

# Desarrollo de un conjunto de datos para la detección de conductas delictivas asociadas con el robo de autopartes

**Resumo** — Este trabajo propone el desarrollo de un conjunto de datos para detectar comportamientos delictivos relacionados con el robo de autopartes. El enfoque se basa en el aprendizaje supervisado, utilizando algoritmos como redes neuronales artificiales (RNA), redes neuronales convolucionales (CNN) y máquinas de vectores soporte (SVM). Estos métodos han mostrado resultados prometedores en diferentes aplicaciones, con precisiones entre el 93% y el 98%. Sin embargo, en la revisión del estado del arte, no se encontraron referencias específicas sobre la identificación del robo de autopartes.

Se propone el uso de un sistema basado en Redes Neuronales Convolucionales 3D (3DCNN) para abordar la detección de robo de autopartes. El sistema se entrena con vídeos de robo y de no robo, y después clasifica los nuevos vídeos en estas dos categorías. De este modo, se pueden identificar comportamientos sospechosos que indiquen un posible robo de piezas de automóvil. El objetivo de esta propuesta es prevenir este delito mediante la detección temprana y la aplicación de medidas preventivas.

**Palabras Clave** - Comportamiento delictivo, aprendizaje supervisado, Support Vector Machine, redes neuronales convolucionales 3D.

**Abstract** — This paper proposes the development of a dataset to detect criminal behavior related to auto parts theft. The approach is based on supervised learning, using algorithms such as artificial neural networks (ANN), convolutional neural networks (CNN) and support vector machines (SVM). These methods have shown promising results in different applications, with accuracies between 93% and 98%. However, in the review of the state of the art, no specific references on the identification of auto parts theft were found.

The use of a system based on 3D Convolutional Neural Networks (3DCNN) is proposed to address auto parts theft detection. The system is trained with theft and non-theft videos, and then classifies new videos into these two categories. In this way, suspicious behavior indicating possible auto parts theft can be identified. The objective of this proposal is to prevent this crime through early detection and implementation of preventive measures.

**Keywords** - Criminal behavior, supervised learning, Support Vector Machine, 3D Convolutional Neural Networks.

## I. INTRODUCCIÓN

La delincuencia, “fenómeno de delinquir o cometer actos fuera de los estatutos previstos por la sociedad” [1], es uno de los principales problemas sociales alrededor del mundo, según la [2] en los últimos años, la delincuencia ha aumentado en distintos países por la proliferación de armas, el desempleo y el abuso de sustancias tóxicas para la salud. Por ello se han observado y estudiado las conductas delictivas que se definen

como “la realización de conductas en contra de las leyes de un país” [3], que encuentran el vandalismo, el hurto y la venta de drogas, por mencionar algunas.

En México existe una clasificación de conductas delictivas que atentan contra el patrimonio, como se muestra en la Fig. 1. Se puede observar que los delitos de robo a casa habitación, robo de vehículo, robo a negocio por mencionar algunos. Estos delitos son frecuentes en el Estado de México, como se indica en el informe del Semáforo Delictivo [4]. En la Fig. 2 se presenta un gráfico que muestra la incidencia del robo de vehículos en los últimos 5 años. Sin embargo, es importante destacar que la falta de denuncias en el robo de autopartes dificulta su inclusión en el análisis posterior.

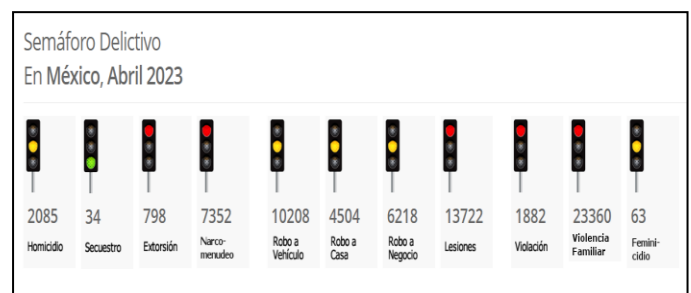


Figura 1. Delitos en México con mayor incidencia.

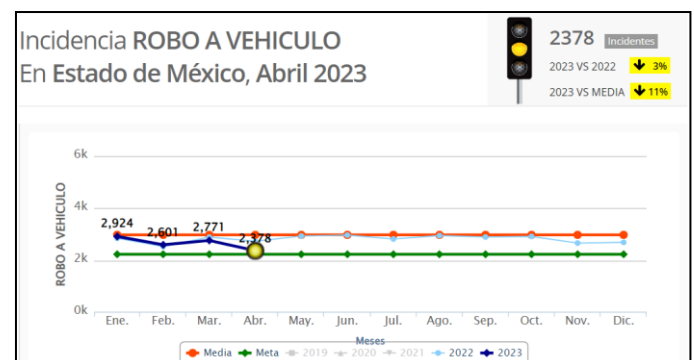


Figura 2. Robo a Vehículo, Estado de México.

Los delitos mencionados, son en esencia, el principal problema de la seguridad ciudadana, aunque actualmente no se ha logrado determinar con precisión el o los factores para estos hechos. Esto motiva que diferentes disciplinas entre ellas la psicología, la sociología, el derecho y la criminología, estudien este gran problema que además de complejo es costoso y dañino para la sociedad.

Por otra parte, el desarrollo tecnológico ha permitido incorporar nuevas herramientas para incrementar la seguridad en zonas de riesgo al enfocarse en la detección y prevención de delitos ya sea de manera masiva o particular. Es bien sabido que se hace uso de dispositivos electrónicos para detectar irregularidades que amenacen la seguridad, por ejemplo, detectores de metales en entradas de lugares públicos como aeropuertos, parques de diversión, estadios; o se emplean cámaras en circuitos cerrados con el fin de identificar personas por medio de reconocimiento facial. Así mismo, se incorporan sensores de movimiento o de proximidad en conjunto con alarmas para enviar alertas a los sectores encargados de la seguridad pública o privada esos lugares. Estos hechos requieren de dichas herramientas para la prevención de actos ilícitos, además, existen investigaciones que proponen un esquema de aprendizaje en línea del contexto para detectar comportamientos sospechosos en videos de vigilancia, lo que permite predecir el ataque de un eventual delincuente antes de que se produzca [5]; por otro lado, el aprendizaje en línea es gradual sabiendo que se tiene que alimentar de manera autónoma con videos en tiempo real. En este punto se abre una brecha donde se puede obtener información errónea, por ejemplo, confundir una escena de robo con otra que presenta solo un juego brusco entre amigos.

Existe también un modelo de aprendizaje profundo de extremo a extremo que se basa en el aprendizaje recurrente bidireccional (BiGRU) y una red neuronal convolucional (CNN) para detectar y prevenir actividades delictivas. La CNN extrae las características espaciales de los fotogramas de video, mientras que BiGRU extrae las características de movimiento temporal y local de múltiples fotogramas de la CNN. Este modelo es capaz de detectar eventos criminales con una eficacia de 98.86% utilizando el conjunto de datos CAVIAR [6]. Esto demuestra que las cámaras de videovigilancia en conjunto con algoritmos desarrollados con aprendizaje profundo son capaces de predecir y prevenir delitos detectando las conductas o comportamientos delictivos.

En estas condiciones se aprecia la importancia de la presente investigación dirigida a la prevención de conductas delictivas que pueden derivar en el robo de autopartes usando aprendizaje supervisado; se obtendrán videos de diferentes fuentes, como YouTube y cámaras de seguridad particulares, que muestren escenas de robos de autopartes para analizar e identificar el comportamiento inusual de un individuo y de esta manera ofrecer una herramienta que ayude a combatir la inseguridad en esta área que actualmente se vive en el país. Esta investigación se inserta en la línea de generación y aplicación del conocimiento de la Inteligencia Artificial (IA).

## II. METODOLOGÍA

En este artículo, se desarrollará el conjunto de datos de un sistema que permitirá detectar conductas delictivas relacionadas con el robo de autopartes mediante el uso de redes neuronales convolucionales 3D, utilizando cámaras de videovigilancia. Para lograrlo, se detalla la primera etapa de las 3 que se proponen en este trabajo, fundamentales para desarrollar un sistema para la detección de conductas delictivas, tal como se muestra en la Fig. 3.

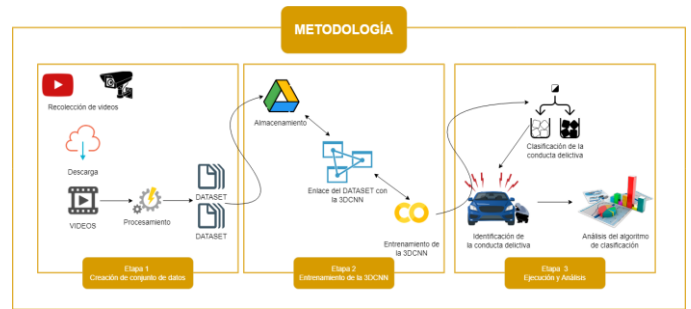


Figura 3. Etapas de la metodología

### Etapas 1. Creación de un conjunto de datos.

*Recolectar videos que contengan escenas de conductas delictivas relacionadas con el robo de autopartes.*

Se creará un conjunto de datos que incluirá videos con escenas de robo de autopartes y otros videos que no incluyan este delito. Cada video se analizará cuidadosamente para identificar las escenas que presenten conductas delictivas, recortando estas escenas en un lapso mínimo de tiempo entre 3 a 30 segundos, para luego almacenarlas en una base de datos. Este proceso se repetirá hasta obtener 1000 videos de conductas delictivas y 1000 videos sin conductas delictivas, los cuales serán utilizados para el futuro entrenamiento. Este enfoque permitirá desarrollar un modelo preciso de clasificación mediante aprendizaje automático, que podrá detectar conductas delictivas en videos relacionados con el robo de autopartes con una alta precisión. Es importante destacar que, al recopilar los videos, es posible que se encuentren con una variedad de resoluciones que van desde 144p hasta 1080p. En la Fig. 4, se muestra un ejemplo de cómo podrían ser las grabaciones necesarias para identificar casos de robo y no robo de autopartes.

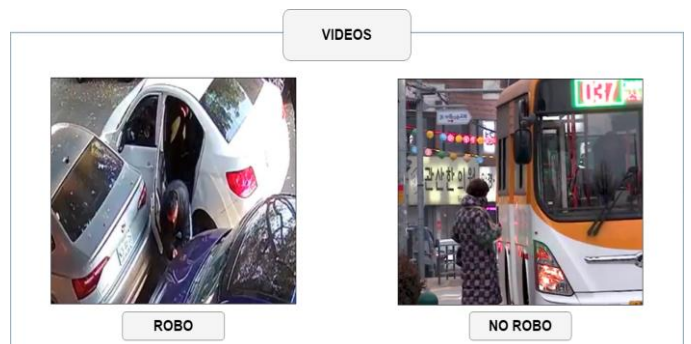


Figura 4. Videos obtenidos de cámaras de seguridad y plataforma YouTube que presentan escenas de robo y no robo de autopartes.

### Etapas 2. Entrenamiento de la 3DCNN

*Procesar los fotogramas obtenidos de cada video a través de la red neuronal convolucional 3D.*

Es imprescindible transferir el conjunto de datos comprimido a un servicio de almacenamiento en la nube. Esto se debe a que hay una gran cantidad de videos recolectados, lo que resulta en un peso considerable que es complicado de manejar. La solución consiste en subir todos los videos a la nube para su posterior almacenamiento. En este caso, se utiliza Google Drive como plataforma de alojamiento para el conjunto de datos y tener el enlace con Google Colab para su posterior uso, como se muestra en la Fig. 5.

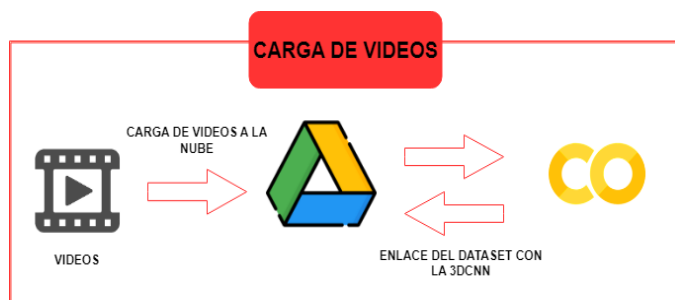


Figura 5. Almacenamiento en la nube

### Clasificación del conjunto de videos.

De los videos almacenados en el conjunto de datos, se extraerán uno por uno alimentando a la red neuronal convolucional 3D, la cual dividirá cada video en fotogramas permitiendo obtener vectores de características convolucionales para cada uno de ellos, estos se concatenarán para formar una matriz vectorial, donde se acomodarán respecto a la información en tiempo y espacio, que representa cada característica, convirtiéndola en una matriz vectorial que en varias dimensiones se representará como otra imagen permitiendo procesarla una vez más por otra red neuronal convolucional 3D. La Fig. 6, representa de manera visual cómo se lleva a cabo el procesamiento del conjunto de datos y su posterior clasificación utilizando una red neuronal convolucional 3D.

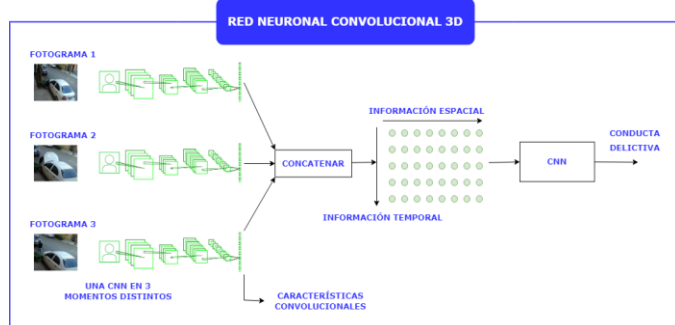


Figura 6. Red neuronal convolucional 3D

Se clasificarán los videos procesados en la red neuronal convolucional 3D en dos categorías: conducta delictiva y conducta no delictiva. Con ello, se formará el conjunto de datos, utilizando un 80% de los videos para entrenamiento y un 20% para pruebas, véase en la Fig. 7. Posteriormente, se evaluará la

eficacia de la red neuronal convolucional 3D en la identificación de conductas delictivas en el robo de autopartes mediante nuevos videos. Cabe mencionar que se presentarán diferentes obstáculos que surgirán durante el entrenamiento, como la posibilidad de perder información relevante y otros problemas relacionados con la clasificación incorrecta de los datos.

Si se realiza una clasificación adecuada, se puede obtener información sobre la ubicación y el momento en que se capturó cada fotograma. Esto nos permite ver cómo estas características cambian con el tiempo e identificar el comportamiento delictivo de un individuo

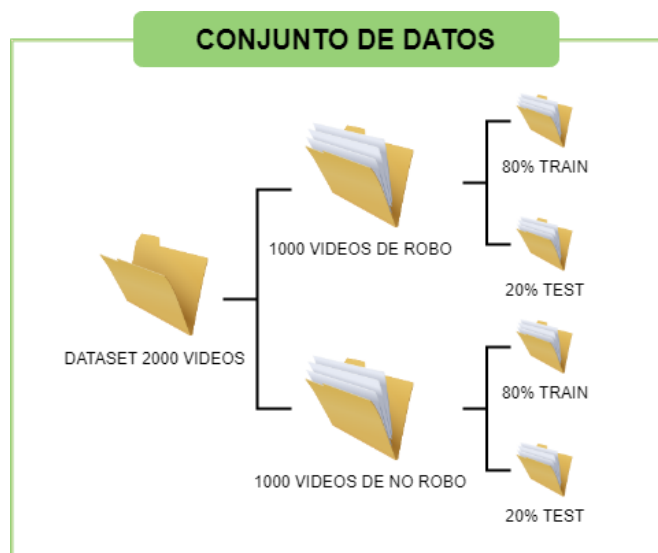


Figura 7. Distribución del conjunto de datos

### Etapas 3. Ejecución y análisis

#### Análisis de resultados.

El análisis de los resultados de la red neuronal convolucional 3D para identificar conductas delictivas en el robo de autopartes implica evaluar qué tan efectivo y preciso es el modelo en la tarea de clasificación. Algunos puntos importantes a considerar en este análisis son los siguientes:

- Exactitud (Accuracy)
- Matriz de confusión (Confusion Matrix)
- Precisión (Precision)
- Recall (Recall)
- F1-Score

Es importante realizar este análisis para evaluar la efectividad de la red neuronal convolucional en la detección de conductas delictivas en el robo de autopartes. Además, se pueden comparar diferentes modelos y técnicas para identificar aquellos que ofrezcan el mejor rendimiento y la mayor precisión en esta tarea específica.

### Construcción del conjunto de datos

En este trabajo se presenta como se generó el conjunto de datos que se utilizará para la detección de conductas delictivas asociadas con el robo de autopartes y que se detalla en los siguientes puntos.

- Descargar videos.

Se realizó una búsqueda de videos en el sitio web YouTube que contenían escenas de robo de autopartes como se muestra en la Fig. 8, copiando la URL de los videos que se deseaban descargar desde la barra de direcciones del navegador. Luego, se utilizó el sitio web savefrom.net para descargar los videos de robos de autopartes, los archivos de video fueron guardados en el dispositivo local en la ruta \Downloads.



Figura 8. Descarga de videos

- Recortar videos.

Se utilizó el software de edición Macrosoft Clipchamp para recortar los videos que contenían escenas de robo de autopartes, véase en la Fig. 9. En la línea de tiempo de los videos, se localizó y seleccionó el punto inicial y final deseado, enfocándose en las escenas específicas del robo de autopartes, y se procedió a realizar los recortes necesarios. Los videos recortados se exportaron en formato (.mp4) y se almacenaron en la ruta \Downloads\Videos-Pre-Procesados.



Figura 9. Recorte de Videos

- Filtrado de videos.

Se empleó el software de edición de video avanzado online-video-cutter.com para depurar y filtrar los videos realizando el análisis minucioso de las características de marcas, mostrado en la Fig. 10, etiquetas y logotipos, considerando su posición, tamaño y apariencia. Posteriormente, se almacenaron en la ruta \Downloads\Videos-Procesados.



Figura 10. Filtrado de videos

- Carga de conjunto de datos a Google Drive.

Se creó una cuenta de Google para poder almacenar los videos, se cargaron desde la PC donde se tenían almacenados con la ruta \Downloads\Videos-Procesados a Google Drive, se crearon dos carpetas adicionales y se agregaron 1000 videos por cada conducta (conducta delictiva y conducta no delictiva) siendo un total de 2000 videos, se muestra en la Fig. 11.

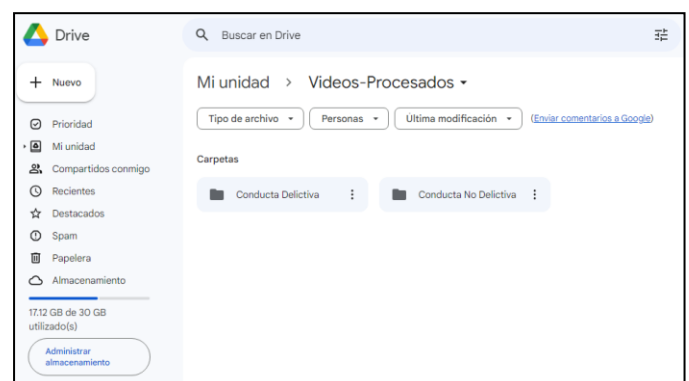


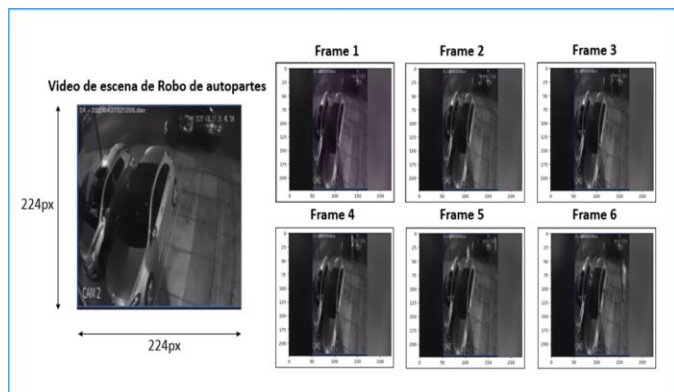
Figura 11. Carga del conjunto de datos a Google Drive

- Extracción de frames de video

Para realizar la extracción de los frames, se estableció una conexión entre Google Colab y Google Drive utilizando la biblioteca drive. Posteriormente, se empleó un algoritmo que utiliza las bibliotecas OpenCV y matplotlib en Python para leer un archivo de video, procesar cada uno de sus fotogramas y mostrarlos como una secuencia de imágenes realizando el recorte estandarizado de cada video almacenado en el conjunto de datos, ya clasificado, esto implicó extraer y guardar cada uno de los frames de los videos en archivos de imagen separados,



convirtiendo los frames en 224 x 224 píxeles, como de observa en la Fig. 12, permitiendo extraer patrones, características, bordes y texturas de los cuadros de video.



**Figura 12.** Extracción de frames de video

### III. RESULTADOS

Para la generación de un conjunto de datos para la detección de conductas delictivas asociadas con el robo de autopartes propuesta, se recopilaron 2000 videos de YouTube, los cuales fueron recortados en segmentos de duración entre 5 y 20 segundos. Cada video tenía un tamaño que variaba de 50 a 8000 KB, dependiendo de su calidad.

Posteriormente, se aplicó un filtro a los videos y se descubrió que 986 de los 2000 contenían marcas que podrían afectar el estudio. Por tanto, se realizaron modificaciones en estos videos para eliminar o reducir al mínimo dichas marcas.

Tanto los videos modificados como los videos sin marcas se almacenaron en Google Drive con el fin de obtener enlaces y facilitar el proceso de extracción de fotogramas. Se utilizó un código que abre el archivo de video utilizando OpenCV y lee los fotogramas secuencialmente. Cada fotograma se convierte de BGR a RGB utilizando la función `cv2.cvtColor()` y se estandariza a 224x224 píxeles usando utilizando `cv2.resize()`. Los fotogramas procesados se guardan en una lista y una vez que se han procesado todos los fotogramas, se muestra cada uno utilizando Matplotlib y `plt.imshow()`. Luego, se utiliza `plt.show()` para visualizar los fotogramas en una ventana.

Estos fotogramas extraídos se utilizarán para el proceso de clasificación mediante una red neuronal convolucional 3D, lo que permitiría realizar una clasificación precisa y eficiente basada en el análisis de los frames extraídos.

### IV. TRABAJOS FUTUROS

Desarrollar el sistema para la detección de conductas delictivas asociadas con el robo de autopartes, continuando la investigación y abordando el desafío de identificar

comportamientos delictivos, realizando la combinación de diversos métodos de aprendizaje supervisado como lo son redes neuronales convolucionales 3D, esto implica utilizar enfoques y técnicas diversas para mejorar la capacidad del sistema y clasificar las conductas delictivas. Para lograrlo, se trabajará en un conjunto de datos más grande que contenga al menos 4000 videos clasificados.

Cabe mencionar que es de suma importancia evaluar el rendimiento del sistema en condiciones reales y realizar los ajustes necesarios para mejorarlo, a través de evaluaciones periódicas, se podrá verificar que el sistema sea efectivo y preciso en la identificación de conductas delictivas en el robo de autopartes.

### V. CONCLUSIONES

Se presenta una propuesta metodológica para el desarrollo de un conjunto de datos para la detección de conductas delictivas asociadas con el robo de autopartes. Se examino que a medida que se recopilen un mayor conjunto de datos, se podrá mejorar el modelo de aprendizaje supervisado y detectar comportamientos delictivos con mayor precisión. Sin embargo, se ha observado que los videos que presentan robos de autopartes a menudo presentan deficiencias, como baja resolución, enfoque deficiente de la cámara, etiquetas o marcas de agua, entre otros. Estos factores pueden introducir ruido durante el entrenamiento del conjunto de datos, lo cual puede afectar la clasificación precisa de los comportamientos delictivos.

Por lo tanto, es necesario garantizar que los videos utilizados en el conjunto de datos sean de la mejor calidad posible. Se debe aplicar un proceso de filtrado exhaustivo para eliminar o reducir al mínimo cualquier elemento que pueda interferir con la clasificación precisa de los comportamientos delictivos en el robo de autopartes. Al hacerlo, se podrá obtener una clasificación más confiable y precisa de las conductas delictivas mediante el entrenamiento de la red neuronal convolucional 3D.

### REFERÊNCIAS BIBLIOGRÁFICA

- [1] M. E. Contreras López, «La Ley Federal contra la Delincuencia Organizada en el contexto global (análisis exploratorio)», 2005.
- [2] «ONU-Habitat - Violencia e inseguridad en las ciudades». <https://onuhabitat.org.mx/index.php/violencia-en-inseguridad-en-las-ciudades> (accedido 20 de noviembre de 2022).
- [3] A. E. Kazdin y G. B. Casal, Conducta antisocial: evaluación, tratamiento y prevención en la infancia y adolescencia. Pirámide, 1999. Accedido: 20 de noviembre de 2022. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/libro?codigo=140340>
- [4] «Semáforo Delictivo». <http://www.semaforo.com.mx/> (accedido 20 de noviembre de 2022).
- [5] Eduardo Francisco Caicedo Bravo, Humberto Loaiza Correa, y Duber Martinez Torres, «Online learning of contexts for detecting suspicious behaviors in surveillance videos», Image and Vision Computing, vol. 89, pp. 197-210, jul. 2019.