunicación. Uno de ellos puede considerarse inseguro y se destina a la comunicación de la imagen portadora. El otro, que se recomienda sea criptográficamente seguro, se usa para comunicar la configuración de la ANN con la que se estima la imagen secreta (sensible) a partir de la imagen portadora. De esta manera, se pretende que terceros tengan una alta complicación para relacionar ambos archivos como parte de una comunicación esteganográfica. El receptor, con la información de configuración de la ANN utilizada y la imagen portadora, está en condiciones de recuperar la imagen secreta. Para ello, no olvidar que es indispensable normalizar la imagen portadora utilizando los mismos patrones usados en la etapa de entrenamiento.

H. Interpretación de las Salidas

En el receptor, la ANN construida a partir de la configuración recibida es capaz de reconocer los patrones de la imagen portadora y obtener la imagen secreta. Para ello, se hace uso del proceso inverso indicado por la ecuación (6).

$$pixel = Out_{pixel} \times \frac{L}{2} + \frac{L}{2}, \quad (6)$$

donde Out_{pixel} es la salida normalizada de la ANN y L es la cuantificación de los niveles de grises en la imagen digital.

Finalmente, cuando todas de columnas de la imagen se hayan procesado, se requiere que sus elementos se conviertan al formato de 8 bits sin signo (*uint8*), lo cual corresponde al nivel de cuantificación en escala de grises de la imagen digital; de esta manera, se obtiene la imagen secreta recuperada.

V. RESULTADOS

En esta sección se muestran los resultados obtenidos al aplicar el método esteganográfico propuesto a imágenes digitales, portadora y secreta, de un tamaño arbitrario, pero siempre del mismo tamaño entre ellas. Nótese que, aunque el tamaño de imágenes empleado es arbitrario, sin pérdida de generalidad, los resultados presentados son representativos de lo que podría pasar para otros casos que empleen otros tamaños, ya que, independientemente del tamaño, el método propuesto buscará obtener el menor MSE y, consecuentemente el mayor PSNR posible para un par dado de imágenes digitales. Para la etapa de comunicación esteganográfica MSE se calcula de acuerdo a la ecuación (5) con I_1 como la imagen secreta original e I_2 es la imagen secreta recuperada a partir de la imagen de transporte. Por otro lado, el PSNR se calcula de acuerdo con la ecuación (7).

$$PSNR = 10\ log_{10}\left(\frac{MAX_I^2}{MSE}\right), \quad (7)$$

donde MAX_I es le máximo valor posible de pixel en la imagen.

Cabe resaltar que los resultados se obtuvieron con una máquina virtual con sistema operativo Windows 7 Ultimate de 32 bits, con 3 Gb de RAM y procesador AMD A6-3430MX APU con Radeon™ HD Graphics a 1.70 GHz.

A. Distintas Arquitecturas

Para mostrar la funcionalidad del proceso esteganográfico, se seleccionaron las imágenes digitales presentadas en la Figura 1. Una de estas imágenes se utilizó como portadora (Lenna) y la otra como imagen secreta (Baboon).

a) b)

Fig. 1. a) Lena: imagen portadora y b) Baboon: imagen secreta original.

Ambas imágenes son BMP de tamaño $M \times N$ pixeles, con $M=N=128$. En cuanto a la ANN, es indispensable establecer el número de entradas y salidas, el número de neuronas de la capa oculta y las funciones de transferencia (o de activación). La Tabla IV describe las arquitecturas de las ANNs utilizadas con 128 entradas y 128 salidas.

TABLA IV
PARÁMETROS EN LA ANN DE UNA CAPA OCULTA CUANDO SE USA LA FUNCIÓN SIGMOIDE TANGENTE HIPERBÓLICA

Arquitectura	Neuronas
1	10
2	100
3	1000
4	2000

De forma general, la Figura 2 muestra una representación simbólica de la ANN descrita por la Tabla IV, n representa el número de neuronas de la capa oculta, y es la única variable que puede modificarse durante el entrenamiento.

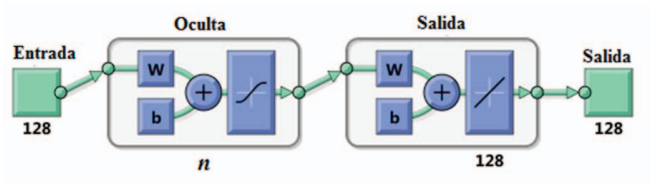


Fig. 2. Representación simbólica de la ANN utilizada.

A continuación, se muestran las Tablas V y VI con los resultados de la implementación del algoritmo SCG con las arquitecturas de la Tabla IV.

Para interpretar estos resultados se debe tener en cuenta que el MSE y el PSNR son dos métricas usadas para comparar la calidad de imágenes digitales cuando se tiene una imagen de referencia (otra imagen digital). Para el cálculo de estas métricas, la imagen digital de referencia es la imagen secreta original y la imagen en comparación es la imagen estimada en cada iteración durante el entrenamiento de la ANN. Así, la Tabla V muestra el MSE y el tiempo de cómputo alcanzado para las arquitecturas de ANN indicadas por la Tabla IV.

TABLA V
MSE Y TIEMPO DE CÓMPUTO PARA LAS DIFERENTES ARQUITECTURAS DE ANN EN LA ETAPA DE ENTRENAMIENTO

n	MSE	Tiempo de cómputo (s)
10	3.53×10^{-2}	16
100	6.06×10^{-3}	30
1000	7.71×10^{-5}	292
2000	1.56×10^{-5}	603

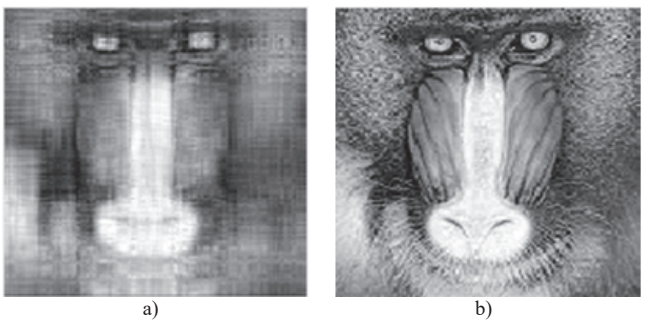
TABLA VI
MSE Y PSNR PARA CUANDO SE HA RECUPERADO LA IMAGEN SECRETA PARA DIFERENTES ARQUITECTURAS DE ANN

n	MSE Imagen	PSNR Imagen
10	578.7	20.50
100	99.07	28.17
1000	1.27	47.06
2000	0.2494	54.16

Note que a medida que se incrementa el valor de n el MSE se hace más pequeño, lo que significa que la ANN ha alcanzado una configuración que permitirá obtener una imagen secreta de mejor calidad respecto a la imagen secreta que se habrá de comunicar esteganográficamente.

Por otro lado, la Tabla VI muestra los valores de MSE y PSNR para cuando en la etapa de recepción se recupera la imagen secreta. En este caso los valores de MSE y PSNR están referidos a la imagen secreta original; de manera que, en cada paso de la ANN, la imagen secreta original se toma como referencia contra la imagen recuperada. De la misma manera que en la Tabla V, se puede notar que a medida que n toma valores más grandes, el MSE se hace más pequeño y el PSNR se hace más grande. Esto significa que la imagen secreta recuperada es más parecida a la imagen secreta original conforme n crece.

La Figura 3 muestra la imagen secreta recuperada con $n=10$ y $n=2000$ neuronas en la capa oculta. Ambos resultados se obtuvieron a partir de la misma imagen portadora ("Lenna").



a) b)

Fig. 3. Imagen secreta ("Baboon") recuperada con a) 10 neuronas y b) 2000 neuronas en la capa oculta.

B. Distintas Imágenes

Para analizar el comportamiento del método de aprendizaje

SCG, y su efectividad ante distintos escenarios, se establecieron tres imágenes portadoras y tres imágenes secretas. Las imágenes portadoras son: “Barbara”, “Macaw” y “Pepper”. En la Figura 4, se muestran dichas imágenes a color, pero para el proceso, y con la finalidad de reducir el tiempo de cálculo se toma sólo la capa R de cada imagen. De la misma forma, se seleccionaron tres imágenes secretas, de las cuales se utilizó una capa de color y se presentan en la Figura 5 en escala de grises. Estas imágenes son: “Lenna”, “Baboon” e “Island”.

a) b) c)

Fig. 4. Imágenes portadoras: a) “Barbara”, b) “Macaw” y c) “Pepper”.

a) b) c)

Fig. 5. Imágenes secretas: a) “Lenna”, b) “Baboon” y c) “Island”.

De esta forma, las ANNs se entrenaron con el algoritmo SCG y con la arquitectura que usa n=2000 neuronas. La Tabla VII muestra los resultados obtenidos para el MSE y el tiempo de cómputo durante la etapa de entrenamiento de la ANN con n=2000 neuronas. Nótese que en todos los casos el MSE es menor a 1×10^{-5}.

TABLA VII
MSE Y TIEMPO DE CÓMPUTO PARA LA ETAPA DE ENTRENAMIENTO DE LA ANN

Imagen secreta original	Imagen Portadora	MSE	Tiempo de cómputo (s)
Lenna	Barbara	4.73×10^{-9}	564
Baboon		3.99×10^{-8}	566
Island		5.07×10^{-7}	616
Lenna	Macaw	6.09×10^{-7}	603
Baboon		2.93×10^{-6}	619
Island		7.56×10^{-6}	646
Lenna	Pepper	7.20×10^{-9}	630
Baboon		1.75×10^{-8}	650
Island		7.57×10^{-7}	657

Por otro lado, la Tabla VIII muestra los resultados para el MSE y el PSNR en la etapa de recepción cuando la ANN emplea n=2000 neuronas en su capa oculta. Se resaltan los casos en que el MSE es cero y el PSNR es infinito, lo que quiere decir que la imagen secreta recuperada es igual a la imagen secreta original.

Note que en este trabajo se calculan dos MSE, uno se calcula para cada iteración durante el proceso de entrenamiento mientras se intenta obtener la imagen secreta a partir de la imagen portadora, y otro para comparar la imagen secreta recuperada contra la imagen secreta original. Durante la etapa de entrenamiento el MSE se determina por el número de neuronas que se utilicen en la ANN, si se quisiera reducir el MSE en la etapa de entrenamiento se debe aumentar el número de neuronas en la ANN, lo que implica que se aumentaría el tiempo de cálculo para obtener una imagen digital que mejor aproxime a la imagen secreta original a partir de la imagen portadora.

TABLA VIII
MSE Y PSNR PARA LA ETAPA DE RECEPCIÓN

Imagen secreta recuperada	Imagen Portadora	MSE Imagen	PSNR Imagen
Lenna	Bárbara	0	Inf
Baboon		1.22×10^{-4}	87.26
Island		5.70×10^{-3}	70.54
Lenna	Macaw	6.50×10^{-3}	69.98
Baboon		40.30×10^{-3}	62.07
Island		11.80×10^{-2}	57.41
Lenna	Pepper	0	Inf
Baboon		0	Inf
Island		9.50×10^{-3}	68.34

Por otro lado, el MSE estimado entre la imagen secreta recuperada y la imagen secreta original establece la posibilidad de tener un algoritmo con o sin pérdidas en la comunicación esteganográfica. Nótese en este caso, que el tiempo requerido para obtener la imagen secreta recuperada es el mismo que se requirió en la etapa de entrenamiento para estimar los pesos en la ANN.

Existe una relación entre el número de neuronas en la arquitectura de la ANN y su efectividad del aprendizaje cuando se usa el algoritmo SCG. Para ello, en la Figura 6 se muestra, en escala logarítmica para la etapa de entrenamiento, que el MSE entre la imagen portadora y la imagen secreta se va reduciendo a medida que n aumenta.

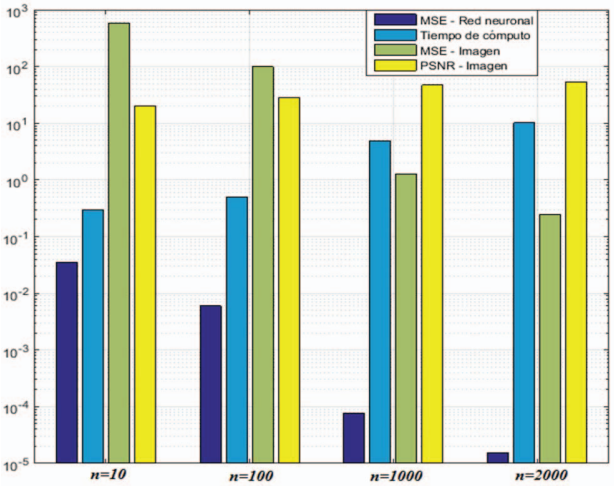


Fig. 6. MSE y tiempo de cómputo en la etapa de entrenamiento de la ANN, así como MSE y PSNR en la etapa de recepción.

Contrariamente, y como era de esperarse, a medida que n aumenta el tiempo de cómputo también aumenta. Ahora bien, nótese que el MSE, que se calcula entre la imagen secreta recuperada y la imagen secreta original, se reduce conforme n

aumenta. De igual forma, el PSNR, que se calcula entre la imagen secreta recuperada y la imagen secreta original, aumenta conforme *n* se aumenta.

Por otra parte, se estimó la efectividad del algoritmo de aprendizaje a partir de la imagen de portadora usada. Como se ve en la Figura 7, también con escala logarítmica, las imágenes portadoras que visualmente presentan un mayor número de tonalidades, o tienen mayor contraste entre sus pixeles (imágenes más alejadas de ser imágenes de un solo tono), hacen que el algoritmo de entrenamiento sea más efectivo; prueba de ello, es que cuando la imagen secreta es "Lenna" o "Baboon" y la imagen de portadora es "Barbara" o "Pepper", el MSE se hace cero, lo que significa que la imagen secreta recuperada resultó ser igual a la imagen secreta original.

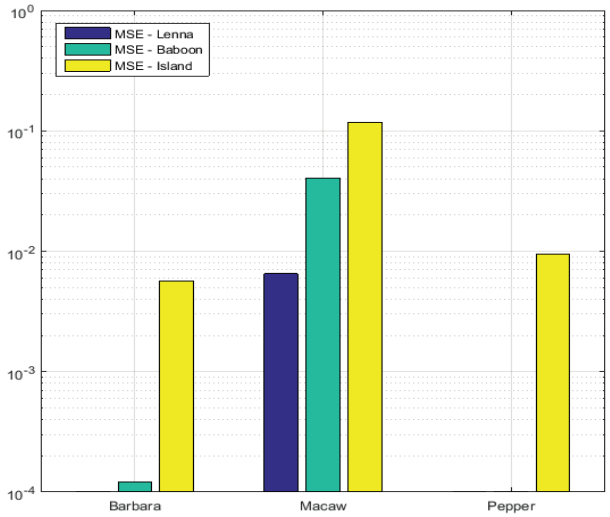


Fig. 7. Eficiencia a partir del cambio de imágenes de prueba en el entrenamiento.

Finalmente, se hizo una prueba que destaca la característica de integridad/autenticación que ofrece el método propuesto. Esta prueba consiste en dos etapas; en la primera se utilizó como imagen portadora a "Pepper", y como imagen secreta a "Lenna" para entrenar una ANN con *n*=2000 neuronas. En la segunda etapa de esta prueba, también con *n*=2000 neuronas, se utilizó como imagen portadora la imagen de "Pepper" pero con un ligero cambio en su brillo/contraste. En este caso, la imagen secreta recuperada no corresponde con la imagen de "Lenna" (ver Figura 8). Esto es una importante ventaja, ya que la imagen recuperada no tiene relación evidente con la imagen secreta original. En esencia, que no es posible obtener una imagen secreta recuperada que se parezca (al menos en apariencia) a la imagen secreta original cuando se hace un ligero cambio en la imagen portadora.

VI. Conclusiones

La capacidad que demostró tener el algoritmo SCG en el reconocimiento de patrones de imágenes digitales (secretas) lo hace una buena opción para aplicarse en el algoritmo esteganográfico del tipo SWE que aquí se propone. Es importante destacar que el algoritmo propuesto puede, en principio, utilizar una ANN de cualquier tipo y tamaño. Eso quedará a elección del usuario en función de la calidad y el tiempo de procesamiento que se tolere para aproximar la imagen secreta a partir de la imagen portadora que se utilice.

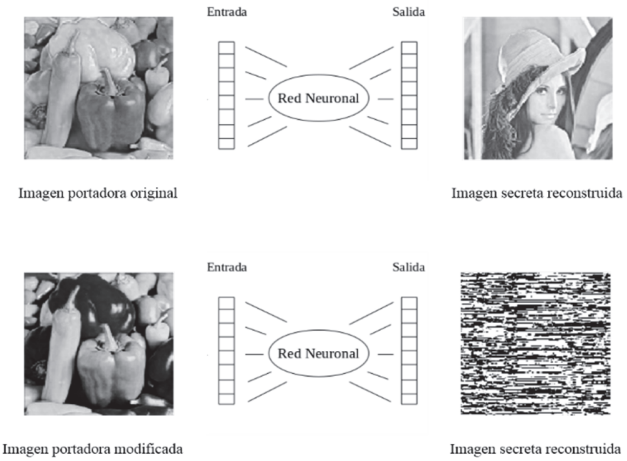


Fig. 8. Demostración de que la imagen secreta recuperada no corresponde con la imagen secreta original cuando se modifica el brillo contraste de la imagen portadora.

Con el algoritmo SCG, y aumentando el número de neuronas en la arquitectura de la ANN se reduce el tiempo de cómputo de la etapa de entrenamiento y se reduce el MSE entre la imagen secreta recuperada y la imagen secreta original. Por otro lado, se pudo mostrar que el algoritmo tuvo un mejor desempeño cuando se usaron imágenes digitales que presentan más cambios de tonalidad. En estos casos, el MSE entre la imagen secreta recuperada y la imagen secreta original llegó a ser cero y el PSNR llegó a ser infinito. Esto significa que la imagen secreta recuperada fue igual a la imagen secreta original; y, por lo tanto, el método esteganográfico no generó pérdida de información entre la imagen secreta recuperada y la imagen secreta original. Esta condición abre la posibilidad de probar el método esteganográfico propuesto para comunicar datos y no solo imágenes digitales. Los resultados obtenidos ponen de manifiesto el cumplimiento de las tres características que posee el algoritmo propuesto. A saber, estas características son: i) *una alta capacidad de comunicación esteganográfica*, ya que la selección del tamaño de la imagen secreta no depende del tamaño de la imagen portadora; así, la única limitante que se identifica está en el nivel de pérdida que se tolere entre la imagen recuperada y la imagen original, ii) *imperceptibilidad del 100% de la imagen secreta en la imagen portadora*, ya que el algoritmo propuesto es del tipo SWE; esto es, esteganografía sin incrustación, iii) *mecanismo de integridad/autenticación*, ya que una vez superada la etapa de entrenamiento, para una imagen secreta particular, el método propuesto mostró una alta dependencia de la imagen de portadora en la obtención de una versión comprensible de la imagen secreta en el proceso de recuperación. Bajo las premisas de funcionalidad y seguridad que definen al algoritmo propuesto, el número óptimo de neuronas de la capa oculta de la ANN cambiará en función del nivel de tolerancia a las pérdidas que se decida entre la imagen

secreta original y la imagen que se desea comunicar subliminalmente. Estas decisiones se deberán tomar durante la *ENI* de la comunicación esteganográfica. Por otro lado, si las imágenes que se desean utilizar son de gran tamaño seguramente se definirá el uso una ANN más grande, de manera que sea más rápido aproximar a la imagen secreta a partir de la imagen portadora. Finalmente, se debe precisar que la seguridad del método propuesto depende de fortaleza en la confidencialidad del intercambio de información durante la *ENI* de la comunicación esteganográfica.

AGRADECIMIENTOS

A. A. López-Hernández agradece al Consejo Nacional de Ciencia y Tecnología [CVU– 951769]. R. Vázquez-Medina agradece el apoyo financiero otorgado por el Instituto Politécnico Nacional [SIP-20196692]. L. Palacios-Luengas agradece a la Universidad Autónoma Metropolitana Unidad Iztapalapa por la estancia como profesor visitante. R. F. Martínez-González agradece al Tecnológico Nacional de México por el apoyo con el Proyecto 6622.18-P: Generación de un esquema esteganográfico caótico de bajo impacto para la imagen portadora.

REFERENCIAS

[1] B. Ramalingam, R. Amirtharajan, and J. B. B. Rayappan, “Multiplexed stego path on reconfigurable hardware: A novel random approach,” *Computers & Electrical Engineering*, Vol. 55, pp. 153-163, 2016. 10.1016/j.compeleceng.2016.02.010.

[2] M. Rashid, M. Imran, A. Jafri and T. Al-Somani, “Flexible Architectures for Cryptographic Algorithms — A Systematic Literature Review,” *Journal of Circuits, Systems and Computers*, Vol. 28 no. 03, pp.1930003, 2019. 10.1142/S0218126619300034.

[3] O-J. Kwon, S. Choi, and B. Lee, “A Watermark-Based Scheme for Authenticating JPEG Image Integrity,” *IEEE Access*, Vol. 6, pp.46194-46205, 2018. 10.1109/ACCESS.2018.2866153.

[4] S. Atawneh, A. Almomani, H. Al Bazar, P. Sumari and B. Gupta, “Secure and imperceptible digital image steganographic algorithm based on diamond encoding in DWT domain,” *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18451-18472, (2016). 10.1007/s11042-016-3930-0.

[5] A. Cheddad, J. Condell, K. Curran and P. M. Kevitt, “Digital image steganography: survey and analysis of current methods,”. *Signal Processing*, vol. 90, pp.727–752, 2010. 10.1016/j.sigpro.2009.08.010.

[6] P-Y. Chen and W-E. Wu, “A DWT based approach for image steganography,”. *International Journal Applied Science Engineering* vol. 4 no. 3, pp. 275–290, 2006.

[7] V. Kumar and D. Kumar, “Performance evaluation of dwt-based image steganography,” in *Proc. 2nd International Conference on Advance Computing*, Patiala, 2010, pp 223–228.

[8] I. Kadhim, P. Premaratne, P. Vial and B. Halloran, “Comprehensive survey of image steganography: Techniques, Evaluations, and trends in future research,” *Neurocomputing* vol. 335, pp.299-326, 2019. 10.1016/j.neucom.2018.06.075.

[9] V. K. Sharma, D. Srivastava and P. Mathur, P. “Efficient image steganography using graph signal processing,” *IET Image Processing*, vol. 12, no. 6, pp.1065-1071, 2018. 10.1049/iet-ipr.2017.0965.

[10] N. F. Johnson and S. Katzenbbeisser, “A survey of steganographic techniques, Information Hiding,” *Artech House*, 43–78.

[11] Y. Ben Ali, “Smell Bees Optimization for new embedding steganographic scheme in spatial domain,” *Swarm and Evolutionary Computation*, vol. 44, pp.584-596, 2019. 10.1016/j.swevo.2018.08.003.

[12] V. Kumar and D. Kumar, “A modified DWT-based image steganography technique,” *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 13279-13308, 2018. 10.1007/s11042-017-4947-8.

[13] M. Y. Rafiq, G. Bugmann and D. J. Easterbrook, “Neural network design for engineering applications,” *Computers & Structures*, vol. 79, no. 17, pp. 1541-1552, 2001. 10.1016/S0045-7949(01)00039-6.

[14] E. A. Ruelas Santoyo, J. A. Vazquez Lopez, J. Yanez Mendiola, I. Lopez Juarez and C. F. Bravo Barrera, "Condition Estimation Of Carbon Steel Using A Neuro-Fuzzy System And Image Processing," in IEEE Latin America Transactions, vol. 13, no. 7, pp. 2322-2328, July 2015. doi: 10.1109/TLA.2015.7273794.

[15] J. Almeida Bessa, D. Almeida Barroso, A. Rego da Rocha Neto and A. Ripardo de Alexandria, "Global location of mobile robots using Artificial Neural Networks in omnidirectional images," in IEEE Latin America Transactions, vol. 13, no. 10, pp. 3405-3414, Oct. 2015. doi: 10.1109/TLA.2015.7387248.

[16] E. Cavalcanti Neto, P. C. Cortez, T. S. Cavalcante, V. E. Rodrigues, P. P. Reboucas Filho and M. A. Holanda, "3D Lung Fissure Segmentation in TC images based in Textures," in IEEE Latin America Transactions, vol. 14, no. 1, pp. 254-258, Jan. 2016. doi: 10.1109/TLA.2016.7430087.

[17] M. Boell, H. Ramos Alves, M. Volpato, D. Ferreira and W. Lacerda, "Exploiting Feature Extraction Techniques for Remote Sensing Image Classification," in IEEE Latin America Transactions, vol. 16, no. 10, pp. 2657-2664, October 2018. doi: 10.1109/TLA.2018.8795147.

[18] C. Ângela, W. Carolina and C. Carlos, "Medical Image Segmentation Using the Kohonen Neural Network," in IEEE Latin America Transactions, vol. 17, no. 02, pp. 297-304, February 2019. doi: 10.1109/TLA.2019.8863176.

[19] C. Bishop, “Pattern recognition and machine learning (information science and statistics),” Springer, New York, Aug. 2006.

[20] Ch-K. Chan and L. M. Cheng, “Hiding data in images by simple LSB substitution,” *Pattern Recognition*, vol. 37, no. 3, pp.469-474, 2004. 10.1016/j.patcog.2003.08.007.

[21] R. Tavares and F. Madeiro, "Word-Hunt: A LSB Steganography Method with Low Expected Number of Modifications per Pixel," in IEEE Latin America Transactions, vol. 14, no. 2, pp. 1058-1064, Feb. 2016. doi: 10.1109/TLA.2016.7437258.

[22] R. Martínez-González, J. A. Díaz-Méndez, L. Palacios-Luengas, J. López-Hernández and R. Vázquez-Medina, “A steganographic method using Bernoulli’s chaotic maps,” *Computers & Electrical Engineering*, vol. 54, pp.435-449, 2016. 10.1016/j.compeleceng.2015.12.005.

[23] K. Jung and K. Yoo, “Steganographic method based on interpolation and LSB substitution of digital images,” *Multimedia Tools and Applications*, vol. 74, no. 6, pp.2143-2155, 2014. 10.1007/s11042-013-1832-y.

[24] K. Muhammad, M. Sajjad S. and Baik, “Dual-Level Security based Cyclic18 Steganographic Method and its Application for Secure Transmission of Keyframes during Wireless Capsule Endoscopy,” *Journal of Medical Systems*, vol. 40, no. 114. 10.1007/s10916-016-0473-x.

[25] A. Abuadbba and I. Khalil, “Wavelet based steganographic technique to protect household confidential information and seal the transmitted smart grid readings,” *Information Systems*, vol. 53, pp.224-236, 2015. 10.1016/j.is.2014.09.004.

[26] S. Braci, C. Delpha and R. Boyer, “How quantization based schemes can be used in image steganographic context,” *Signal Processing: Image Communication*, vol. 26, no. 8-9, pp.567-576, 2011. 10.1016/j.image.2011.07.006.

[27] Cox, I., et al. *Digital Watermarking and Steganography*. Morgan Kaufmann, 2 Edition (2007). ISBN-13: 978-0123725851.

[28] Chaumont, M., *Deep Learning in steganography and steganalysis from 2015 to 2018*, Chapter in Digital Media Steganography: Principles, Algorithms, Advances, Elsevier 2019.

[29] G. Villarrubia, J. De Paz, P. Chamoso. and F. la Prieta, “Artificial neural networks used in optimization problems,” *Neurocomputing*, vol. 272, pp.10-16, 2018. 10.1016/j.neucom.2017.04.075.

[30] J. Zeng, X. Zhao, J. Gan, C. Mai, Y. Zhai. and F. Wang, “Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition,” *Computational Intelligence and Neuroscience*, vol. 2018, pp.1-11. Art. ID. 3803627, 2018. 10.1155/2018/3803627.

[31] R. Jarusek, E. Volna and M. Kotyrba, “Robust steganographic method based on unconventional approach of neural networks,” *Applied Soft Computing*, vol. 67, pp. 505-518, 2018. 10.1016/j.asoc.2018.03.023.

[32] S. Baluja, “Hiding Images in Plain Sight: Deep Steganography,” *Advances in Neuronal Information Processing Systems 30*, Curran Associates. Inc, 2017, 2069-2079.

[33] P. Wu, Y. Yang and X. Li, "StegNet: Mega Image Steganography Capacity with Deep Convolutional Network," *Future Internet*, Multidisciplinary Digital Publising Institute, 2018, 10,54.

[34] Goodfellow, I; Pouget-Abadie, J; Mirza, M; Xu, B; WardeFarley, D; Ozair, S; Courville, A and Bengio; Y. *Generative adversarial nets*. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence et K. Q. Weinberger, ´editeurs: *Advances in Neural Information Processing Systems* 27, pages 2672–2680. Curran Associates, Inc., 2014.

[35] Shi, H; Dong, J; Wang, W; Qian, Y and Zhang, J. *SSGAN: Secure Steganography Based on Generative Adversarial Networks*. In Proceedings of the 18th Pacific-Rim Conference on Multimedia, PCM'2017, volume 10735 de Lecture Notes in Computer Science, pages 534–544. Springer, september 2017.

[36] S. Husien and H. Badi, "Artificial neural network for steganography," *Neural Computing and Applications*, vol. 26, no. 1, pp. 111-116, 2014. 10.1007/s00521-014-1702-1.

[37] A. Santos Brandao and D. Calhau Jorge, "Artificial Neural Networks Applied to Image Steganography," *IEEE Latin America Transactions*, vol. 14, no. 3, pp.1361-1366, 2016. 10.1109/TLA.2016.7459621.

[38] R. K. Chawla and S. K. Muttoo, "Steganography using bit plane complexity segmentation and artificial neural network," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, pp. 1-9, 2017. 10.26483/ijarcs.v8i5.4017.

[39] S. Alam, T. Ahmad and M. N. Doja, "A Novel Edge Based Chaotic Steganography Method Using Neural Network. Advances in Intelligent Systems and Computing," in *Proc. 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, pp. 467–475, 2017. 10.1007/978-981-10-3156-4_48.

[40] A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," *2010 20th International Conference on Pattern Recognition*, Istanbul, 2010, pp. 2366-2369. doi: 10.1109/ICPR.2010.579.

[41] J. Søgaard, S. Forchhammer and J. Korhonen, "No-Reference Video Quality Assessment Using Codec Analysis," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 10, pp. 1637-1650, Oct. 2015. doi: 10.1109/TCSVT.2015.2397207.

[42] J. Zhu, R. Kaplan, J. Johnson and L. Fei-Fei, "Hiding data with deep networks," *Proceeding of the European Conference on Computer Vision* (ECCV), 2018, 657-672.

[43] X. Duan, K. Jian and B. Li, "Guo, D., Zhang, E., Qin, C., Reversible image steganography scheme based on a U-Net structure," *IEEE Access*, IEEE, 2019, 7, 9314-9323.

[44] D. Hu, L. Wang, E. Jiang, S. Zheng and B. Li, "A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks," *IEEE Access,* 6 (2018) 38303-38314.

[45] A. Moller, "Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural Networks*, vol. 6, no.4, pp.525-533, 1993. 10.1016/S0893-6080(05)80056-5.

[46] N. Seth, A. Ubrani, S. Mane, and F. A. Kazi, "Comparative analysis of major jacobian and gradient backpropagation optimizers of ann on svpwm," in *Soft Computing and Signal Processing*. Springer, 2019, pp. 345-357.

[47] J. Orozco and C. A. R. García, "Detecting pathologies from infant cry applying scaled conjugate gradient neural networks," in *Proc. European Symposium on Artificial Neural Networks, Bruges (Belgium)*, pp. 349-354. ISBN 2-930307-03-X.

[48] Mathworks. (2011). *Global Optimization Toolbox: User's Guide* (r2011b). Retrieved November 10, 2011 from www.mathworks.com/help/pdf_doc/gads/gads_tb.pdf

Ricardo Francisco Martínez-González. Nació en Veracruz, México. Obtuvo su título de Ingeniero en Electrónica por el Instituto Tecnológico de Veracruz. Posteriormente estudió su Maestría en el Instituto Nacional de Astrofísica, Óptica y Electrónica; lugar donde también estudió su Doctorado, especializándose en el campo de las Aplicaciones Caóticas. Hoy en día trabaja en el Departamento de Ingeniería Eléctrica-Electrónica del Instituto Tecnológico de Veracruz.

Andrés Ali López-Hernández. Nació en Veracruz, México. Obtuvo su título de Ingeniero en Electrónica por el Instituto Tecnológico de Veracruz en el año 2018. Posteriormente, inició sus estudios de posgrado en la Maestría en Eficiencia Energética y Energías Renovables; siguiendo como línea de investigación el Control Energético por el Instituto Tecnológico de Veracruz.

José Antonio Hernández-Reyes. Nació en Veracruz, México. Obtuvo el título de Ingeniero Industrial en Electrónica por el Instituto Tecnológico de Veracruz. Realizó la Maestría en Ciencias en Ingeniería Electrónica en el Centro Nacional de Investigación y Desarrollo Tecnológico, especializándose en la aplicación de Redes Neuronales Artificiales en el Control de Procesos. Actualmente, es profesor del Departamento de Ingeniería Eléctrica – Electrónica del Instituto Tecnológico de Veracruz.

Leonardo Palacios Luengas es miembro del IEEE. Doctor en Comunicaciones y Electrónica en el año 2016 en la Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME) Unidad Culhuacan del Instituto Politécnico Nacional (IPN). Actualmente, es profesor visitante en la Universidad Autónoma Metropolitana, Unidad Iztapalapa (UAM-I). Sus áreas de interés están en la criptografía, esteganografía, sistemas embebidos y diseño electrónico digital.

Rubén Vázquez–Medina. Nació en México. En 1988 obtuvo el grado de ingeniero en electrónica por la Universidad Autónoma Metropolitana (UAM–Iztapalapa), en 1993 obtuvo el grado de maestro en ciencias por el Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (IPN). En 2008, obtuvo el grado de doctor en ciencias por la UAM–Iztapalapa. Hizo una estancia en el Centro de Física Aplicada y Tecnología Avanzada de la Universidad Nacional Autónoma de México (UNAM) Campus Juriquilla. Actualmente, es investigador en el IPN. Sus áreas de interés están en sistemas bioinspirados, sistemas no lineales y seguridad de la información, así como la esteganografía y la criptografía no convencional.