DETECCIÓN DE ATAQUES DE MALWARE OBFUSCATED MEDIANTE UNA RED NEURONAL FEEDFORWARD INCORPORANDO UN MECANISMO DE ATENCIÓN TIPO SQUEEZE-AND-EXCITATION (SE)

Obfuscated Malware attack detection using a feedforward neural network incorporating a Squeeze-and-Excitation (SE) attention mechanism

VÍCTOR ALFONSO GUZMÁN-BRAND[1], LAURA ESPERANZA GELVEZ-GARCÍA[2]

Recibido: 22 de abril de 2025. Aceptado:17 de junio de 2025.
DOI: https://doi.org/10.21017/rimci.1155

Resumen

Objetivo: detectar ataques de Malware Obfuscated mediante una red neuronal feedforward incorporando un mecanismo de atención tipo Squeeze-and-Excitation (SE).

Metodología: se emplea la metodología *Knowledge Discovery in Databases* (KDD) como marco analítico para la minería de datos, esta metodología permite examinar grandes volúmenes de información, identificar patrones y convertirlos en conocimiento para la toma de decisiones. Tomando para la experimentación el conjunto de datos Malware Memory Analysis CIC-MalMem-2022, publicado por el Instituto Canadiense de Ciberseguridad.

Resultados: los modelos de redes neuronales artificiales han demostrado un desempeño sólido en tareas de clasificación evaluado mediante la matriz de confusión. Arquitecturas como las redes neuronales recurrentes (RNN) y las unidades de puerta recurrente (GRU) presentan ventajas significativas. No obstante, el modelo de red neuronal feedforward con mecanismo de atención Squeeze-and-Excitation (RNF+SE) destaca por su precisión y por la reducción de falsos positivos y negativos.

Discusión: al aplicar las métricas generales de evaluación la red neuronal con mecanismo de atención Squeeze-and-Excitation (RNF+SE) sobresale con una exactitud, F1-score y ROC-AUC destacables, demostrando ser el modelo preciso y robusto.

Conclusiones: el estudio demuestra que la integración del mecanismo de atención Squeeze-and-Excitation (SE) en una red neuronal profunda (RNF+SE) mejora significativamente la detección de malware ofuscado, superando modelos tradicionales en precisión. A pesar de su mayor costo computacional, su capacidad para recalibrar dinámicamente la importancia de las características lo posiciona como una alternativa eficiente para entornos de ciberseguridad.

Palabras Claves: detección; malware obfuscated; red neuronal feedforward; mecanismo de atención; squeeze-and-excitation (SE).

Abstract

Objective: to detect Obfuscated Malware attacks by means of a feedforward neural network incorporating a Squeeze-and-Excitation (SE) type attention mechanism.

1 Profesional en Psicología. Especialista en Desarrollo Integral de la Infancia y Adolescencia. Especialista en analítica de datos. Investigador Junior (Colciencias). ORCID: https://orcid.org/0000-0002-6051-3153 Correo electrónico: victora.guzman@cun.edu.co

2 Licenciada en Lengua Castellana. Magíster en Lingüística Española. Doctora en Ciencias de la Educación. ORCID: https://orcid.org/0000-0003-0164-2972 Correo electrónico: laura_garcia@cun.edu.co

Methodology: the Knowledge Discovery in Databases (KDD) methodology is used as an analytical framework for data mining. This methodology allows examining large volumes of information, identifying patterns and converting them into knowledge for decision making. Taking for experimentation the Malware Memory Analysis CIC-MalMem-2022 dataset, published by the Canadian Cybersecurity Institute.

Results: Artificial neural network models have demonstrated robust performance in classification tasks evaluated using the confusion matrix. Architectures such as recurrent neural networks (RNN) and recurrent gating units (GRU) have significant advantages. However, the feedforward neural network model with Squeeze-and-Excitation (RNF+SE) attention mechanism stands out for its accuracy and reduction of false positives and negatives.

Discussion: when applying the general evaluation metrics, the neural network with Squeeze-and-Excitation attention mechanism (RNF+SE) stands out with outstanding accuracy, F1-score and ROC-AUC, proving to be the accurate and robust model.

Conclusions: The study demonstrates that the integration of the Squeeze-and-Excitation (SE) attention mechanism in a deep neural network (RNF+SE) significantly improves the detection of obfuscated malware, surpassing traditional models in accuracy. Despite its higher computational cost, its ability to dynamically recalibrate feature importance positions it as an efficient alternative for cybersecurity environments.

Keywords: Detection; Obfuscated Malware; feedforward neural network; attention mechanism; Squeeze-and-Excitation (SE).

I. Introducción

El crecimiento a gran escala de las redes de computadoras, junto con la expansión de los servicios que estas ofrecen, ha generado una demanda cada vez mayor por garantizar la confiabilidad, integridad y disponibilidad de la información transmitida. Este panorama ha situado la seguridad de los sistemas de cómputo como una prioridad[1]. Paralelamente, el incremento constante de los ataques dirigidos a estos sistemas ha intensificado los desafíos asociados a la protección de la información, convirtiéndose en una problemática de gran relevancia en el ámbito tecnológico y de ciberseguridad[2].

El malware ofuscado representa una amenaza cibernética altamente sofisticada, caracterizada por el empleo de avanzadas técnicas de evasión diseñadas para ocultar su presencia. Esta capacidad le permite eludir los mecanismos de detección tradicionales, lo que dificulta significativamente su identificación mediante métodos de seguridad convencionales[3]. El malware es un tipo de software malicioso diseñado para infiltrarse en un sistema informático con el propósito de dañar, robar información o comprometer la seguridad del dispositivo. Una vez que infecta un equipo, puede propagarse a otros dispositivos conectados, amplificando su impacto[4].

La detección de malware ofuscado mediante minería de datos se ha convertido en un campo de investigación en ciberseguridad de creciente relevancia. Al aplicar técnicas avanzadas de análisis de datos junto con algoritmos de aprendizaje automático, es posible identificar patrones ocultos y anomalías en el tráfico de red y la ejecución de programas[5]. Este enfoque basado en Inteligencia Artificial (IA) permite desentrañar las estrategias de evasión utilizadas por el malware y detectar cambios sutiles en el entorno digital[6].

Entre los antecedentes relevantes se encuentran estudios especializados en la identificación de malware ofuscado mediante técnicas de aprendizaje automático. Empleando redes neuronales convolucionales (CNN) en combinación con arquitecturas de propagación hacia adelante, logrando una precisión del 93 % en tareas de clasificación[7]. De igual manera, se aplica redes neuronales basadas en similitudes alcanzando la tasa de error del 8%[8]. Por otro lado, la detección y clasificación de malware mediante el uso de XGBoost, LightGBM y Random Forest[9]. Entre estos modelos, LightGBM obtuvo el mejor desempeño, alcanzando una precisión del 94.1% y un F1-score del 93.9%[10].

El objetivo de esta investigación es desarrollar un modelo para detectar ataques de Malware Obfuscated mediante una red neuronal feedforward incorporando un mecanismo de atención tipo Squeeze-and-Excitation (SE). Además, se compara el desempeño de esta arquitectura con otras redes neuronales artificiales para evaluar su eficacia en la identificación de amenazas. Para el entrenamiento y validación del modelo, se utiliza el conjunto de datos Malware Memory Analysis CIC-MalMem-2022, publicado por el Instituto Canadiense de Ciberseguridad.

II. Metodología

En este proyecto, se emplea la metodología *Knowledge Discovery in Databases* (KDD) como marco analítico para la minería de datos en entornos de Big Data. Esta metodología permite examinar grandes volúmenes de información, identificar patrones ocultos y convertirlos en conocimiento significativo para la toma de decisiones[11]. El proceso de minería de datos basado en KDD sigue una secuencia estructurada de cinco etapas[12]:

- Selección: donde se determinan y extraen los datos relevantes para el análisis.
- Preprocesamiento: que involucra la limpieza, transformación y normalización de los datos para mejorar su calidad y reducir sesgos.
- Minería de datos: en la que se aplican algoritmos de aprendizaje automático y técnicas estadísticas para identificar patrones y relaciones.
- Evaluación: donde se analizan los resultados obtenidos para determinar su validez y relevancia.
- Interpretación: en la que se traducen los hallazgos en conocimiento útil y aplicable.

Este ciclo iterativo permite optimizar continuamente los resultados a través de la retroalimentación, asegurando que el conocimiento extraído sea preciso y relevante[13]. La aplicación de la metodología KDD es esencial en la investigación científica, y también en diversos campos como el análisis empresarial, la ciberseguridad y la medicina, donde la toma de decisiones basada en datos es clave para la innovación y el desarrollo estratégico[14].

Etapa uno: selección de los datos

En esta etapa se selecciona el data set Malware Memory Analysis CIC-MalMem-2022 publicado por el Instituto Canadiense de Ciberseguridad, este ha sido desarrollado con el propósito de evaluar métodos de detección basados en el análisis de memoria. Su diseño busca representar escenarios realistas al incluir muestras de malware predominantes en entornos reales[15] (Tabla I).

Este conjunto de datos contiene 58.596 registros, de los cuales el 50% son tráfico normal y el otro 50% se refieren a ataques. De igual manera, contiene diversas categorías de software malicioso, como spyware, ransomware y troyanos, proporcionando un equilibrio adecuado para la evaluación de sistemas de detección de malware ofuscado[2]. Los cuales se clasifican y se distribuyen de la siguiente forma (Fig. 1).

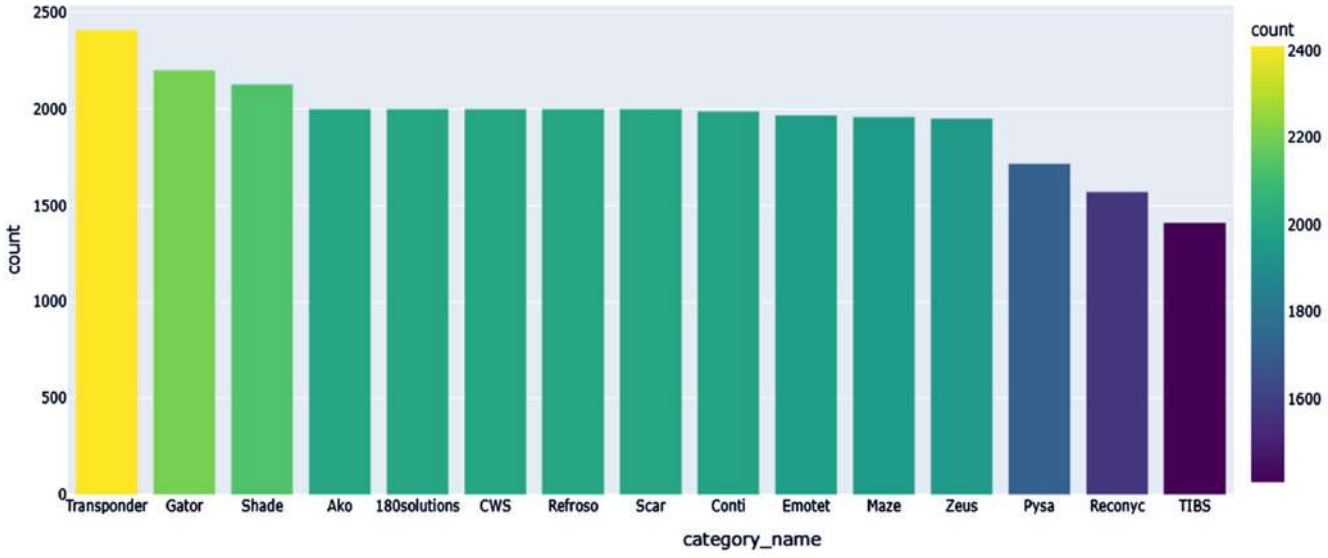
Etapa dos: preprocesamiento

En esta fase, se llevan a cabo diversas técnicas de limpieza y transformación de los datos para garantizar su calidad y adecuación al modelo de aprendizaje automático. Inicialmente, se procede con la eliminación de valores nulos y el tratamiento de datos faltantes mediante estrategias de imputación. Posteriormente, se aplica un proceso de normalización de características mediante la función StandardScale, lo que permite estandarizar la escala de las variables y mejorar la estabilidad del modelo. Además, se realiza la conversión de variables categóricas a valores numéricos utilizando LabelEncoder, lo que facilita su procesamiento en modelos de clasificación binaria.

Tabla I. Descripción de la base de datos

Ubicación	Enlace	Nombre del repositorio	Tipo de licencia	Peso	Origen	Formato
Instituto Canadiense de Ciberseguridad	https://www.unb.ca/cic/datasets/malmem-2022.html	Malware Memory Analysis	Attribution 4.0 International (CC BY 4.0)	4.26 MB	Canadá	.csv

Nota: elaboración propia.



Nota: en la gráfica se observan los diferentes tipos de malware en estudio y registrados en el conjunto de datos. Elaboración propia.

Fig. 1. Discriminación de malware

Se empleó el análisis de componentes principales (PCA) para reducir la dimensionalidad del dataset, preservando inicialmente el 95% de la varianza con 18 componentes principales, lo que optimizó la complejidad computacional y minimizó el riesgo de sobreajuste. Para determinar el número óptimo de componentes (k), se realizó una validación cruzada estratificada con 10 particiones, empleando un clasificador de regresión logística y el F1-score como métrica principal por su equilibrio entre precisión y sensibilidad. Evaluando valores de k de 1 a 50, se identificó que k=41 maximizaba el rendimiento, alcanzando un F1-score promedio de 0.999334 (desviación estándar= 0.000421), lo que refleja estabilidad.

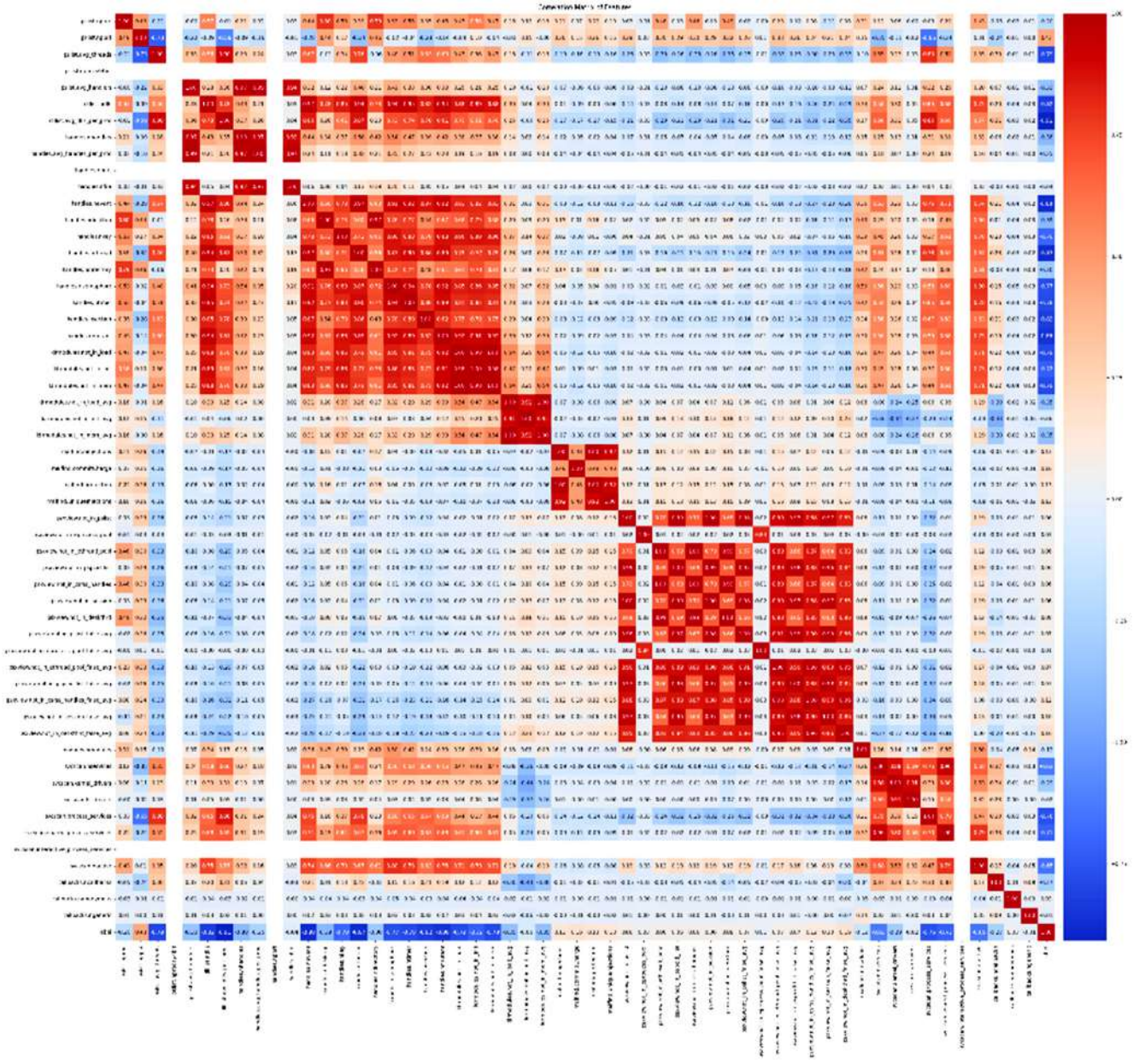
III. Resultados

Etapa tres: minería de datos

La fase de experimentación se lleva a cabo en el entorno de desarrollo Google Colaboratory, utilizando el lenguaje de programación Python y aprovechando la capacidad de procesamiento de la GPU T4 para mejorar la eficiencia computacional. Para el análisis y la minería de datos, se emplean bibliotecas especializadas como pandas, numpy, seaborn, matplotlib, scikit-learn, TensorFlow, Keras, SHAP, entre otras, facilitando la manipulación, visualización y modelado de los datos. El experimento se basa en el conjunto de datos Malware Memory Analysis, el cual comprende diversas categorías de software malicioso.

El análisis de la matriz de correlación del conjunto de datos CIC-MalMem-2022 revela relaciones entre variables clave que describen el comportamiento del malware ofuscado. Las correlaciones más fuertes se observan en métricas relacionadas con el consumo de recursos del sistema, como el tamaño del conjunto de trabajo (*working_set_size*), los bytes privados (*private_bytes*) y el recuento de hilos (*thread_count*). Estas variables presentan una interdependencia significativa, indicando que el malware, particularmente ransomware y troyanos, tiende a consumir grandes cantidades de memoria y a generar múltiples hilos para ejecutar sus rutinas de manera encubierta.

Asimismo, las variables relacionadas con la actividad de red, como los bytes enviados y recibidos por la red (*network_sent_bytes - network_received_bytes*), y las operaciones de lectura y escritura de archivos (*read_operations -write_operations*), muestran correlaciones relevantes con los



Nota: la matriz de correlación arriba muestra las relaciones entre las variables, representando el comportamiento del malware ofuscado. Los valores varían de -1 (correlación negativa) a 1 (correlación positiva). Las áreas en rojo indican una fuerte relación positiva entre variables, como el consumo de memoria (working_set_size) y los hilos generados (thread_count), típicos en ransomware y troyanos. Las áreas en azul reflejan relaciones negativas, útiles para identificar patrones divergentes. Elaboración propia.

Fig. 2. Matriz de correlación del conjunto de datos Malware Memory Analysis

eventos de persistencia, como la creación de procesos secundarios (*process_creation_events*). Este patrón es característico de malware como el spyware, que accede de forma constante a archivos y utiliza conexiones de red para transferir información a servidores remotos.

Redes neuronales artificiales

Por otra parte, se estructuran los algoritmos de redes neuronales artificiales, estas son modelos computacionales inspirados en el cerebro humano, diseñados para procesar información de manera

distribuida[16]. La información fluye a través de neuronas interconectadas, cuyos enlaces están regulados por pesos sinápticos, que determinan su influencia en la red. Su capacidad de aprendizaje radica en la optimización de estos pesos mediante algoritmos de entrenamiento, como la retropropagación del error[17]. Dado esto se cuenta con la aplicación a los datos de las siguientes redes neuronales artificiales:

- Perceptrón Multicapa (MLP): es una extensión del Perceptrón Simple que permite resolver problemas no linealmente separables. Su arquitectura se organiza en tres tipos de capas**:** entrada**,** ocultas y salida, lo que facilita el procesamiento jerárquico de los datos y la extracción de características complejas[18].

- Redes Neuronales Recurrentes (RNN): se caracterizan por su capacidad para procesar y extraer información de datos secuenciales. Su principal ventaja radica en la compartición de parámetros a lo largo de la secuencia, lo que permite modelar relaciones temporales y generalizar a secuencias de longitud variable. Sin esta compartición, cada elemento de la secuencia requeriría parámetros independientes, limitando la capacidad de inferencia del modelo[19].

- Unidades recurrentes cerradas (GRU): son una variante de las RNN que incorporan puertas de actualización y reinicio para optimizar el flujo de información, reduciendo el problema del desvanecimiento del gradiente y mejorando la eficiencia en el procesamiento de secuencias[20].

- Redes Neuronales Profundas (DNN): son un tipo de red neuronal con múltiples capas ocultas, diseñadas para modelar relaciones complejas en los datos mediante el aprendizaje jerárquico de representaciones[21].

- Red Neuronal Feedforward Profunda (RNF): es un tipo de red neuronal artificial en la que la información fluye en una única dirección, desde la capa de entrada hacia la capa de salida, atravesando múltiples capas ocultas sin ciclos ni retroalimentación. Su profundidad permite modelar relaciones complejas y mejorar la capacidad de aprendizaje en tareas como clasificación y regresión[22].

- Rede neuronal convolucional (CNN): son un tipo de redes neuronales profundas diseñadas para procesar datos estructurados en forma de rejilla, como imágenes o series temporales. Utilizan capas convolucionales que aplican filtros para extraer características relevantes y capas de agrupación para reducir la dimensionalidad, preservando información clave[23].

- Red neuronal Transformer: son un tipo de arquitectura de aprendizaje profundo diseñada para procesar secuencias de datos, como texto o series temporales, utilizando un mecanismo de atención que captura relaciones entre elementos sin depender de estructuras recurrentes[24].

Para entrenar las redes neuronales en el dataset Obfuscated-MalMem2022, se emplearon 100 épocas como máximo, utilizando una función de parada temprana (early stopping) para detener el proceso cuando el modelo no mejora, evaluando la pérdida en un conjunto de validación. Los modelos tienen los siguientes parámetros generales: (Tabla II).

Arquitectura de red neuronal con mecanismo de atención Squeeze-and-Excitation (SE)

La arquitectura en este estudio se centra en una red neuronal artificial que incorpora un mecanismo de atención tipo Squeeze-and-Excitation (SE), un componente diseñado para mejorar la capacidad del modelo para recalibrar dinámicamente la importancia de las características. Este enfoque combina atención global y local, lo que permite capturar patrones sutiles en los datos y mejorar significativamente el rendimiento en tareas de minería de datos[27].

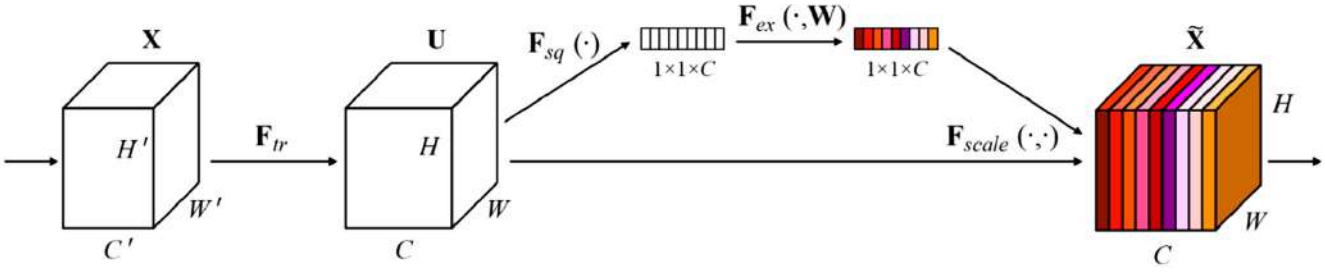
Funcionamiento del Bloque Squeeze-and-Excitation (SE)

El mecanismo Squeeze-and-Excitation (SE) es un tipo de mecanismo de atención que se integra en las redes neuronales convolucionales (CNN) para mejorar su capacidad de representación. Su objetivo es recalibrar las respuestas de los canales de características (feature maps) para destacar los más informativos y suprimir los menos relevantes[28] (Fig. 3).

Tabla II. Parámetros de las redes neuronales artificiales empleadas para el análisis de los datos

Parámetro	MLP	RNN	GRU	DNN	RNF+SE	CNN	Transformer
Arquitectura	2 capas densas (256, 64)	2 capas SimpleRNN (64, 32)	2 capas GRU (64, 32)	3 capas densas (512, 256, 128)	2 capas densas (256, 128)	3 capas Conv2D (32, 64, 128), 2 densas (256, 64)	2 capas Transformer (4 cabezas, 128), 1 densa (64)
Función de Activación	ReLU	Tanh	Tanh	ReLU	ReLU/Sigmoid	ReLU	ReLU
Capa de Salida	Sigmoid	Sigmoid	Sigmoid	Sigmoid	Sigmoid	Sigmoid	Sigmoid
Dropout	0.5	No	No	0.5	0.3	0.4	0.2
Batch Normalization	Sí	No	No	Sí	No	Sí	No
Optimizador	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Tasa de Aprendizaje	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Función de Pérdida	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy
Épocas	100	100	100	100	100	100	100
Batch Size	64	64	64	64	64	64	64

Nota: estructura con base a Goodfellow et al.[25] y Aggarwal[26].



Nota: en la imagen tomada de[28], se observa el proceso de Squeeze-and-Excitation (SE).

Fig. 3. Estructura de la actividad de Squeeze-and-Excitation (SE)

Este se compone de diferentes pasos[29]:

a. Squeeze (Compresión): en esta etapa, se captura información global de cada canal de características. Para ello, se aplica una operación de promedio global (global average pooling) sobre cada feature map, convirtiendo cada canal en un único valor escalar. Esto resume la información espacial de cada canal en un vector de dimensión C (donde C es el número de canales).

- Entrada: Un feature map de tamaño $H \times W \times C$ (altura, ancho, canales).
- Salida: Un vector de tamaño

b. Excitation (Excitación): en esta etapa, se aprende una función de dependencia entre los canales. Para ello, se utiliza una pequeña red neuronal con dos capas fully connected (FC) y una función de activación no lineal (como ReLU) para capturar relaciones no lineales entre los canales.

- Primero, se reduce la dimensionalidad del vector C a C/r (donde r es un factor de reducción) mediante una capa FC.
- Luego, se expande de nuevo a C mediante otra capa FC.
- Finalmente, se aplica una función de activación sigmoide para obtener pesos en el rango[0, 1], que indican la importancia de cada canal.

c. Escalado: los pesos obtenidos en la etapa de excitación se utilizan para recalibrar los feature maps originales. Cada canal se multiplica por su peso correspondiente, destacando los canales más importantes y suprimiendo los menos relevantes.

En términos prácticos, el bloque SE actúa como una capa de atención que modula las características de entrada antes de que sean procesadas por las capas subsiguientes de la red. Esto mejora la representación de las características y también introduce una forma eficiente de regularización implícita, ya que las características irrelevantes son suprimidas durante el entrenamiento[30].

La arquitectura puede clasificarse como una red neuronal feedforward profunda con un mecanismo de atención integrado. Aunque no sigue la estructura recurrente o convolucional típica de otras redes especializadas, su diseño modular y flexible la hace adecuada para una amplia gama de aplicaciones en minería de datos, especialmente en dominios donde la relevancia de las características varía significativamente entre diferentes instancias. Ejemplos de tales aplicaciones incluyen la detección de fraudes, la clasificación de texto y la predicción de eventos raros en series temporales.

Etapa cinco: evaluación de rendimiento del modelo

Matriz de confusión

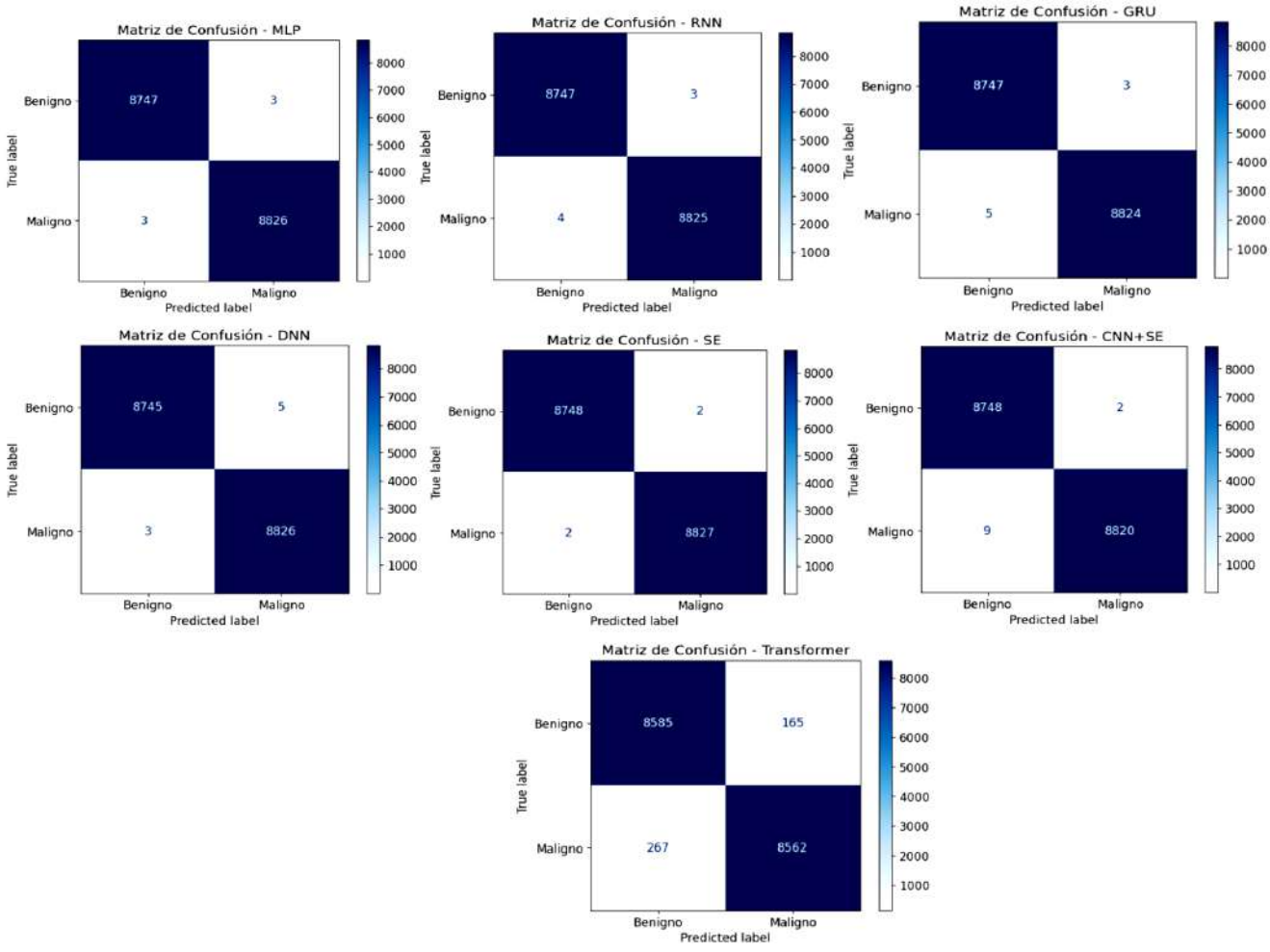
Luego de su entrenamiento con el conjunto de datos se compara el rendimiento de las redes neuronales artificiales en razón a la matriz de confusión siendo una herramienta fundamental para evaluar el desempeño de un modelo de clasificación, ya que compara las predicciones del modelo con las clases reales[18]. Esta representación tabular permite identificar el número de instancias clasificadas correctamente e incorrectamente, proporcionando información detallada sobre la precisión del modelo y facilitando la detección de posibles áreas de mejora[31] (Fig. 4).

El modelo red neuronal con mecanismo de atención Squeeze-and-Excitation (RNF+SE) destaca por su capacidad para recalibrar dinámicamente la importancia de las características, lo que se traduce en una mejora significativa en la precisión y la reducción de falsos positivos y negativos. Los modelos perceptrón multicapa (MLP) y redes neuronales recurrentes (RNN) muestran una buena capacidad de clasificación, reflejada en matrices de confusión con una diagonal predominante. Sin embargo, el modelo de redes neuronales Transformers muestran bajo desempeño.

Métricas generales de evaluación

Siguiendo con la evaluación, se utilizan las métricas generales para el análisis de rendimiento de los modelos de aprendizaje supervisado, estas se comparan en la siguiente tabla III.

En la evaluación de las métricas generales se observa un buen desempeño de todos los modelos. Esto indica una capacidad para clasificar correctamente las muestras. El modelo red neuronal con mecanismo de atención Squeeze-and-Excitation (RNF+SE) lidera con un Accuracy de 0.999772, un F1-score de 0.999773 y un ROC-AUC de 0.999992, reflejando un equilibrio entre precisión y sensibilidad, mientras que MLP, DNN y CNN+SE están en un rango de 0.999 en la mayoría de las métricas, con tiempos de ejecución competitivos (31.5-34.6 segundos). El modelo Transformer, con métricas significativamente bajas y un tiempo de ejecución elevado (183.9 segundos), indica desafíos en la adaptación a los



Nota: se compone de cuatro elementos clave: Verdaderos Positivos (TP): Casos correctamente clasificados como positivos. Falsos Positivos (FP): Casos incorrectamente clasificados como positivos. Falsos Negativos (FN): Casos incorrectamente clasificados como negativos. Verdaderos Negativos (TN): Casos correctamente clasificados como negativos. Elaboración propia.

Fig. 4. Conjunto de evaluación mediante matriz de confusión en el caso de clasificación binaria de los modelos.

Tabla III. Métricas de evaluación para diferentes modelos

	MLP	RNN	GRU	DNN	RNF+SE	CNN+SE	Transformer
Accuracy	0.999659	0.999602	0.999545	0.999545	0.999772	0.999374	0.975425
Precision	0.99966	0.99966	0.99966	0.999434	0.999773	0.999773	0.981093
Recall	0.99966	0.999547	0.999434	0.99966	0.999773	0.998981	0.969759
F1-score	0.99966	0.999604	0.999547	0.999547	0.999773	0.999377	0.975393
ROC-AUC	0.999772	0.99999	0.999997	0.999772	0.999992	0.999993	0.996033
MCC	0.999317	0.999204	0.99909	0.99909	0.999545	0.998749	0.950916
T (S)	33.5	43.8	78.8	31.5	50.5	34.6	183.9

Nota: en la tabla se compara el rendimiento de los modelos de redes respecto a las métricas de evaluación

datos tabulares reformulados como secuencias, destacando una brecha de rendimiento frente a los demás modelos.

Por otro lado, se realiza una evaluación gráfica del desempeño de los modelos de redes neuronales, basada en curvas de pérdida, que grafican la función de pérdida a lo largo de las épocas, estas son clave para evaluar la convergencia del modelo; una disminución constante indica aprendizaje efectivo, mientras que un estancamiento o aumento en la validación puede señalar sobreajuste. Las curvas de precisión, que muestran la evolución del puntaje de precisión, complementan este análisis al reflejar la capacidad del modelo para clasificar correctamente con el tiempo (Fig. 5a y 5b).

a)

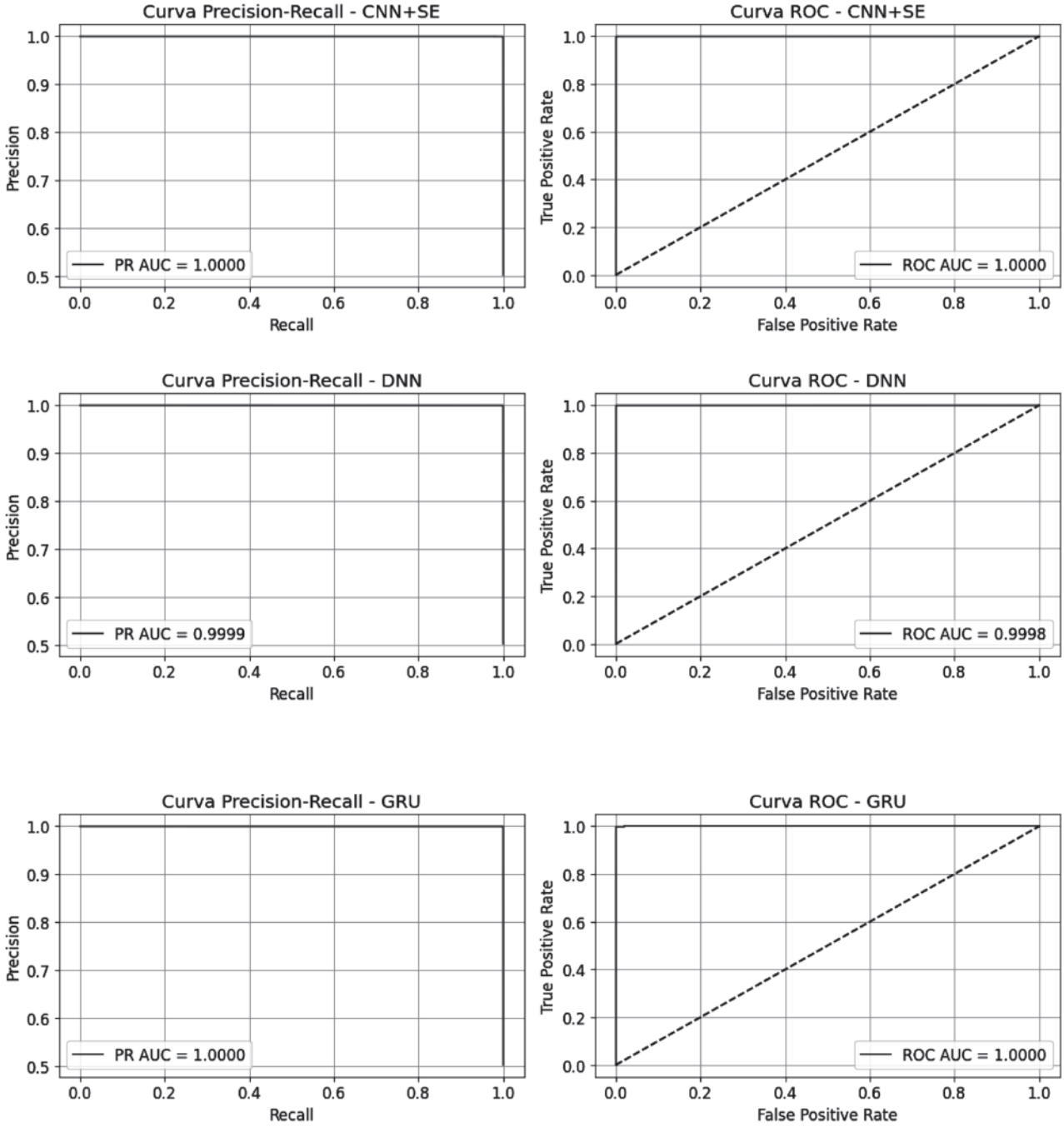
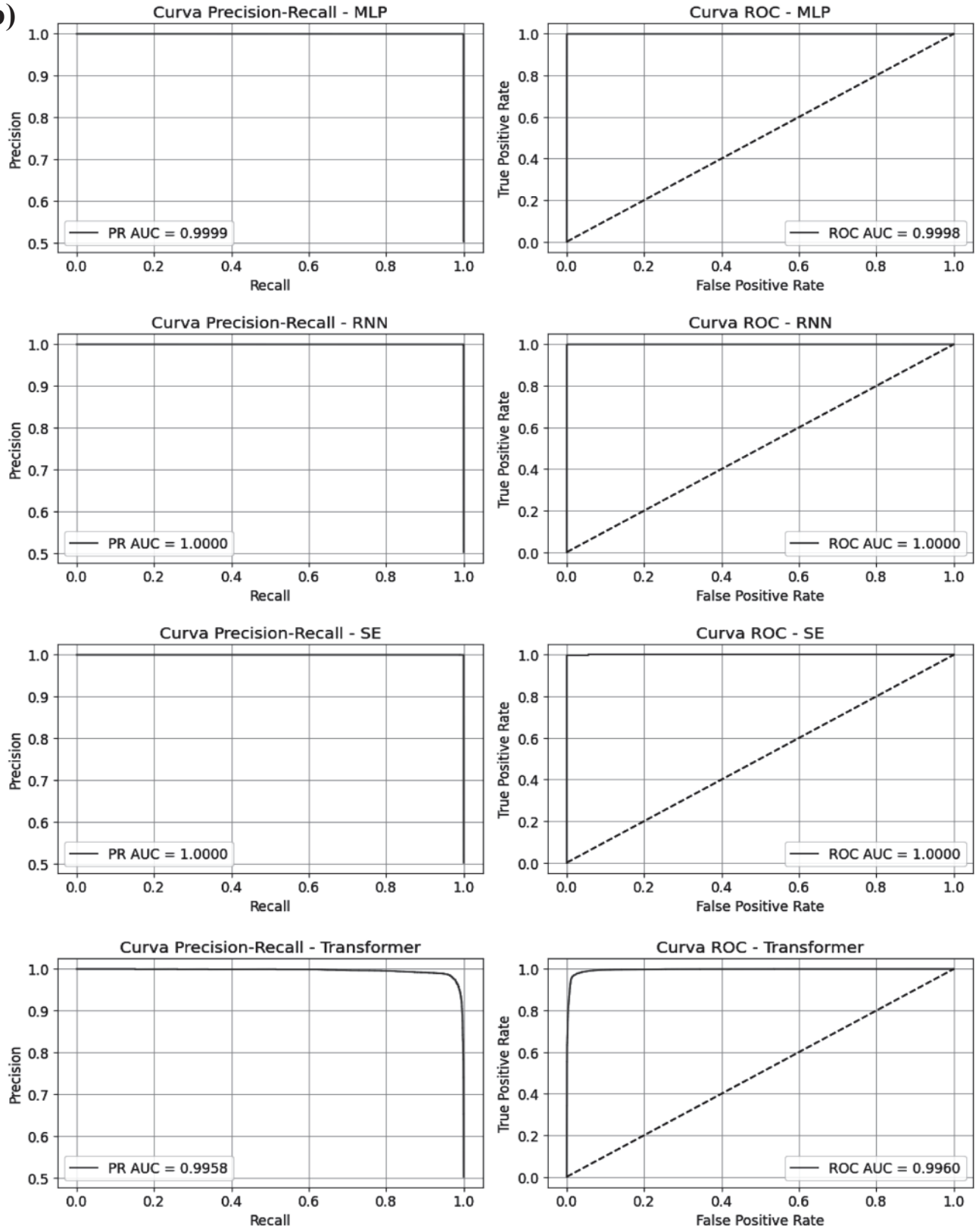


Fig. 5a y b. Evaluación del desempeño de los modelos de redes neuronales mediante curvas Precision-Recall / ROC-AUC.

b)



Nota: las gráficas muestran el desempeño de distintos modelos de redes neuronales en términos de las métricas Precision-Recall (PR AUC) y Receiver Operating Characteristic (ROC AUC). Estas métricas permiten evaluar la capacidad de los modelos para distinguir entre clases.

En términos generales, los modelos CNN+SE, GRU, RNN y RNF+SE se destacan por su desempeño, alcanzando valores de Área Bajo la Curva iguales a 1.0 en ambas curvas, lo cual sugiere una capacidad destacada para la clasificación binaria. Por su parte, los modelos DNN y MLP exhiben un rendimiento bueno, aunque ligeramente inferior, con valores de AUC cercanos a 0.9998 (ROC) y 0.9999 (Precision-Recall). El modelo Transformer, aunque presenta el menor desempeño en comparación con los demás, sigue mostrando resultados satisfactorios. Además, la evolución de la pérdida durante el entrenamiento y la validación refleja una convergencia estable en la mayoría de los casos, lo que indica un aprendizaje efectivo y la ausencia de sobreajuste significativo.

Valores SHAP

Los valores SHAP (SHapley Additive exPlanations) son una técnica basada en la teoría de juegos que permite interpretar la contribución de cada característica en la predicción de un modelo de aprendizaje automático[32] (Fig. 6).

Se obtienen los valores SHAP para un modelo de red neuronal con un mecanismo de atención Squeeze-and-Excitation, identificando las características más influyentes en las predicciones. La primera gráfica (A), identifica las características más relevantes, destacando *svcscan.process_services*, *handles.mmutant* y *handles.nthread* como las de mayor impacto en el rendimiento del modelo. La segunda gráfica (B), detalla cómo los valores individuales de estas características influyen en las predicciones, diferenciando entre valores altos y bajos (representados por los colores rojo y azul, respectivamente). En conjunto, estas visualizaciones priorizan las variables clave y también ofrecen una perspectiva granular sobre su contribución al modelo (Fig. 7).

Los diagramas de fuerza presentados ilustran cómo las características individuales contribuyen a las predicciones de un modelo de aprendizaje automático. En el primer caso (A), con una tendencia hacia "uno", se observa que las variables como *handles.nmutant* y *handles.ntimer* tienen un impacto positivo significativo, impulsando la predicción hacia un valor mayor. En contraste, el segundo diagrama (B) muestra una predicción con tendencia hacia "cero", donde factores como *svcscan.nactive* tienen un efecto negativo predominante, disminuyendo el valor de la predicción.

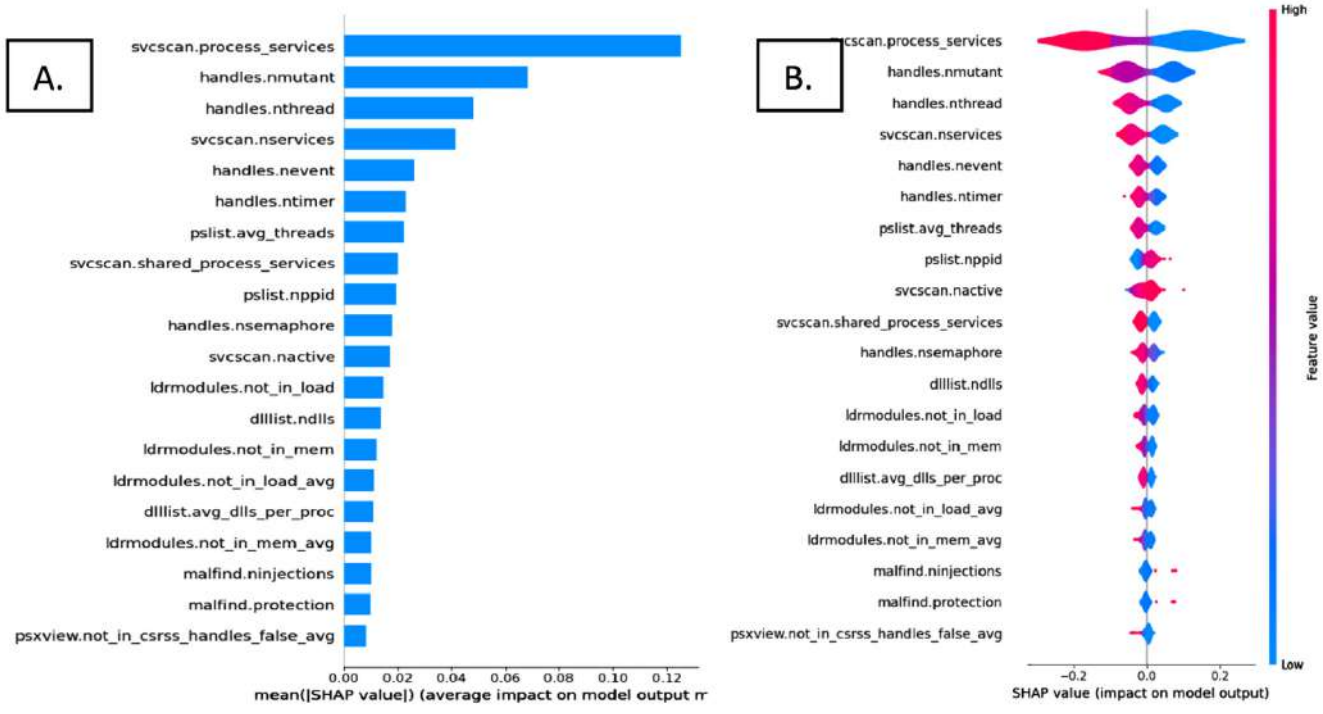


Fig. 6. Resumen que exponen las importancias de las características en el modelo. Elaboración propia.

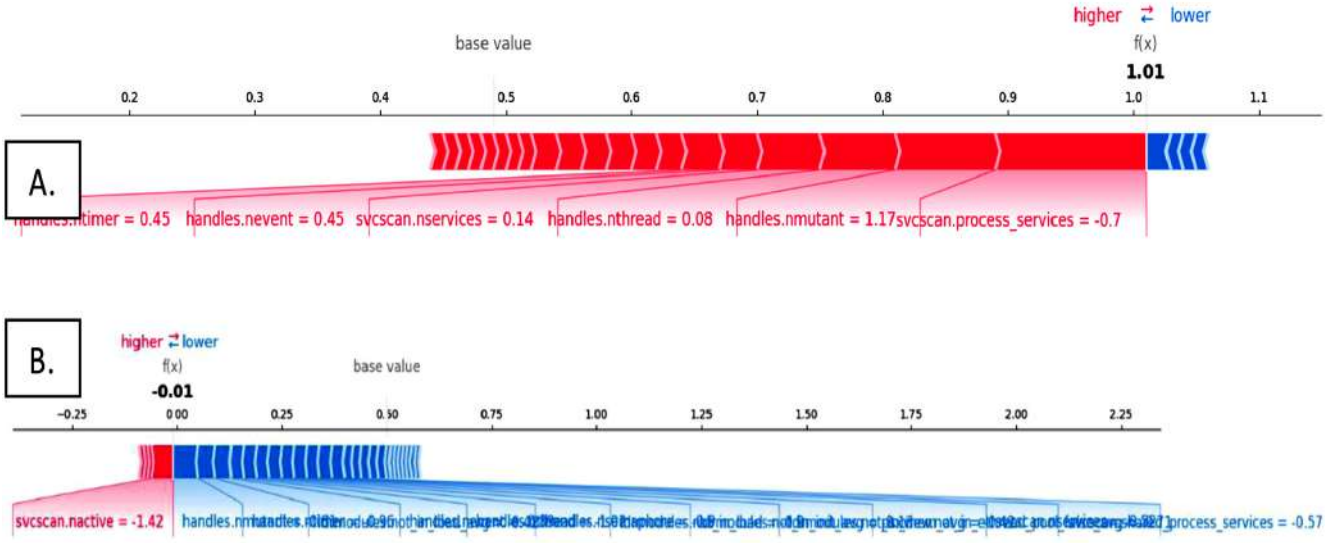


Fig. 7. Diagrama de fuerzas que representan las variables que aportan a una dirección específica. Elaboración propia.

IV. Discusiones

Etapa Cinco: Interpretación

La detección de malware ofuscado representa un desafío creciente en el campo de la ciberseguridad, impulsado por el desarrollo continuo de técnicas avanzadas de evasión utilizadas por los creadores de malware. Estas técnicas, como la manipulación de código, el cifrado y la modificación dinámica de estructuras de datos, dificultan significativamente la identificación mediante métodos tradicionales basados ??en firmas o comportamiento. En este contexto, la incorporación de mecanismos de atención, como Squeeze-and-Excitation (SE), en redes neuronales feedforward emerge como una estrategia prometedora. Estos mecanismos recalibran dinámicamente las características de entrada al identificar y priorizar las más relevantes para la tarea de detección, lo que permite mejorar la precisión y robustez del modelo.

El análisis de la matriz de correlación del conjunto de datos CIC-MalMem-2022 destaca fuertes relaciones entre variables como el tamaño del conjunto de trabajo, los bytes privados y el recuento de hilos, evidenciando que malware como ransomware y troyanos consumen mucha memoria y generan múltiples hilos para operar encubiertamente. El malware ofuscado posee la capacidad de evadir con facilidad los sistemas de detección basados en firmas o comportamientos predefinidos[33]. Esto se debe a que dichas técnicas dependen de patrones específicos que pueden ser modificados o alterados por el propio malware[34].

El análisis de las redes neuronales artificiales evidencia que estas arquitecturas tienen un desempeño notable en tareas de clasificación binaria aplicadas al conjunto de datos. En particular, las redes neuronales recurrentes (RNN) y unidades de corrientes cerradas (GRU) destacan por ofrecer ventajas significativas en este ámbito. Investigaciones previas han demostrado la eficacia de las redes neuronales convolucionales (CNN) en la detección de malware ofuscado, mediante la conversión de ejecutables en imágenes de Markov, alcanzando altos niveles de precisión[35]. Además, los modelos híbridos que combinan CNN con redes de memoria a corto plazo bidimensionales (LSTM) han mostrado ser especialmente efectivos para identificar malware en dispositivos IoT, proporcionando soluciones compactas y de alta velocidad[36].

La red neuronal artificial con mecanismo de atención Squeeze-and-Excitation (RNF+SE) sobresale por su precisión y robustez, logrando un ROC-AUC considerable y destacándose como ideal para problemas complejos que requieren alta precisión, aunque con mayor tiempo de ejecución. Este mecanismo puede mejorar la capacidad de las redes

neuronales para enfocarse en características relevantes del malware, aumentando la precisión de detección al resaltar patrones importantes en los datos de entrada[37].

V. Conclusiones

El presente estudio demuestra la eficacia de la incorporación del mecanismo de atención Squeeze-and-Excitation (SE) en una red neuronal feedforward profunda (RNF+SE) para la detección de malware ofuscado. A través del análisis del conjunto de datos CIC-MalMem-2022, se identificaron correlaciones significativas entre diversas métricas de consumo de recursos del sistema y actividad de red, lo que permitió una caracterización importante del comportamiento malicioso de diferentes tipos de malware, como ransomware, troyanos y spyware.

La evaluación de múltiples arquitecturas de redes neuronales artificiales evidencia que los modelos tradicionales, como el perceptrón multicapa (MLP) y las redes neuronales profundas (DNN), muestran una buena capacidad de clasificación, pero su desempeño se ve limitado en problemas que requieren la identificación de patrones secuenciales o la atención a características específicas. En este contexto, las redes neuronales recurrentes (RNN) y las unidades de corrientes cerradas (GRU) presentan ventajas en la modelización de relaciones temporales, aunque con un costo computacional elevado.

La inclusión del mecanismo de atención SE en la arquitectura de la red neuronal feedforward profunda (RNF+SE) permite una recalibración dinámica de la importancia de las características, optimizando el aprendizaje y reduciendo la tasa de falsos positivos y negativos. Los resultados experimentales muestran que este modelo alcanza un desempeño destacable en comparación de los modelos convencionales, con una exactitud, precisión y un F1-score excepcional, consolidándose como la opción confiable para la detección de malware ofuscado.

A pesar de su mayor tiempo de ejecución en comparación con otras arquitecturas, la capacidad de RNF+SE para identificar patrones sutiles y relevantes en los datos lo posiciona como una alternativa idónea en aplicaciones de seguridad informática donde la precisión y la interpretabilidad son críticas. En trabajos futuros, se sugiere la optimización de la eficiencia computacional del modelo y la evaluación de su desempeño en entornos de producción con datos en tiempo real para validar su aplicabilidad en escenarios prácticos de detección de amenazas cibernéticas. Asimismo, validar el modelo en data sets adicionales para evaluar su generalización.

Fuentes de financiamiento

La presente investigación no recibió financiamiento de entidades gubernamentales, comerciales o sin fines de lucro.

Conflictos de interés

Todos los autores declaramos no poseer conflictos de intereses de índole financiero, profesional o personal que puedan influir de manera inapropiada en los resultados obtenidos o en las interpretaciones formuladas.

Referencias

[1] J. Rivero, «Técnicas de aprendizaje automático para la detección de intrusos en redes de computadoras», *Rev. Cuba. Cienc. Informáticas*, vol. 8, n.º 4, pp. 52-73, 2014, doi: https://www.redalyc.org/journal/3783/378368201003/html/.

[2] O. Cumbicus, P. Ludeña, y L. Neyra, «Técnicas de machine Learning para la detección de Ransomware: Revisión sistemática de Literatura», *J. Sci. Res.*, vol. 7, n.º 3, Art. n.º 3, jul. 2022

[3] M. Hossain y M. Islam, «Enhanced detection of obfuscated malware in memory dumps: a machine learning approach for advanced cybersecurity», *Cybersecurity*, vol. 7, n.º 16, pp. 1-20, 2024, doi: https://doi.org/10.1186/s42400-024-00205-z.

[4] J. Kim y S. Cho, «Obfuscated Malware Detection Using Deep Generative Model based on Global/Local Features», *Comput. Secur.*, vol. 112, n.º 102501, pp. 30-45, ene. 2022, doi: 10.1016/j.cose.2021.102501.

[5] W. Martínez, «Análisis de técnicas de Machine Learning aplicadas a la ciberseguridad informática para mejorar la detección de intrusiones y comportamientos anómalos en la Web», #*ashtag*, vol. 2, n.º 17, Art. n.º 17, 2020, doi: 10.52143/2346139X. 829.

[6] F. Palacios, E. Moya, y W. Paredes, «Análisis de Memoria de Malware Ofuscado en el Conjunto de Datos CIC- MALMEM-2022», *Rev. Multidiscip. Desarro. Agropecu. TECNOLÓGICO Empres. HUMANISTA*, vol. 6, n.º 1, Art. n.º 1, feb. 2024, doi: 10.61236/dateh.v6i1.870.

[7] B. Kolosnjaji, G. Eraisha, G. Webster, A. Zarras, y C. Eckert, «Empowering Convolutional Networks for Malware Classification and Analysis», *Proc. 30th Int. Jt. Conf. Neural Netw. IJCNN*, vol. 2, n.º 2, pp. 3838-3845, 2017, doi: 10.1109/IJCNN.2017.7966340.

[8] S. Kolli, P. Balakesavareddy, y D. Saravanan, «Obfuscated Memory Malware Detection», *Int. Conf. Syst. Comput. Autom. Netw. ICSCAN*, vol. 1, n.º 1, pp. 1-25, ago. 2024, doi: 10.48550/arXiv.2408.12866.

[9] A. Tiwari y N. Chaudhari, «Obfuscated Memory Malware Detection», *ArXiv ArXiv240812866*, vol. 2, n.º 2, pp. 1-20, 2024, doi: doi: 10.48550/arXiv.2408.12866.

[10] E. Villarroel y J. Gutiérrez-Cárdenas, «Dynamic Malware Analysis Using Machine Learning-Based Detection Algorithms», *Interfases*, vol. 19, n.º 19, Art. n.º 019, jul. 2024, doi: 10.26439/interfases2024.n19.7097.

[11] D. Ríos y D. Gomez, *Big data Conceptos, tecnologías y aplicaciones*. Editorial CSIC, 2019.

[12] V. Guzman-Brand y L. Gelvez-García, «Análisis Estadístico y Predictivo de los Datos de Eventos, Víctimas y Desminado Humanitario de las Minas Antipersonal (MAP) en Colombia», *Ciudad Paz-Ando*, vol. 17, n.º 1, Art. n.º 1, jun. 2024, doi: 10.14483/2422278X.21706.

[13] M. L. Avila y J. Medina, «Minería de datos para la predicción de la deserción estudiantil en la Universidad Nacional Abierta y a Distancia», *Doc. Trab. ECBTI*, vol. 1, n.º 2, Art. n.º 2, dic. 2020, doi: 10.22490/ECBTI.4354.

[14] J. R. García-González, P. A. Sánchez-Sánchez, M. Orozco, S. Obredor, y J. R. García-González, «Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia», *Form. Univ.*, vol. 12, n.º 4, pp. 55-62, ago. 2019, doi: 10.4067/S0718-50062019000400055.

[15] Canadian Institute for Cybersecurity, «Malware Memory Analysis», Canadian Institute for Cyber security. Accedido: 7 de febrero de 2025.[En línea]. https://www.unb.ca/cic/datasets/malmem-2022.html

[16] W. Campos y Y. Trujillo, «Redes Neuronales Artificiales en la estimación del esfuerzo», *Rev. Cuba. Cienc. Informáticas*, vol. 15, n.º 2, pp. 183-198, jun. 2021, doi: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2227-18992021000200183&lng=es&nrm=iso&tlng=es.

[17] W. A. Castañeda, B. R. Polo, y F. Vega, «Redes neuronales artificiales: una medición de aprendizajes de pronósticos como demanda potencial», *Univ. Cienc. Tecnol.*, vol. 27, n.º 118, pp. 51-60, mar. 2023, doi: 10.47460/uct.v27i118.686.

[18] F. Berzal, *Redes neuronales & Deep learning*. Editorial Editorial Universidad de Granada, 2018.

[19] C. Arana, «Redes neuronales recurrentes: Análisis de los modelos especializados en datos secuenciales», Serie Documentos de Trabajo, Working Paper 797, 2021. Accedido: 8 de febrero de 2025.[En línea]. https://www.econstor.eu/handle/10419/238422

[20] L. J. Ibáñez, «Arquitectura de Red Neuronal para el Desarrollo de Agentes Conversacionales destinados a la Atención al Cliente en las Redes Sociales», *Cienc. Tecnol.*, vol. 2, n.º 2, pp. 37-53, doi: 10.18682/cyt.vi0.4308.

[21] M. J. Suarez, J. S. Gonzalez, y J. E. Espindola, «Deep Neural Network (DNN) Applied to the Analysis of Student Dropout in a Higher Education Institution», *Investig. E Innov. En Ing.*, vol. 10, n.º 1, Art. n.º 1, jun. 2022, doi: 10.17081/invinno.10.1.5607.

[22] O. A. Soto-Orozco, A. D. Corral-Sáenz, C. E. Rojo-González, y J. A. Ramírez-Quintana, «Análisis del desempeño de redes neuronales profundas para segmentación semántica en hardware limitado», *ReCIBE Rev. Electrónica Comput. Informática Bioméd. Electrónica*, vol. 8, n.º 2, pp. 1-21, 2019.

[23] I. Sakshi, K. Anil, S. Mishra, y A. Pooja, «Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach», *Procedia Comput. Sci.*, vol. 132, n.º 132, pp. 679-688, ene. 2018, doi: 10.1016/j.procs.2018.05.069.

[24] H. Fúquene, «Procesamiento de Lenguaje Natural, los Transformers y los Bots Conversacionales», *XIKUA Bol. Científico Esc. Super. Tlahuelilpan*, vol. 12, n.º 12, pp. 151-160, jul. 2024, doi: 10.29057/xikua.v12iEspecial.12904.

[25] I. Goodfellow, Y. Bengio, y A. Courville, *Deep Learning*. Independiente, 2016.[En línea]. http://alvarestech.com/temp/deep/Deep%20Learning %20by%20Ian%20Goodfellow,%20Yoshua% 20Bengio,%20Aaron%20Courville%20(z-lib.org).pdf

[26] C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*. Springer, 2018.

[27] X. Jin, Y. Xie, X.-S. Wei, B.-R. Zhao, Z.-M. Chen, y X. Tan, «Delving deep into spatial pooling for

squeeze-and-excitation networks», *Pattern Recognit.*, vol. 121, n.º 12, p. 108159, ene. 2022, doi: 10.1016/j.patcog.2021.108159.

[28] J. Hu, L. Shen, S. Albanie, G. Sun, y E. Wu, «Squeeze-and-Excitation Networks», *Arxiv Cornell Univ.*, vol. 2, n.º 2, pp. 1-20, may 2019, doi: 10.48550/arXiv.1709.01507.

[29] M. Cabral, «Redes Convolucionais aplicadas à Segmentação Semântica de Imágenes Sísmicas», Maestría, Pontificia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2021.[En línea]. https://www.maxwell.vrac.puc-rio.br/54148/54148.PDF

[30] R. Chappa y M. El-Sharkawy, «SqueezeNext: una red neuronal profunda eficiente para la implementación de hardware», *10th Annu. Comput. Commun. Workshop Conf. CCWC*, vol. 10, n.º 10, pp. 0691-0697, 2020, doi: doi: 10.1109/CCWC47524.2020.9031119.

[31] J. Gironés, J. Casas, J. Minguillón, y R. Caihuelas, *Minería de datos. Modelos y algoritmos*. Editorial UOC, 2017.

[32] A. Mora, «Explicabilidad de modelos de Machine Learning: Valores SHAP». Accedido: 9 de febrero de 2025.[En línea]. https://blog.damavis.com/explicabilidad-de-modelos-de-machine-learning-valores-shap/

[33] A. Altaher, «An improved Android malware detection scheme based on an evolving hybrid neuro-fuzzy classifier (EHNFC) and permission-based features», *Neural Comput. Appl.*, vol. 28, n.º 12, pp. 4147-4157, dic. 2017, doi: 10.1007/s00521-016-2708-7.

[34] S. Kolli, P. Balakesavareddy, y D. Saravanan, «Neural Network based Obfuscated Malware detection», en *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, India: Puducherry, jul. 2021, pp. 1-5. doi: 10.1109/ICSCAN53069.2021.9526496.

[35] D. K. Dhanya *et al.*, «Obfuscated Malware Detection in IoT Android Applications Using Markov Images and CNN», *IEEE Syst. J.*, vol. 17, nº 2, pp. 2756-2766, jun. 2023, doi: 10.1109/JSYST.2023.3238678.

[36] S. S. Shafin, G. Karmakar, y I. Mareels, «Obfuscated Memory Malware Detection in Resource-Cons trained IoT Devices for Smart City Applications», *Sensors*, vol. 23, nº 11, Art. nº 11, ene. 2023, doi: 10.3390/s23115348.

[37] L. E. González-Medina y R. A. Vázquez, «Clasificación de Malware mediante Redes Neuronales Artificiales», *Rev. Cent. Investig. Univ. Salle*, vol. 11, nº 44, pp. 69-102, 2015, doi: https://www.redalyc.org/articulo.oa?id=34242142004.