Inteligencia Artificial aplicada al Análisis Forense Digital: Una revisión preliminar

J. J. Cano, J. D. Miranda y S. Pinzón

***Resumen*—The digital forensic analysis seeks the application of scientific and statistical techniques to identify, collect, preserve and present the relevant digital evidence that allows the hypothesis to be affirmed or rejected against a possible criminal act. The current methods of digital forensic analysis, although effective for the visual analysis of the material evidence, do not allow to execute in an automated way and for large volumes of data, correlation studies on the obtained files, validation of metadata and identification of anomalies in files of text, graphics or audiovisual. It is for this reason that artificial intelligence techniques have been proposed for data processing, identifying patterns and trends that allow noticing aspects that are not visually perceptible. This paper discusses the role that artificial intelligence can play in digital forensic analysis, proposing a review of the literature, in order to illustrate the areas of computer forensics in which artificial intelligence techniques have been used to date. This, to identify a new work niche in this area, hoping that the ideas in this document can represent promising directions for the development of more efficient and effective computer forensic tools.**

Palabras clave — Forensic Computing, Forensic Analysis, Artificial Intelligence, Vector Support Machines, Artificial Neural Networks, Autonomous Systems, Intelligent Agents.

I. Introducción

Debido al rápido crecimiento en la aplicación de tecnologías digitales en diferentes entornos de la sociedad, son cada vez más los casos de ciberdelincuencia y mayor la cantidad de información que se logra extraer como rastro de un acto de delincuencia digital.

En complemento, los ataques informáticos se han vuelto más sofisticados, no solo por la experticia de los atacantes, sino por las herramientas que usan estos para acometer las acciones no autorizadas. En consecuencia, se hace necesario el uso de instrumentos que asistan al personal especializado para el estudio de los elementos materiales probatorios recopilados frente a la ocurrencia de un evento de carácter informático y que favorezcan en términos de efectividad y velocidad a la resolución de los diferentes casos forenses.

Sin embargo, a través de los años la intervención humana ha sido en algunos casos insuficiente para el análisis oportuno de un ataque y su consiguiente respuesta. Estos ataques son encabezados por atacantes habilidosos que emplean herramientas cada vez más sigilosas y perfeccionadas para esta labor [1], tales como agentes inteligentes, gusanos o virus informáticos, que son analizados por personal humano que desconoce su comportamiento y que busca de forma superficial rastros de una dinámica no conocida subyacente. Adicionalmente, existen otros desafíos, tal como lo documenta [2], que hacen que el análisis forense digital sea complejo, tedioso y en ocasiones infructuoso:

- La complejidad del problema y la heterogeneidad de los datos en su adquisición.
- El volumen excesivo de datos procedentes de múltiples fuentes y la falta de técnicas estandarizadas para procesarlos.
- La falta de técnicas que encuentren la correlación en la información contenida en la evidencia digital.
- La falta de estandarización en las zonas horarias y los registros de tiempo de los eventos encontrados.

Estos factores hacen que se requiera la aplicación de extensas etapas adicionales de preprocesamiento de los datos, se aumente el tiempo de respuesta y no se termine de efectuar un proceso eficiente de análisis forense, dejando al atacante en ocasiones dentro del sistema, lo que aumenta los riesgos para la empresa, los costos en reparaciones y la recuperación de los daños informáticos.

Un reporte del Ponemon Institute [3] revela que, a 2014, se ha aumentado el costo del cibercrimen en más del 9% con respecto a años anteriores, y el tiempo para resolver un ataque a incrementado a 45 días, lo que representa un alza de 40% respecto a las mediciones de los años previos. Esto significa que las técnicas actuales utilizadas para el análisis forense digital no son suficientes para el análisis oportuno y concluyente de los eventos informáticos que se presentan.

Es por esta razón que es propicio el uso adicional de agentes semiautónomos inteligentes que puedan aportar en la eficacia del análisis forense y en la toma de decisiones, con base en la experiencia. Esto puede hacerse mediante la implementación de métodos de Inteligencia Artificial (AI) como las Máquinas de

J. J. Cano, Facultad de Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana de Bucaramanga. Colombia. jjcano@yahoo.com.

J. D. Miranda, Facultad de Electrónica e Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana de Bucaramanga. Colombia. juliandariomiranda@gmail.com. Corresponding author.

S. Pinzón. Facultad de Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana de Bucaramanga. Colombia. spinzonsarmiento@gmail.com.

https://doi.org/10.12804/si9789587844337.02

Soportes Vectoriales (SVM), Redes Neuronales (NN), Agentes Inteligentes y Aprendizaje de Máquina (ML), entre otros, aplicados a los sistemas que asisten al personal especializado en la detección, prevención y mitigación del cibercrimen [4].

En el campo del análisis forense automatizado se utilizan métodos de AI con el objetivo de automatizar los diferentes procesos y análisis que se realizan en los dispositivos, teniendo en cuenta un amplio volumen de datos. Además, estos datos son inocuos por sí mismos, ya que provienen de diferentes fuentes y para ser útiles debe eliminarse la alta correlación existente, con el fin de descartar los que son intrascendentes o similares entre sí, y ejecutar un análisis pertinente de la dinámica no visible de estos.

Tal como lo documenta Laurance Merkle [5], las herramientas existentes en el mercado que realizan análisis estadísticos, análisis de trafico de red y análisis de sesiones, entre otras, lo hacen de forma superficial y mediante el estudio de umbrales, razón por la cual resultan insuficientes para procedimientos de análisis forense multinomial (múltiples variables de entrada).

El objetivo de este artículo es presentar los resultados de un breve estudio del estado del arte de referentes de interés en el campo de la implementación de técnicas de inteligencia artificial, aprendizaje automático, máquinas de soportes vectoriales y sistemas expertos, puestos a disposición del análisis forense digital, de tal forma que, se refuerce la identificación de patrones y componentes que no son visualmente perceptibles, se correlacionen eventos para encontrar una secuencia no evidente y se puedan tomar decisiones oportunas frente a los eventos informáticos. A continuación, se detalla la fundamentación teórica que enmarca el contexto de esta investigación, seguida de la metodología propuesta para la búsqueda de información, los resultados obtenidos y la discusión de estos.

II. Fundamentación Teórica

Esta sección se divide en cinco temáticas de estudio, consideradas relevantes para el desarrollo de la investigación: inteligencia artificial, aprendizaje automático, máquinas de soportes vectoriales, redes neuronales artificiales, y sistemas expertos y agentes inteligentes.

A. Inteligencia Artificial

La Inteligencia Artificial (AI) está definida como la capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación dinámica y flexible [6]. Según [7], una máquina es considerada como inteligente cuando se trata de un agente dinámico que percibe su entorno y lleva a cabo acciones que maximicen sus posibilidades de éxito en algún objetivo o tarea. En otras palabras, la AI involucra un agente dinámico flexible que aprenda de las observaciones con el fin de presentar una respuesta cada vez más acertada y acorde con el aprendizaje.

Son muchos los campos en los que se aplica AI, entre los cuales destacan la lingüística computacional, la medicina, la robótica, los videojuegos, la domótica, la automatización y la seguridad informática; buscando resolver problemas relacionados con la búsqueda de nuevas heurísticas, la representación del conocimiento, la planeación de estrategias, el procesamiento del lenguaje, la percepción de los patrones y el análisis forense. En este último contexto, la AI ha puesto especial atención en el aprendizaje de máquina, tal como lo muestra la Fig. 1, una ramificación de la AI en la que se busca el desarrollo de técnicas estadísticas que le permiten a la máquina mejorar la ejecución de las tareas con la experiencia.

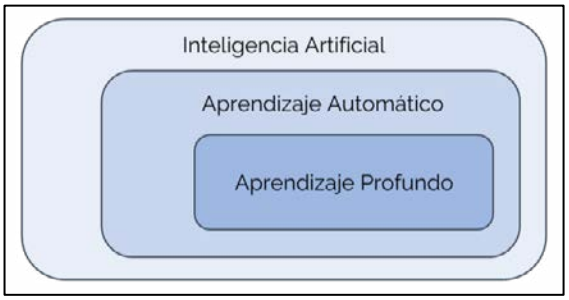

Fig. 1. Relación entre la inteligencia artificial, el aprendizaje de máquina y el aprendizaje profundo. Elaboración propia**.**

Dentro del aprendizaje de máquina existen desarrollos algorítmicos específicos que permiten que el software se entrene a sí mismo para llevar a cabo tareas como el habla y el reconocimiento de patrones, mediante la exposición a gran cantidad de datos, en lo que se conoce como el aprendizaje profundo.

De esta forma, para resolver problemas mediante métodos computacionales, se requiere de secuencias de instrucciones que transformen un conjunto de variables de entrada conocidas, en un grupo de variables de salida deseadas. Para procesos en los cuales no se conoce su comportamiento en su totalidad, se construye una aproximación. Ésta puede no explicar la totalidad de los eventos consultados, pero permite la detección de patrones que sirven como indicios para su identificación y clasificación. Este es el nicho del Aprendizaje Automático, los sistemas expertos y agentes inteligentes.

B. Aprendizaje Automático

En general, el aprendizaje automático es el proceso de programación de una unidad computacional para optimizar los criterios de rendimiento, haciendo uso de muestras recolectadas de experiencias pasadas [8]. Un modelo de aprendizaje puede ser supervisado o no supervisado. En el aprendizaje supervisado se orientan las predicciones del algoritmo mediante un conjunto de categorías o etiquetas. Los algoritmos de clasificación y regresión son ejemplos de aprendizaje supervisado. En la clasificación se intenta hacer una predicción de la categoría a la que corresponde una observación mientras que, en regresión, se intenta hacer una predicción de un valor numérico.

Por otro lado, en el aprendizaje no supervisado no existen categorías adjuntas a los datos de entrada. Los algoritmos de Clustering, estimación de densidad y reducción de dimensionalidad son ejemplos de aprendizaje no supervisado. En Clustering se intenta agrupar datos con características similares, mientras que en estimación de densidad se pretende encontrar valores estadísticos que describan al conjunto de datos. En cuanto a la reducción de dimensionalidad, se busca reducir la cantidad de características de los datos para reducir el costo computacional de su procesamiento y posibilitar su representación más fácilmente [9].

Para que el modelo aprenda de los datos se ejecuta un proceso inicial de entrenamiento que se realiza de forma iterativa. Cuando el modelo de aprendizaje ha sido expuesto a todo el conjunto de entrenamiento, se dice que ha pasado un *epoch* [10]. Existen tres conjuntos de datos con los que se ejecutan las fases de entrenamiento, validación y prueba: el training set es el conjunto de datos usados como referencias para seleccionar los pesos asociados a las unidades de procesamiento del modelo de aprendizaje y crear las conexiones entre las mismas; el *validation set* corresponde al conjunto de datos usados durante el proceso de entrenamiento para contrastar el desempeño por cada ciclo de aprendizaje, clasificando datos desconocidos; y el test set es el conjunto de datos desconocidos para el modelo de aprendizaje, con los que se prueba el desempeño.

C. Máquinas de Soportes Vectoriales (SVM)

Las máquinas de soportes vectoriales (SVM) son una familia de algoritmos de aprendizaje automático de tipo supervisado que generalmente se emplea en procesos de clasificación y regresión de dos conjuntos de datos. Estos algoritmos han sido aplicados ampliamente en áreas como el reconocimiento de escritura, la detección de rostros, categorización de texto, entre otras. El método combina el aprendizaje estadístico y la optimización convexa, combinando una máquina de soporte vectorial para clasificación y una para regresión. Contiene una base de conocimiento o ejemplos de entrenamiento que generalmente pertenecen a una de dos categorías, siendo esta una clasificación de tipo binaria [11].

El nombre se deriva del conjunto de puntos de datos o *vectores de soporte* que contienen la información. Una SVM es básicamente una máquina de aprendizaje lineal diseñada para resolver problemas de clasificación usando el principio de separación de clases. El objetivo es encontrar un hiperplano, un plano en un espacio multidimensional de separación lineal que separe dos clases de interés. El hiperplano se ubica entre clases para cumplir dos condiciones: que todos los vectores de datos que pertenezcan a una misma clase se ubiquen del mismo lado del hiperplano y maximizar la distancia entre los vectores de datos más cercanos en ambas clases [11].

La función del SVM es la de encontrar la mejor línea, plano o hiperplano que divida el grupo de datos en dos clases, utilizando un kernel (núcleo) que se compone de un conjunto de funciones matemáticas para transformar los datos de entrada en la forma deseada. En ese sentido, un hiperplano óptimo es aquel que deja el margen máximo entre las dos clases de salida, siendo el margen la distancia entre el hiperplano y el dato más cercano de cualquiera de los conjuntos.

Las máquinas SVM tiene un buen comportamiento para conjuntos de datos pequeños, pero al aumentar la cantidad de datos es más difícil encontrar una posición del hiperplano que divida adecuadamente las dos clases de datos, además del incremento de carga computacional y aumento de la complejidad en el aprendizaje conforme aumenta el número de muestras [12].

D. Redes Neuronales Artificiales (ANN)

Las redes neuronales artificiales (ANN: Artificial Neural Networks) se plantean como una alternativa computacional para la toma de decisiones dentro del dominio del Aprendizaje Automático. Esta técnica busca emular la capacidad de aprendizaje natural de los seres vivos, la cual es atribuida al sistema neuronal de su cerebro.

La unidad básica de procesamiento de una red neuronal artificial es la neurona. Las redes neuronales funcionan mediante la interacción de conjuntos de neuronas con características diferentes. Dichas características son explicadas a continuación:

- Función de propagación: cada neurona tiene una serie de entradas provenientes de otras neuronas. Estas señales de entrada son atenuadas o amplificadas por un factor de peso y son operadas en conjunto por una función de propagación que es comúnmente una suma ponderada.

- Función de activación: es la función de umbral que determina la acción de una neurona, dependiendo del valor de entrada proveniente de la función de propagación [8].

- Función de salida: es la función encargada de calcular el valor de salida de la neurona para ser transferido como entrada a otras neuronas [13].

La topología de una red neuronal está ligada con el algoritmo de aprendizaje usado para entrenar la red. Los factores que definen la topología de la red son las capas y la naturaleza de las conexiones entre las neuronas. En una red neuronal, una capa es un conjunto de nodos (pueden ser neuronas o fuentes de datos) con características similares, subdivididas es en capas de entrada, capas ocultas y capas de salida. Las conexiones entre las neuronas y las capas pueden ser unidireccionales (feedforward) o recurrentes (con al menos un lazo de realimentación o feedback).

La capacidad de aprendizaje mediante un proceso de entrenamiento y la aplicación de la experiencia adquirida en este proceso le otorga a las ANN la capacidad de responder apropiadamente a situaciones a las que no había sido expuesta

[14]. Los sistemas de aprendizaje cambian sus características para poder adaptarse al problema que se está afrontando y conseguir la generalización de la comprensión del problema. Las ANN pueden aprender mediante el desarrollo de nuevas conexiones, del cambio de la ponderación de sus conexiones, la creación de nuevas neuronas y el cambio los valores de umbral en la función de activación, entre otros.

E. Sistemas expertos y agentes inteligentes

Los sistemas inteligentes consisten en softwares de respuesta a preguntas en un determinado dominio, emulando la capacidad de toma de decisiones de un experto humano. Son utilizados generalmente para apoyar la decisión de un experto en diferentes áreas del conocimiento. El sistema contiene una base de conocimientos sobre un dominio específico y se conecta con un motor de inferencia para la derivación de las respuestas basadas en este conocimiento [15].

Los agentes inteligentes son componentes de software que poseen características de comprensión de un lenguaje, capacidad para tomar decisiones y actuar según corresponda, usualmente usados para la defensa contra ataques frecuentes. Junto con un sistema de redes neuronales, estos agentes pueden conformar un método de detección eficaz [16].

III. METODOLOGÍA

El análisis del estado del arte concerniente con la aplicación de inteligencia artificial en el análisis forense digital se hizo mediante consultas a la base de datos de Scopus de Elsevier, utilizando la siguiente sentencia de búsqueda:

TITLE-ABS-KEY ("Digital Forensics" AND "Forensic Analysis" AND ("intelligent agent" OR "Support Vector Machines" OR "Artificial Neural Networks")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "ENGI"))

En esta consulta se hizo una búsqueda todos aquellos documentos que contuvieran las frases “Forense Digital” y “Análisis Forense”, y cualquiera de las técnicas de inteligencia artificial: “Redes Neuronales Artificiales”, “Máquinas de Soportes Vectoriales” o “Agentes Inteligentes”. La búsqueda se enmarca en las ciencias de la computación y la ingeniería, con el fin de obtener aplicaciones de la inteligencia artificial dentro de la computación, tal que muestre resultados preliminares algorítmicos y de desempeño de modelos de aprendizaje. La consulta se hizo para documentos publicados a partir del año 2000 en adelante, excluyendo patentes o cualquier documento no publicado.

La temática analizada en esta investigación es de carácter novedoso y se encuentra poca literatura disponible y un número limitado de artículos que detallen los modelos, propuestas y resultados preliminares de desarrollo e implementación. Inicialmente fueron encontrados 161 documentos que involucraban la temática de forense digital, de los cuales 17 hacían énfasis en el análisis forense y el uso de técnicas de inteligencia artificial, siendo seleccionados 11 que documentaban puntualmente la información requerida: técnicas implementadas preliminarmente, tamaño de las observaciones de entrenamiento, validación y prueba de la técnica, y desempeño del modelo implementado. En la siguiente sección se detallan los resultados del análisis de referentes ejecutado.

IV. RESULTADOS DEL ANÁLISIS

Un breve estudio del estado del arte fue ejecutado con el fin de identificar las diversas aplicaciones de la Inteligencia Artificial al análisis forense digital, haciendo especial énfasis en el uso de técnicas de Aprendizaje Automático, entre las que se encuentran las Máquinas de Soportes Vectoriales y las Redes Neuronales Artificiales, y el uso de sistemas expertos y agentes inteligentes.

En la Tabla I se ha detallado el estado del arte relevante para la investigación. De esta tabla se puede observar que se han desarrollado diversos trabajos en el ámbito del análisis forense digital que incluyen: la detección de alteraciones a imágenes por medios digitales, la identificación de atributos y características físicas y químicas de objetos y la determinación del contenido de una imagen forense para su análisis automatizado, todos enmarcados en el uso de técnicas de aprendizaje automático e inteligencia artificial. A continuación, se detallan los trabajos relevantes para la investigación por temática.

A. Máquinas de Soportes Vectoriales

Algunos expertos han aplicado técnicas de aprendizaje automático en informática forense, en particular máquinas de soportes vectoriales (SVM), que se aplican generalmente al reconocimiento de escritura, detección de rostros, procesamiento del lenguaje natural y visión artificial. Las SVM pertenecen a una clase de algoritmos de Machine Learning denominados métodos kernel que se usan continuamente en la detección de intrusiones, debido a su alta velocidad de aprendizaje y escalabilidad. A continuación, se detalla el proceso y resultados de autores que han implementado estas técnicas para ejecutar un análisis forense, enmarcado en distintas áreas de estudio y con diversos objetivos.

S. Mukkamala y A. Sung [17] estudiaron en 2003 la implementación de SVM para el análisis forense de redes de datos, sobre una base de datos de 494,021 eventos procedentes del ataque de DARPA en 1999, considerando cuatro tipos de ataques: análisis de vulnerabilidades invasivo, denegación de servicios (DOS), acceso no autorizado con privilegios de súper usuario (U2Su) y acceso no autorizado desde máquina remota (R2L). El desempeño del clasificador por SVM logra una precisión ligeramente superior al 99% para las cuatro clases de ataques evaluadas.

Por otro lado, A. Mikkilineni, et. al. [18] en 2005, implementaron una clasificación multiclase utilizando SVM, para la identificación de impresoras a partir de documentos impresos, considerando el tamaño de letra, el tipo de fuente, el tipo de papel y la edad de la impresión. Se consideraron 5,000 datos de entrenamiento y prueba con 10 clases de impresoras de

TABLA I. ESTADO DEL ARTE DE INTERÉS DEL USO DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL EN ANÁLISIS FORENSE.

Método	Ref	Objetivo	Dataset de observaciones	Desempeño
Máquinas de Soportes Vectoriales	[17]	Análisis forense de redes de datos considerando cuatro tipos de ataques: análisis de vulnerabilidades invasivo, denegación de servicios (DOS), acceso no autorizado con privilegios de súper usuario (U2Su) y acceso no autorizado desde máquina remota (R2L).	Entrenamiento (50%): 247,010 Prueba (50%): 247,010 Total dataset: 494,021	99.5% de precisión
	[18]	Identificación de impresoras a partir de documentos impresos, considerando el tamaño de letra, el tipo de fuente, el tipo de papel y la edad de la impresión.	Entrenamiento (50%): 5,000 Prueba (50%): 5,000 Total dataset: 10,000	84% al 93% de precisión
	[19]	Determinación del tipo de archivos contenidos en el material probatorio de una investigación.	Entrenamiento (90%): 3,240 Prueba (10%): 360 Total dataset: 3,600	81% al 98% de precisión
	[20]	Desarrollo de un modelo de aprendizaje para la identificación de modificaciones de filtrado y re-muestreo en imágenes digitales.	Entrenamiento (25%): 1,250 Prueba (75%): 3,250 Total dataset: 5,000	99.35% de precisión
	[21]	Desarrollo de un modelo para la clasificación de bloques de archivos con formatos conocidos.	Entrenamiento (20%): 50,400 Prueba (80%): 201,600 Total dataset: 252,000	85% al 95% de precisión
Redes Neuronales Artificiales	[23]	Identificación de clases de vidrio con base en atributos químicos (índice de refracción, cantidad de sodio, magnesio, aluminio, silicio, potasio, calcio, bario, hierro y tipo de vidrio) de muestras encontradas en las escenas donde ocurrieron los eventos criminales.	Entrenamiento (80%): 192 Prueba (20%): 38 Total dataset: 240	73% al 85% de precisión
	[24]	Identificación de atributos espaciales en imágenes con esteganografía LSB con cinco densidades de contenido embebido (payload): 0.1, 0.2, 0.3, 0.4 y 0.5 bpp (bits por píxel).	Entrenamiento (70%): 7,000 Prueba (30%): 3,000 Total dataset: 10,000	84% al 86% de precisión
	[25]	Desarrollo de un modelo para la detección de posible manipulación de imágenes digitales a color.	Entrenamiento (80%): 160,000 Prueba (20%): 40,000 Total dataset: 200,000	95% de precisión
	[26]	Análisis forense de imágenes JPEG para notar transformaciones espaciales.	Entrenamiento (65%): 284,000 Prueba (35%): 150,000 Total dataset: 434,000	84% al 99% de precisión
	[27]	Desarrollo de un algoritmo para la detección y localización de falsificaciones de imágenes mediante funciones de remuestreo y aprendizaje profundo.	Entrenamiento (80%): 100,000 Prueba (20%): 25,000 Total dataset: 125,000	95% de precisión
Sistemas expertos y agentes inteligentes	[28]	Desarrollo de una propuesta de software MADIK para la investigación digital multi agente que permita suplir la dificultad que tiene un experto de determinar de forma rápida qué evidencia es relevante cuando se analiza un crimen.		De 69% a 74% de cubrimiento

salida para la clasificación, obteniendo un desempeño en la clasificación de entre el 84% y el 93% de precisión.

Con el fin de determinar el tipo de archivos contenidos en el material probatorio de una investigación, Q. Li y A. Ong [19] en 2010 clasificaron los fragmentos de archivos utilizando SVM, analizando cinco clases de salida: archivos JPEG (800 ejemplos), archivos MP3 (800 ejemplos), archivos PDF (800 ejemplos), archivos DLL (800 ejemplos) y archivos binarios ejecutables de Windows (400 ejemplos). La distribución entre entrenamiento y prueba fue del 90% y 10%, respectivamente, obteniendo un desempeño del clasificador medido en precisión en el intervalo de 81% a 98%.

Identificar modificaciones ocultas en imágenes o develar alteraciones en el formato de los archivos son otros campos trabajados en el análisis forense digital, pues puede existir contenido embebido no visible en las imágenes que pueda ser de utilidad, o los archivos pueden estar particionados para evitar mostrar su contenido al abrirlos, haciendo que parezcan dañados.

En este primer campo, Cai, et. al. [20] desarrollaron un modelo de SVM que permite identificar si una imagen ha sido alterada por medios digitales, incluyendo operaciones como submuestreo, compresión, filtrado de mediana y filtrado de media. Esto, a partir de una base de datos de 1,000 imágenes a color por cada operación de modificación, divididas en 25% para entrenamiento y 75% para pruebas del modelo. El desempeño del clasificador multiclase fue medido en precisión y alcanzó el 99.35%.

De forma complementaria, Sportiello y Zanero [21] desarrollaron un modelo de SVM que, mediante descriptores frecuenciales, permite clasificar los bloques de archivos que han sido separados o malformados, haciendo énfasis en los formatos: bmp, doc, exe, gif, jpg, mp3, odt, pdf y ppt. Esto, a partir de una base de datos heterogénea de archivos en estos formatos que fueron seccionados en bloques de 512 bytes para un total de 252,000 bloques que fueron analizados para el entrenamiento (20% de las observaciones) y la prueba (80% de las observaciones) del modelo. El desempeño final del mismo medido en precisión alcanza un máximo de 95% con un mínimo de clasificación de 85%.

B. Redes Neuronales Artificiales

Una segunda técnica utilizada son las redes neuronales artificiales, consisten en una colección de elementos que están interconectados y se transforman en un conjunto de salidas deseadas. La red neuronal realiza un análisis de la información y proporciona una estimación de probabilidad que coincide con los datos para los cuales ha sido entrenada para reconocer. La

red neuronal va obteniendo eficacia al irse entrenando el sistema con la entrada y la salida del problema deseado. A partir de ese entrenamiento se obtienen mejores resultados y la configuración de la red se refina conforme pasa el tiempo [17].

R. Mohammad [22] ha demostrado que es posible identificar si el sistema operativo ha sufrido modificaciones no autorizadas a partir de la implementación de un modelo de red neuronal.

A. Tallón-Ballesteros y J. Riquelme [23] en 2014, por otro lado, han aplicado métodos de clasificación multiclase usando redes bayesianas, árboles de decisión y redes neuronales artificiales, para la identificación de clases de vidrio con base en atributos químicos (índice de refracción, cantidad de sodio, magnesio, aluminio, silicio, potasio, calcio, bario, hierro y tipo de vidrio) de muestras encontradas en las escenas donde ocurrieron los eventos criminales, a partir de una base de datos del USA Forensic Science Service con 240 observaciones. Los algoritmos implementados fueron validados haciendo uso de validaciones cruzadas de 4 folds, es decir, subdividiendo el conjunto de datos de prueba en cuatro subconjuntos homogéneos y ejecutando las pruebas pertinentes sobre los datos. Los autores obtienen un desempeño en el mejor clasificador (el de redes neuronales) de entre 73% y 75%.

Qian, et. al. [24] en 2016, desarrollaron un modelo de aprendizaje basado en Redes Neuronales Convolucionales (CNN) para la identificación de atributos espaciales en imágenes con esteganografía LSB, a partir de imágenes procedentes de la base de datos BOSSbase (Break Our Steganography System) v1.01, una base de datos de imágenes a escala de grises diseñada para ejecutar pruebas de estegoanálisis. Esta base de datos contiene 10,000 imágenes portadoras y con esteganografía con una distribución del 50% y con resolución espacial de 512x512. El 70% de este conjunto de datos fue asignado para entrenamiento del modelo, el 10% para validación y el 20% para pruebas. Se tuvieron en cuenta cinco densidades de contenido embebido (payload): 0.1, 0.2, 0.3, 0.4 y 0.5 bpp (bits por píxel). El clasificador logró un desempeño medido en precisión de entre el 84% y 86%.

D. Kim y H. Lee [25] en 2017 han utilizado un enfoque de redes neuronales convolucionales, un tipo especializado de redes neuronales que permite aprender de una mayor cantidad de información, almacenando conocimiento de los entrenamientos. Esta característica les permite a los autores desarrollar un modelo para la detección de posible manipulación de imágenes digitales a color. El modelo fue entrenado con ejemplos gráficos procedentes de una base de datos estandarizada con dimensión espacial 256x256, que contienen cuatro tipos de ruido, procesando las imágenes con filtros de mediana y Gaussianos. El modelo fue entrenado con 160,000 imágenes (80% del dataset) y probado con 40,000 imágenes (20% del dataset), consiguiendo un desempeño del detector de 95% de precisión.

De forma similar, N. Bonettini [26] en 2017 ha usado un enfoque de redes neuronales convolucionales para el análisis forense de imágenes JPEG, teniendo en cuenta una base de datos RAISE de imágenes RAW de 284,000 ejemplos que fue transformada en imágenes a escala de grises que fueron comprimidas, con resoluciones espaciales variadas equitativamente desde 64x64 hasta 256x256. Para el entrenamiento, el 70% del dataset original fue tenido en cuenta, con un tamaño de batch de 128 (cantidad de ejemplos por subconjunto de entrenamiento) y 30 epochs de entrenamiento (cantidad de iteraciones por unidad de entrenamiento). La validación del algoritmo fue ejecutada con el 30% del dataset y la prueba con 150,000 imágenes independientes del conjunto de entrenamiento y validación. El desempeño del clasificador está entre el 84% y 99% de precisión.

J. Bunk et. al. [27] en 2017, desarrollaron un algoritmo para la detección y localización de falsificaciones de imágenes mediante funciones de remuestreo y aprendizaje profundo, usando redes neuronales LSTM, una tipología de redes neuronales especializada en almacenar memoria de corto y largo plazo en cada iteración de entrenamiento, haciendo que sea más rápido el aprendizaje con una base de conocimiento existente. Para esto, tuvieron en cuenta un dataset de 100,000 ejemplos de las bases de datos de UCID y RAISE, y para la prueba la base de datos de NIST Nimble 2016 con tres clases de modificaciones en las imágenes: copiado y clonado, remoción y sobreposición. Los autores obtienen un 95% de precisión con el modelo de LSTM ANN en la detección de falsificaciones.

C. Sistemas expertos y agentes inteligentes

Una propuesta de software presentada en el escenario internacional es la de MADIK, un Toolkit para la investigación digital multi agente (de sus siglas en inglés Multi-Agent Digital Investigation ToolKit) elaborado por B. Hoelz, et. al. [28] en 2008. Esta es una apuesta que se ha sido desarrollada con el fin de suplir la dificultad que tiene un experto de determinar de forma rápida qué evidencia es relevante cuando se analiza un crimen. Ante la falta de herramientas que colaboren en el preanálisis de las evidencias y su correlación, nace este marco de trabajo basado en un conjunto de herramientas usando un sistema de multi agente. Es implementado usando JADE, marco de trabajo basado en lenguaje Java muy común en el desarrollo de sistemas multi agente.

La arquitectura definida (ver Fig. 2) se divide en roles distribuidos en cuatro niveles de agentes. Se definen agentes autónomos especializados cada uno con distintos objetivos, que pueden colaborar con el trabajo de otros agentes en un único espacio denominado *Blackboard*. Un sistema multi agente permite un uso de los recursos más eficiente y a los agentes operar de forma autónoma en diferentes máquinas y entornos.

La arquitectura de cuatro capas incluye un nivel operativo donde encontramos a los agentes especializados, y unos niveles superiores en donde se encuentran los agentes de gestión, encargados de las decisiones estratégicas y tácticas. Estos últimos están encargados a su vez de la distribución y coordinación de las tareas que ejecutarán los agentes especializados de la capa operacional. La comunicación es jerárquica, ya que los agentes especializados se comunican solo con el gerente operacional, el gerente operacional con el gerente táctico y el gerente táctico con el gerente estratégico.

La plataforma se puede distribuir entre máquinas diferentes y la configuración se realiza desde una interfaz GUI remota. Dentro de las opciones principales que permite la configuración

están las de definir tiempos donde se ejecutan procesos y mover a los agentes de una máquina a otra.

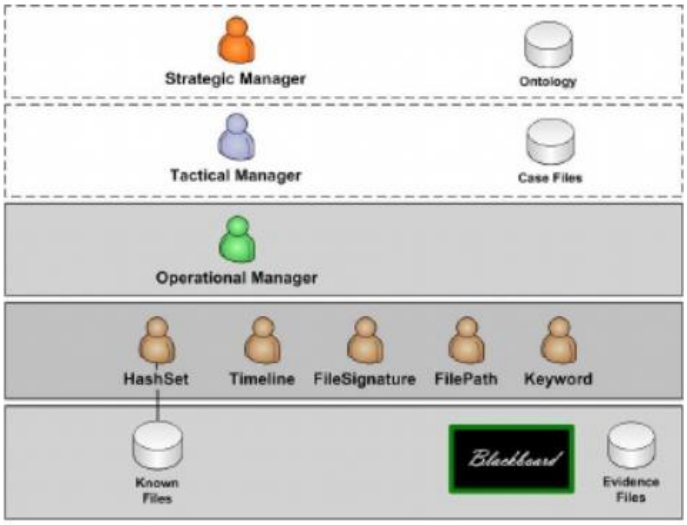

Fig. 2. Arquitectura MADIK. Tomado de: [28].

En el nivel operativo se encuentra la mayor cantidad de agentes desplegados, entre los que se encuentran:

- HashSetAgent: calcula el hash MD5 de un archivo y lo compara con su base de conocimiento.
- FilePathAgent: mantiene en su base de conocimiento carpetas que son usadas habitualmente por aplicaciones que pueden ser de interés para la investigación.
- FileSignatureAgent: analiza el encabezado del archivo para determinar su extensión. Usado para determinar si el atacante cambia la extensión de los archivos.
- TimelineAgent: Examina fecha de creación, de acceso y modificación de los archivos.

Después de varios experimentos realizados con las herramientas en múltiples casos, los autores pueden determinar buenos resultados al compararlos con las herramientas disponibles evaluadas. En comparación con el trabajo realizado por un humano experto, los autores observaron un menor factor de reducción para el sistema de múltiples agentes, aunque en términos del tiempo requerido observaron una reducción de tiempo hasta por seis veces, considerando el mismo contenido examinado por parte humana y parte artificial. El factor de cobertura alcanzó el 80%, una buena cifra, según los autores, que se puede mejorar con bases de conocimientos robustas y con el uso de más agentes especializados.

D. Otras aplicaciones

La Inteligencia Artificial ha sido usada en otros campos en los que se puede extrapolar un estudio forense, como lo son: el reconocimiento de patrones en línea [29] [30], el descubrimiento de conocimiento [31], la computación forense sobre Internet de las Cosas (IoT) [32], la correlación de eventos almacenados en los elementos materiales probatorios [33], entre otras aplicaciones.

El reconocimiento de patrones es un campo particularmente utilizado en el análisis forense digital de elementos materiales probatorios, pues permite identificar grupos de datos y dinámicas no visibles. Es ampliamente utilizado en informática forense para el reconocimiento de patrones en imágenes, tal como se ha precisado, en el que una sección de software realiza la identificación de componentes espaciales y frecuenciales de una imagen [34] [35]. Otra forma de reconocimiento de patrones se aplica en mensajes de correo electrónico para encontrar aquellos con SPAM o que utilicen phishing [36] [37]. También se puede utilizar para reconocimiento de patrones de audio en pistas o sectores de disco [38].

Un tercer campo de la AI que se puede aprovechar en el ámbito forense es la minería de datos y descubrimiento de conocimiento en bases de datos. Aunque ambos conceptos son diferentes, se usan para referirse al mismo procedimiento de recolección de grandes cantidades de datos [31].

V. DISCUSIÓN

Las técnicas actuales de análisis forense digital están basadas en la identificación de variaciones de variables individuales que permiten una identificación de posibles actividades irregulares mediante métodos manuales y visuales. Sin embargo, el exceso en el volumen de datos que deben ser analizados, los patrones que no son identificables al aplicar técnicas manuales y visuales, y las diversas dinámicas de texto, gráficas y audiovisuales que deben ser analizadas, hacen que la complejidad temporal y espacial de los algoritmos crezca, aumentando el tiempo requerido para ejecutar el análisis forense y limitando las capacidades computacionales en la búsqueda de información concluyente. Para esto se han desarrollado técnicas de inteligencia artificial que permiten entrenar un modelo que identifique de forma correlacionada e integrada la actividad irregular de múltiples variables, archivos y medios de forma simultánea, entreviendo patrones en dinámicas ocultas que no son visualmente apreciables.

El uso de técnicas de inteligencia artificial se encuentra en constante expansión en la actualidad, debido a su eficiencia para afrontar problemas complejos en diferentes áreas de estudio, sumado a que las herramientas computacionales modernas permiten una implementación viable de esta técnica. En el campo del análisis forense digital, tal como se evidenció en los resultados del análisis, las técnicas de inteligencia artificial permitieron un entrenamiento de modelos de aprendizaje a partir de la experiencia y la identificación de patrones en los posibles vectores de ataque y alteración de la información, evidenciando las siguientes ventajas:

- Capacidad de realizar tareas con base en criterios adquiridos a partir de un entrenamiento.
- Abstracción de una representación de la información y organización de acuerdo con esta.
- Tolerancia a fallos y composición modular de las técnicas de aprendizaje que permiten que los servicios, a pesar de sufrir daños parciales, tengan capacidades que se puedan conservar.

- Las operaciones y cálculos pueden hacerse en paralelo y tiempo real, lo que permite la identificación en línea de patrones en vectores de ataque.

Por otro lado, se evidencia que la inteligencia artificial es aplicable a múltiples contextos dentro del análisis forense digital, lo que permite reforzar esta disciplina. Desde la identificación de aspectos físicos mediante el procesamiento de imágenes de las escenas, hasta la detección de alteraciones en archivos y datos. Esto demuestra que la inteligencia artificial es un área flexible y heterogénea en técnicas que se pueden aplicar a variados contextos, que está en constante evolución y que ha colaborado con la solución de algunos inconvenientes que se presentan a la hora de realizar tareas humanas.

Sin embargo, aunque las técnicas de inteligencia artificial presentadas resultan eficientes para el análisis y correlación de grandes volúmenes de datos, siguen presentando una tasa de falsos positivos y falsos negativos alta y su implementación implica un alto consumo de recursos computacionales. Adicionalmente, se trata de técnicas basadas en la experiencia, razón por la cual ataques de días cero o vectores de ataque no previstos pueden no ser detectados correctamente al analizar los datos recabados. Por lo tanto, debe hacerse un esfuerzo adicional por seguir implementando la inteligencia artificial de forma integrada con el análisis manual y visual de los datos, pues la base de conocimientos es fundamental para un mejor desempeño de los modelos de detección implementados.

Aunque este estudio presenta reflexiones sobre la problemática planteada, existen algunas limitaciones a destacar: poca literatura que documente el desarrollo y resultados preliminares de pruebas de aplicación de la inteligencia artificial en el ámbito forense digital. Así mismo, se incluyeron técnicas específicas usadas para la solución de problemáticas enmarcadas en el análisis forense digital, obviando el uso de otras técnicas como los modelos de Markov, estadísticas Bayesianas y modelos de redes neuronales más robustos, entre otros. El análisis del desarrollo e implementación de técnicas de inteligencia artificial se hace de forma limitada, sin entrar en detalle en áreas específicas del análisis forense digital, como el almacenamiento en la nube, operaciones transaccionales y bancarias, o procesos cuánticos criptográficos, entre otros.

VI. CONCLUSIONES

En este documento se detallan las características y propiedades de los modelos de inteligencia artificial que son aplicados al análisis forense digital. Se presenta una breve revisión de la literatura para ilustrar en qué áreas del análisis forense digital se ha utilizado recientemente la inteligencia artificial, proporcionando una visión de las múltiples disciplinas de trabajo del aprendizaje automático, las máquinas de soportes vectoriales y las redes neuronales artificiales, así como los agentes inteligentes.

Muchos son los retos que aún debe resolverse con el uso de técnicas automatizadas que permitan la identificación de patrones nacientes, ataques e información extraíble de dinámicas ocultas no visibles mediante la implementación de métodos manuales.

En este sentido, este breve estudio establece un punto de referencia para habilitar nuevas perspectivas del uso de la inteligencia artificial en el análisis forense digital, que permitan orientar investigaciones posteriores y propuestas novedosas de aplicaciones emergentes, que atendiendo las ventajas y limitaciones de las disciplinas de aprendizaje automático y los agentes inteligentes revisadas, puedan facilitar la convergencia entre los métodos manuales y las posibilidades que plantean las estrategias automáticas revisadas.

AGRADECIMIENTOS

Esta investigación estuvo soportada por la Universidad Pontificia Bolivariana de Bucaramanga. Gracias al Maestro Diego Javier Parada, Coordinador del Programa de Especialización en Seguridad Informática, por el apoyo y colaboración en el desarrollo de esta investigación.

REFERENCIAS

[1] M. López Delgado, Análisis Forense Digital, España: CriptoRed, 2007.

[2] S. Raghavan, «Digital forensic research: Current state of the art,» *CSI Transactions,* pp. 91-114, 2013.

[3] M. J. Rivas Sández, «A Review of Technical Problems when Conducting an Investigation in Cloud-Based Environments,» *arXiv,* p. 16, 2014.

[4] H. Ç. a. M. A. S. Dilek, «Applications of Artificial Intelligence Techniques to combating Cyber Crimes: a review,» *International Journal of Artificial Intelligence & Applications (IJAIA),* vol. 6, pp. 21-23, 2015.

[5] L. D. Merkle, «Automated Network Forensics,» pp. 1929-1931, 2008.

[6] . A. Kaplan y M. Haenlein, «Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations,» *Illustrations and Implications of Artificial Intelligence, Business Horizons,* vol. 62, nº 1, pp. 15-25, 2019.

[7] D. Poole, Computational Intelligence: A Logical Approach, Nueva York: Oxford University Press, 2018.

[8] N. Nilsson, Introduction to Machine Learning, Standford, California: Department of Computer Science, Standford University, 2005.

[9] P. Harrington, Machine Learning in Action, United States of America: Manning publications Co,, 2012, p. 382.

[10] D. Kriesel, A Brief Introduction to Neural Networks, Bonn, Germany: Dkriesel, 2005, p. 244.

[11] A. Mammone, M. Turchi y N. Cristianini, «Support Vector Machines,» vol. 1, pp. 283-288, 2009.

[12] A. Navia Vazquez y E. Parrado Hernandez, «Support vector machine interpretation,» *Neurocomputing,* pp. 1754-1759, 2006.

[13] D. Matich, Redes Neuronales: Conceptos Básicos y Aplicaciones, Rosario: Universidad Tecnológica Nacional, 2001, p. 55.

[14] C. A. Ruiz y M. S. Basualdo, Redes Neuronales: conceptos básicos y aplicaciones, Rosario: Universidad Tecnológica Nacional, Facultad Regional Rosario, 2001, p. 55.

[15] C. S. Krishnamoorthy y S. Rajeev, Artificial Intelligence and Expert Systems for Engineers, Boca Ratón, Florida: CRC Press, 2000, p. 190.

[16] L. C. Jain, Z. Chen y N. Ichalkaranje, Intelligent Agents and Their Applications, New York: Physica-Verlag, Springer, 2002.

[17] S. M. a. A. H. Sung, «Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligent Techniques,» *International Journal of Digital Evidence,* vol. 1, nº 4, 2003.

[18] A. Mikkilineni, O. Arslan, P.-J. Chiang, R. Kumontoy, J. Allebach, G. Chiu y E. Delp, *Purdue University,* p. 4, 2005.

[19] Q. Li y A. Ong, «A Novel Support Vector Machine Approach to High Entropy Data Fragment Classification,» *Proceedings of the South African Information Security Multi-Conference,* p. 10, 2010.

[20] K. Cai, X. Lu, J. Song y X. Wang, «Blind Image Tampering Identification Based on Histogram Features,» de *Third International Conference on Multimedia Information Networking and Security*, Shanghai, China, 2011.

[21] L. Sportiello y S. Zanero, «File Block Classification by Support Vector Machines,» de *Sixth International Conference on Availability, Reliability, and Security*, Milan, Italy, 2011.

[22] R. Mohammad, «A Neural Network-based Digital Forensics Classification,» *IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA),* p. 7, 2018.

[23] A. Tallón-Ballesteros y J. Riquelme, «Data Mining Methods Applied to a Digital Forensics Task for Supervised Machine Learning,» *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications,* vol. 555, pp. 413-428.

[24] Y. Qian, J. Dong, W. Wang y T. Tan, «Learning and Transferring Representations for Image Steganalysis using Convolutional Neural Network,» de *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.

[25] D.-H. Kim y H.-Y. Lee, «Image Manipulation Detection using Convolutional Neural Network,» *International Journal of Applied Engineering Research,* vol. 12, nº 21, pp. 11640-11646, 2017.

[26] N. Bonettini, «JPEG-based Forensics through Convolutional Neural Networks,» Milano, 2017.

[27] J. Bunk, J. Bappy, T. Mohammed, L. Nataraj, A. Flenner, B. Manjunath, S. Chandrasekaran, A. Roy-Chowdhury y L. Peterson, «Detection and Localization of Image Forgeries using Resampling Features and Deep Learning,» de *CVPR Workshop on Media Forensics*, University of Maryland, 2017.

[28] B. Hoelz, C. Ralha, R. Geeverghese y H. Junior, «A Cooperative Multi-Agent Approach to Computer Forensics,» *IEEE International Conference on Web Intelligence and Intelligent Agent Technology,* pp. 1-7, 2008.

[29] S. Qatawneh, S. Ipson, R. Qahwaji y H. Ugail, «3D face recognition based on machine learning,» *Eighth IASTED International Conference on Visualization, Imaging and Image Processing (VIIP 2008),* pp. 362-366, 2008.

[30] D. Saez Trigueros, L. Meng y M. Hartnett, «Face Recognition: From Traditional to Deep Learning Methods,» *Cornell University Press,* p. 13, 2018.

[31] F. Mitchell, «THE USE OF ARTIFICIAL INTELLIGENCE IN DIGITAL FORENSICS: AN INTRODUCTION,» *Digital Evidence and Electronic Signature Law Review,* vol. 7, pp. 35-41, 2010.

[32] F. Spencer, «Digital Forensics with Artificial Intelligence Internet of Things,» p. 6, 2018.

[33] B. Hoelz, R. Geeverghese y C. G. Ralha, «Artificial intelligence applied to computer forensics,» *Proceedings of the 2009 ACM Symposium on Applied Computing (SAC),* p. 6, 2009.

[34] J. Li, S. You y A. Robles-Kelly, «A Frequency Domain Neural Network for Fast Image Super-resolution,» *International Joint Conference on Neural Networks,* p. 9, 2017.

[35] F. Petroski Such, S. Sah, M. Dominguez, S. Pillai, C. Zhang, A. Michael, N. Cahill y R. Ptucha, «Robust Spatial Filtering with Graph Convolutional Neural Networks,» *IEEE Journal of Selected Topics in Signal Processing,* p. 14, 2017.

[36] L. Özgür, T. Gungor y F. Gurgen, «Spam Mail Detection Using Artificial Neural Network and Bayesian Filter,» *Intelligent Data Engineering and Automated Learning - IDEAL, 5th International Conference,* p. 6, 2004.

[37] S. Sekhar Roy y M. Viswanatham, «Classifying Spam Emails Using Artificial Intelligent Techniques,» *International Journal of Advanced Computer Technology (IJACT),* p. 5, 2016.

[38] K. Al Smadi, H. A. Al Issa, I. Trrad y T. Al Smadi, «Artificial Intelligence for Speech Recognition Based on Neural Networks,» *Journal of Signal and Information Processing,* p. 7, 2006.

Jeimy J. Cano. Egresado del Programa de Ingeniería y Maestría en Sistemas y Computación de la Universidad de Los Andes. Doctor en Filosofía de la Administración de Negocios, de Newport University en California, Estados Unidos. Certificado como Examinador Certificado de Fraude - en inglés CFE. Es profesor e investigador a nivel nacional y latinoamericano en temas de seguridad informática, computación forense y sistemas de información. Actualmente, es director de la revista SISTEMAS, de la Asociación Colombiana de Ingenieros de Sistemas (ACIS).

Julián D. Miranda. Ingeniero Electrónico (2016), Ingeniero de Sistemas e Informática (2018) y Especialista en Seguridad Informática (2019) de la Universidad Pontificia Bolivariana de Bucaramanga, Colombia. Cuenta con experiencia en el desarrollo de proyectos de investigación con metodologías Ágiles, aplicando aprendizaje automático y procesamiento digital de imágenes y señales, enfocadas hacia la Inteligencia Artificial y las ciencias de datos. Actualmente, es docente de pregrado y posgrado en las áreas de computación, sistemas operativos, criptografía, esteganografía y ciencia de datos.

Sergio A. Pinzón. Ingeniero de Telecomunicaciones (2013) de la Universidad Santo Tomas de Aquino de Bucaramanga, Colombia. Cuenta con experiencia en Arquitectura de redes móviles, redes cableadas de fibra óptica y cableado UTP, configuración de dispositivos de comunicaciones y seguridad perimetral, con intereses en temas de Machine Learning aplicado a la seguridad de la información, automatización de dispositivos de red y Ethical Hacking.