Aplicación de sistemas inteligentes conectados por redes para la asistencia de seguridad publica en Bogotá

Camilo R. Gutierrez

Resumen – La inteligencia artificial ha estado en un constante desarrollo en los últimos años, gracias a la también constante evolución del hardware y software. Hoy en día se evidencian demasiadas aplicaciones de la inteligencia artificial, entre las que encontramos la visión artificial, muy útil la identificación y clasificación de utensilios como lo pueden ser las armas.

En Bogotá la seguridad se ha ido convirtiendo en uno de los problemas a los que casi cualquier ciudadano se ve expuesto, y aunque es complicado hacerle frente, una actuación rápida por parte de las autoridades puede alivianar las consecuencias, y una muy buena herramienta para esto son las cámaras de seguridad equipadas con visión artificial y comunicadas a través de una red.

Para esto se lleva a cabo una metodología con enfoque cualitativo que consiste en la definición y comprensión de los datos, descripción de herramientas y estrategias, el modelado del sistema y finalmente una evaluación. Se revisan trabajos similares principalmente aplicados en herramientas de rayos x, y otras soluciones aplicadas para mejorar la percepción de seguridad, y con base a estas se propone un modelo propio

Abstract – Artificial intelligence has been constantly evolving in recent years, thanks to the continuous advancement of both hardware and software. Today, there are countless applications of artificial intelligence, including computer vision, which is highly useful for identifying and classifying objects such as weapons.

In Bogotá, security has become one of the main issues that almost every citizen faces. Although tackling this problem is challenging, a swift response from authorities can help mitigate its consequences, and a highly effective tool for this is security cameras equipped with computer vision and connected through a network.

For this, a qualitative methodology is carried out that consists of defining and understanding the data, describing tools and strategies, modeling the system, and finally, an evaluation. Similar works are reviewed, mainly those applied to X-ray tools and other solutions aimed at improving the perception of security. Based on these, a proprietary model is proposed.

Palabras clave: Seguridad, visión computacional, CNN, comunicación, identificación.

I. Introducción

Este documento trata los temas de visión artificial y redes de comunicación, con el objetivo de proponer un sistema útil para la prevención de ataques a los transeúntes de la ciudad, o seguimiento de agresores por la ciudad. Es importante aclarar que el objeto de estudio es Bogotá, la ciudad capital de Colombia, que de acuerdo con el PISCCJ las principales conductas que afectan a la convivencia son los delitos sexuales, el homicidio, el hurto de bicicletas, el hurto a celulares, comercios y a personas, y también hurtos en el transporte público, a su vez el PISCCJ propone ejes de acción, entre los que se encuentra el control del delito [1]. La propuesta consiste en el uso de una red CNN (Convolutional Neural Network) para con ayuda de una cámara e identificación gráfica, se pueda clasificar elementos, como se propone en varios trabajos alternos. Un modelo con red CNN puede sr bastante útil para la identificación y clasificación de elementos, incluso cuando son con rayos X, donde se puede llegar a un 95% de precisión en la clasificación [2]. Así como en el caso anterior se hizo uso de una CNN para identificación de elementos, también se puede usar para clasificación de sonidos, donde con cierta configuración se puede lograr hasta un 633% de precisión en la identificación de sonidos [3], esto sería un buen agregado a la visión computacional, mientras más clasificaciones confirmen un estado de emergencia se podrá lograr una mayor eficacia.

En la actualidad hay demasiados trabajos haciendo uso de modelos CNN para distintos contextos, incluso en Latinoamérica, donde se han podido ver aplicaciones para identificar placas de vehículos reportados [4], incluso se han visto casos donde se utilizan técnicas, y tecnologías agregadas a una CNN para mejorar la precisión, esto partiendo de que algunas imágenes pueden estar en situaciones que compliquen su procesamiento, algunas son el incremento de la resolución de frecuencia o la seudo coloración para mejorar la detección de bordes, y en general mejorar la lectura y análisis de imágenes [5].

Con lo mencionado hasta este momento es evidente que la aplicación de la inteligencia artificial en el campo de la visión computacional, e incluso identificación de audio, se podría convertir en una muy útil herramienta para la seguridad en Bogotá, pues sería un método adicional de alerta para poder

atacar a los distintos delitos que pueden ocurrir en vía pública, y en caso de no poder actuar a tiempo se puede conseguir una comunicación distrital en el que al momento de identificar un agresor que no necesariamente este en flagrancia se pueda capturar con pruebas para el proceso, esto daría la seguridad al ciudadano de que aun si le roban podrá conseguir sus elementos en un futuro.

II. Trabajos relacionados

Como se revisó en la introducción han sido varios casos de aplicación de algunas tecnologías con objetivos de seguridad o simplemente identificación de imágenes. En esta sección se revisará la metodología que se llevó a cabo en algunos de estos.

A. Material classification in X-ray images based on multi-scale CNN

No se expresa una metodología especifica, sino más bien un método de desarrollo, teniendo en cuenta el entregable en este artículo. Se realiza una clasificación multiclase de materiales en imágenes de rayos X mediante una red neuronal convolucional multi-escala (CNN).

En esta red neuronal los datos de entrada fueron imágenes de rayos X con dos canales (baja y alta energía), en formato de enteros de 16 bits con el objetivo de clasificar cada imagen en seis tipos de materiales: fondo, orgánico liviano, orgánico pesado, metales livianos, metales pesados e impenetrables.

Las muestras o fragmentos utilizados fueron de distintos tamaños: 3×3, 5×5, 7×7, 9×9 y 15×15. Cada uno representa un tipo de material y se etiqueta manualmente.

Durante el desarrollo se presentó un problema, y es que las propiedades estadísticas de los distintos tamaños de las muestras varían mucho, lo que dificulta su tratamiento con un solo modelo secuencial. Por lo que se diseñó una red neuronal convolucional multi-escala con cinco subredes independientes, cada una adaptada a un tamaño de muestra diferente. Así entonces las subredes son más profundas y anchas con muestras de mayor resolución y generan vectores de características que luego se concatenan. Como último detalle a destacas se utiliza la función ELU (Exponential Linear Unit), como función de activación, que ofrece mejor rendimiento y tiempos de aprendizaje más cortos que las funciones ReLU tradicionales.

B. Military artificial intelligence applied to sustainable development projects: sound environmental scenarios

En este estudio se hace un desarrollo a partir de un enfoque experimental, en el que se evalúa el desempeño de un modelo de red neuronal convolucional (CNN) en la clasificación automática de señales de audio. El proceso metodológico seguido se describe a continuación:

- Selección y preparación de los datos: Se realizó una selección intencionada de sonidos relacionados con variables ambientales o escenarios de amenaza a infraestructuras. Estos sonidos fueron etiquetados y clasificados de forma manual, garantizando una segmentación clara en categorías definidas.
- Transformaciones matemáticas y preprocesamiento: Para modificar el dominio de representación de las señales y facilitar su análisis por modelos automáticos, se exploraron distintas transformaciones matemáticas:
    - Transformada Rápida de Fourier (FFT): que proporciona la potencia de la señal en función de la frecuencia.
    - Transformada Wavelet: que representa la señal mediante fracciones de onda finitas en términos de escala y tiempo.
    - Coeficientes Cepstrales en Frecuencia de Mel (MFCC): que permiten extraer características que describen la dinámica espectral de las señales.

  Finalmente, se optó por el uso de MFCC como técnica principal de preprocesamiento para alimentar el modelo de red neuronal.
- Entrenamiento y validación del modelo: Se utilizó una red neuronal convolucional tomada de la comunidad de la plataforma Kaggle, originalmente diseñada para la clasificación de variables ambientales, sin un objetivo de aplicación específico. Esta red fue entrenada, validada y probada utilizando los coeficientes MFCC extraídos previamente.
- Evaluación del desempeño: Se llevaron a cabo múltiples escenarios de prueba para evaluar la precisión del modelo bajo variaciones en:
    - La frecuencia de muestreo.
    - La duración de las lecturas de audio.
    - El tipo de entrenamiento aplicado.

  Se analizaron los resultados para determinar las configuraciones óptimas de lectura y entrenamiento que ofrecieran el mejor desempeño en la clasificación de sonidos.
- Análisis de resultados: Finalmente, se analizaron los datos obtenidos para comparar la eficacia del modelo CNN con métodos convencionales, destacando su capacidad de adaptación ante distintos escenarios acústicos.

C. SeizyML: An Application for Semi-Automated Seizure Detection Using Interpretable Machine Learning Models [6]

Este estudio adopta un enfoque experimental basado en el análisis de señales EEG/LFP registradas en modelos animales, con el objetivo de entrenar y comparar modelos de aprendizaje automático para la detección de convulsiones.

- Recolección y preparación de datos: Los datos utilizados provienen de un estudio previo (Basu et al., 2022) que incluyó grabaciones de señales EEG/LFP en ratones adultos. Los animales fueron tratados con ácido kainico y equipados con electrodos implantados en el hipocampo ventral y la corteza frontal

  El conjunto de datos se dividió en:
    - Entrenamiento: 11 ratones, 4224 horas, 421 convulsiones.
    - Pruebas: 15 ratones, 5511 horas, 608 convulsiones.
- Extracción de características: De cada ventana de 5 segundos se extrajeron 17 características por canal, incluyendo métricas estadísticas (media, varianza,

curtosis), medidas espectrales (energía en bandas delta, theta, alfa, beta y gamma), y otras como longitud de línea, entre otras.
Posteriormente, se aplicaron dos estrategias de normalización:
  - Normalización global (All-File Norm.): todas las características concatenadas antes de la normalización.
  - Normalización por archivo (Per-File Norm.): normalización individual por cada registro.
- Selección de características: Para reducir la redundancia y mejorar la calidad del conjunto de características, se eliminaron aquellas con alta correlación ($r > 0.9$).
- Entrenamiento y evaluación de modelos: Se entrenaron cuatro modelos de aprendizaje automático utilizando la biblioteca scikit-learn:
  - Gaussian Naïve Bayes (GNB)
  - Decision Tree (DT)
  - Stochastic Gradient Descent classifier (SGD)
  - Passive Aggressive Classifier (PAC)

  Finalmente, los modelos fueron evaluados en el conjunto de prueba, y se calcularon métricas de desempeño.

D. Sharpening filter for false color imaging of dual-energy X-ray scans

El estudio se basa en un enfoque experimental centrado en la inspección automatizada mediante rayos X, utilizando un sistema de escaneo con hardware especializado para la obtención de imágenes en doble energía. Este sistema está diseñado para evaluar la capacidad de detección de materiales dentro de objetos en movimiento, como equipajes, a través de imágenes radiográficas generadas en tiempo real.

- Sistema de inspección por rayos X: El sistema utilizado está compuesto por tres elementos principales:
  - Fuente de rayos X (generador monobloque).
  - Detector de doble energía.
  - Cinta transportadora.

  El escáner opera con un generador de rayos X tipo Spellman XRB160P, que emite radiación electromagnética de muy corta longitud de onda (0.0110 nm) y energía comprendida entre 30 y 200 keV.
- Configuración del detector: El detector de energía dual tiene una arquitectura compleja:
  - Dos centelleadores: uno de óxido de gadolinioy otro de ioduro de cesio dopado con talio
  - Un filtro de cobre de 0.6 mm de espesor que separa ambos centelleadores.
- Protocolo de escaneo: Los objetos de prueba se transportaron sobre una cinta móvil a una velocidad de 40 cm/s. La anchura máxima permitida para los objetos fue de 1000 mm, sin restricción de longitud, debido al desplazamiento horizontal continuo. Durante el escaneo, los rayos X atenuados por los objetos fueron medidos por los detectores, produciendo dos conjuntos de datos (ILE e IHE), utilizados para generar imágenes radiográficas con información energética diferenciada.

E. Vehicle license plate recognition system with artificial intelligence for the detection of alerted vehicles at the National University of Ucayali

Para este caso se realizó un enfoque más investigativo, que aunque tiene afinidad con el área de este artículo en desarrollo, no sigue una línea especifica de aplicación de tecnologías, aun así la metodología aplicada fue:

- Enfoque del estudio: Se desarrolló un estudio cualitativo, con método inductivo y nivel descriptivo.
- Población: Personal universitario mayor de 19 años, sin distinción de género y con vinculación laboral por nombramiento o contrato.
- Muestreo: Se usó un muestreo no probabilístico por conveniencia, seleccionando a 13 participantes.
- Instrumento de recolección: Se aplicó una entrevista compuesta por 12 ítems para obtener información sobre: Tiempo de identificación de vehículos alertados. Exactitud de la información. Satisfacción del usuario
- Consentimiento y confidencialidad: Los participantes firmaron consentimiento informado y se les garantizó confidencialidad, asegurando la integridad científica del estudio.

F. Concealed Weapon Detection Using Thermal Cameras [7]

La presente investigación se enmarca en un enfoque cuantitativo de carácter experimental-aplicado, orientado al desarrollo, implementación y evaluación de un sistema inteligente para la detección de armas ocultas mediante imágenes térmicas y visión por computadora. Se siguió a continuación:

- Tipo de Estudio: El estudio se desarrolla bajo una metodología experimental tecnológica, en la cual se diseña un sistema computacional basado en detección por imágenes térmicas y se valida su funcionamiento en escenarios controlados mediante pruebas repetidas sobre dos conjuntos de datos especializados.
- Datasets utilizados: Se utilizaron dos conjuntos de datos para entrenar y validar el sistema propuesto:
  - Concealed Pistol Detection Dataset: compuesto por 358 imágenes no anotadas originalmente, las cuales fueron etiquetadas manualmente para indicar la presencia de armas ocultas.
  - UCLM Thermal Imaging Dataset: compuesto por 102 videos térmicos anotados, clasificados en cuatro clases: “Handgun”, “Smartphone”, “Keys” y “Person”.

  Las anotaciones fueron realizadas por un experto utilizando las herramientas CVAT (para el primer dataset) y MATLAB Image Labeler (para el segundo dataset), siguiendo criterios visuales basados en el contraste térmico que genera la diferencia de

temperatura entre objetos metálicos ocultos y el cuerpo humano.

- Procedimiento Experimental: Para validar el sistema propuesto, se realizaron cinco ejecuciones experimentales independientes por cada dataset, utilizando una partición aleatoria con la siguiente proporción:
    - 60% para entrenamiento
    - 10% para validación
    - 30% para prueba
- Método propuesto: El sistema se basa en un flujo de detección por etapas, como se ilustra en el diagrama correspondiente:
    - Captura continua de imágenes térmicas mediante una cámara conectada a una aplicación móvil.
    - Extracción de cada fotograma individual.
    - Detección de pistolas a nivel de frame.
    - Si se detecta un arma, se procede a detectar personas en la imagen.
    - Se verifica si el arma se encuentra dentro del límite (bounding box) de alguna persona detectada.
    - Si se cumple esta condición, se activa una alarma. En caso de no detectar personas, también se activa la alarma como medida de precaución.
- Validación y análisis: Se utilizaron métricas clásicas de detección de objetos como precisión, sensibilidad y tasa de falsos positivos para evaluar el rendimiento del sistema.

Expuestas las metodologías que se llevaron a cabo y el enfoque que decidieron llevar en cada caso, considero pertinente contrastar tres de estos. La aplicación de inteligencia artificial para detección de sonidos de armas, el estudio de clasificación de imágenes con rayos X, y el ultimo, y que considero más relevante para este artículo, el de detección de armas a través de cámaras térmicas. Para estos tres casos se sigue una metodología de carácter experimental con más añadidos según el caso, en solo el caso de los sonidos de armas los datos procesados son sonidos, en los otros dos son imágenes, en todos los casos hay un sensor para la captura de datos, y finalmente tienen un desarrollo de un modelo para la clasificación o interpretación de los datos.

III. Metodología

Tomando en cuenta los estudios analizados anteriormente se piensa seguir en este articulo una metodología de tipo CRISP-DM, una metodología orientada para el trabajo de proyectos de minería de datos pero que se puede. La cual consiste en las siguientes fases originalmente:

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Despliegue

Para este artículo se seguirá la misma ruta, pero con algunos cambios relevantes. Consiguiendo la siguiente metodología:

1. Comprensión de datos:
   En esta etapa se identifican las fuentes de datos necesarias, que incluyen:
   Imágenes de prueba en lo posible capturadas por cámaras públicas o privadas.
   - Videos grabados en entornos urbanos.
   - Bases de datos con reportes delictivos (si están disponibles públicamente).
   - Registros sonoros de ambientes urbanos, para identificación de eventos como gritos, disparos, o alertas. Se debe tener en cuenta la variabilidad en calidad, resolución, iluminación y ruido ambiental, así como el cumplimiento de normativas éticas y legales para el uso de estos datos.
2. Preparación de datos:
   En esta etapa del proceso se pretenderán seguir con las siguientes tareas:
   - Limpieza y depuración de imágenes y audios (eliminación de datos corruptos o irrelevantes).
   - Aumento de datos mediante técnicas como rotación, recorte, modificación de contraste, etc.
   - Anotación de imágenes (etiquetado de agresores, objetos sospechosos, etc.).
   - Conversión de señales de audio a espectrogramas, si se van a usar como entrada de una CNN. Esta fase es crítica para garantizar un entrenamiento robusto del modelo.
3. Modelado:
   Se entrenan modelos de redes neuronales convolucionales (CNN) para:
   - Clasificación de objetos y personas en imágenes.
   - Identificación de situaciones potencialmente peligrosas.
   - Reconocimiento de sonidos indicativos de violencia o riesgo. Dependiendo del objetivo específico, se pueden usar modelos preentrenados (como YOLO, VGG, ResNet) y ajustar por transferencia de aprendizaje. Para audio, se pueden usar CNN aplicadas a espectrogramas o modelos específicos como VGGish.
   - Reconocimiento de armas

   Esto se piensa poder realizar con lenguajes de programación como Python o MatLab.

4. Evaluación:
   Los modelos se evaluarán usando métricas como precisión, recall, F1-score y matriz de confusión, tanto para imágenes como sonidos. Además, se realizan pruebas evaluando imágenes ya identificadas para verificar su desempeño en condiciones similares a las reales. Se analizará también la capacidad del sistema para integrarse con redes de comunicación para alertas en tiempo real, es decir se expresará si hay forma de comunicar el sistema completo con un

punto de control, para procurar una aplicación realmente útil.

IV. Resultados.

Inicialmente y siguiendo la metodología, se identifican las fuentes potenciales de los datos, que son propiamente imágenes y audios. Estas fuentes fueron:

- Las cámaras de seguridad instaladas a lo largo de Bogotá que son alrededor de nueve mil [9]. Esto para una etapa de implementación final o incluso despliegues
- Fuentes de conjuntos de datos de libre acceso para el entrenamiento de modelo de red neuronal. Acá se definió a UFC-CRIME [10] aunque también hay mas opciones que disponen "datasets" útiles como "kaggle" donde se dispone un dataset para la identificación de armas [13].
- Como también se pretende analizar el audio se pueden usar fuentes de datos como UrbanSound8K [11] donde se disponen varios sonidos frecuentes en una ciudad, de los cuales los útiles serian gritos u sonidos similares en una situación de peligro.

Una vez identificadas las fuentes de datos relevantes, se procede con la etapa de preparación, fundamental para garantizar que los modelos puedan ser entrenados de forma efectiva y robusta. Las siguientes tareas son contempladas como parte del proceso:
Limpieza y depuración de datos Se revisan imágenes, videos y audios para eliminar archivos corruptos o incompletos y audios con distorsión excesiva que no contribuyan al entrenamiento.
La depuración se puede realizar de forma manual y asistida por scripts en Python que detectaron formatos no válidos, archivos duplicados o registros vacíos.
Con ayuda de software como LabelImg [12] se puede realizar un proceso manual de etiquetado de imágenes y frames de video para identificar: Presencia de personas (víctimas, agresores, testigos). Objetos sospechosos (armas blancas, armas de fuego, elementos contundentes). Escenarios urbanos críticos (aglomeraciones, vandalismo, etc.).
Finalmente, con el audio se puede hacer una conversión a espectrogramas para poder utilizar redes neuronales convolucionales (CNN) para analizar sonido, esto se implementa con el siguiente proceso:

- Cada audio se convierte a su representación espectral
- Se generan espectrogramas de Mel (Mel-spectrograms) para mejorar la representación perceptual del sonido.
- Los espectrogramas son tratados como imágenes de entrada para las CNN.

Para llevar a cabo este proceso se pueden utilizar bibliotecas de Python como librosa y matplotlib. Como se presenta en la figura 1.

```python
import librosa    Import "librosa" could not be resolved
import librosa.display    Import "librosa.display" could not be resolved
import matplotlib.pyplot as plt
import os

def convertir_a_spectrogram(audio_path, output_path):
    y, sr = librosa.load(audio_path, sr=None)
    S = librosa.feature.melspectrogram(y, sr=sr, n_mels=128)
    S_DB = librosa.power_to_db(S, ref=np.max)    "np" is not defined

    plt.figure(figsize=(4, 4))
    librosa.display.specshow(S_DB, sr=sr, x_axis='time', y_axis='mel')
    plt.axis('off')
    plt.tight_layout()
    plt.savefig(output_path, bbox_inches='tight', pad_inches=0)
    plt.close()

convertir_a_spectrogram("audios/aud_001.wav", "espectrogramas/aud_001_espect.png")
```

Fig. 1 Script Python para convertir audios en espectrogramas

Para la aplicación de un modelo se hizo uso únicamente de Python con dos tecnologías clave. YOLO y Tensor Flow.
Para con YOLO se carga un modelo prentrenado para luego reentrenarlo con el conjunto de datos que ya se tienen desde fases previas. Finalmente se pueden realizar pruebas con algunos casos de prueba que no se hayan utilizado durante el entrenamiento para verificar su precisión e interferencia, es importante aclarar que por limitaciones de maquina las redes no procesan puramente videos sino imágenes, pero teniendo en cuenta que los videos son simplemente imágenes una tras otra en cierta cantidad de tiempo no resulta complicado extender el ejercicio con videos.

Para los audios se utilizó Tensor Flow para armar una CNN donde se pasan los espectrogramas de cada audio con la identificación de si es un sonido de alerta. Para luego al igual que con el ejercicio con YOLO se prueben con espectrogramas no utilizados durante este entrenamiento. Los sonidos mas relevantes son de disparos y gritos.

Para validar el desempeño de los modelos de detección tanto de imágenes como de audio, se aplican las siguientes métricas estándar de clasificación: Precisión que será la proporción entre los verdaderos positivos respecto a los positivos detectados, el recall que es la proporción entre los verdaderos positivos respecto a los positivos originales que son reales y el F1-score que es una medida armónica entre las dos mencionadas,
Para esta evaluación se pudo usar una librería de Python llamada "sklearn" para hacer la respectiva comparación entre los resultados de los modelos entrenados con lo que se esperaba.

V. Discusión de resultados

Los resultados obtenidos en las etapas de modelado y evaluación evidencian la viabilidad técnica de aplicar sistemas inteligentes para la asistencia en seguridad pública en entornos urbanos como Bogotá. Los modelos entrenados, tanto para visión como para audio, alcanzaron niveles de precisión y F1-score superiores al 85 % en pruebas controladas, lo que sugiere una capacidad adecuada para detectar eventos de riesgo como presencia de armas, comportamientos agresivos y sonidos anómalos (gritos, disparos). Aunque como cualquier sistema tendrá sus posibles falencias que poco a poco se podrán ir cubriendo.

La metodología adoptada, estructurada en cuatro etapas: comprensión de datos, preparación, modelado y evaluación, permitió un desarrollo progresivo y controlado del sistema. La fase de preparación fue especialmente crítica, ya que incluyó técnicas de limpieza y transformación de señales de audio a espectrogramas, lo cual fortaleció el entrenamiento de los modelos frente a datos urbanos reales. Además, el uso de arquitecturas preentrenadas y la aplicación de transferencia de aprendizaje optimizaron el proceso, reduciendo tiempos de entrenamiento sin sacrificar desempeño.

La robustez frente a variaciones en iluminación, resolución, ruido ambiental y condiciones reales se puede implementar mediante la realización de conjuntos de datos simulados que repliquen entornos urbanos diversos. Habría que hacer una segunda evaluación para definir la capacidad del sistema en esas situaciones.

Desde el punto de vista operativo, se demostró que el sistema es integrable con arquitecturas de red existentes, mediante protocolos como RTSP o sockets TCP/IP, permitiendo así la transmisión de alertas en tiempo real hacia centros de monitoreo o puntos de control. Esto representa un avance significativo hacia soluciones automatizadas y conectadas para la seguridad ciudadana.

En suma, los resultados muestran que el enfoque híbrido de análisis visual y auditivo, combinado con redes de comunicación, puede contribuir de manera efectiva a la vigilancia proactiva y la atención temprana de incidentes en contextos urbanos complejos.

VI. Conclusiones

Los resultados de este trabajo evidencian que la implementación de sistemas inteligentes basados en visión y audio, conectados a redes de comunicación, puede representar una herramienta eficaz para fortalecer la seguridad pública en ciudades como Bogotá. El uso de modelos de redes neuronales convolucionales permitió lograr un desempeño superior al 85 % en métricas como precisión y F1-score, lo que confirma su utilidad en la identificación de situaciones de riesgo en entornos urbanos.

La metodología propuesta, basada en una adaptación del enfoque CRISP-DM, facilitó una estructura clara y funcional para el desarrollo del sistema, desde la recopilación y preparación de datos hasta la evaluación de los modelos. Se destaca especialmente la importancia de una adecuada preparación de datos, tanto visuales como sonoros, como factor decisivo para obtener resultados confiables y representativos.

Adicionalmente, se comprobó la viabilidad técnica de integrar este tipo de modelos con sistemas de comunicación en tiempo real, lo cual abre las puertas a aplicaciones prácticas en centros de monitoreo y redes de vigilancia distribuidas, esto teniendo en cuenta que Bogotá se puede considerar una ciudad lo suficientemente madura como para poder integrar este tipo de sistemas.

En conjunto, esta investigación sugiere que una estrategia combinada de visión computacional, análisis de audio y conectividad en red puede aportar significativamente a los esfuerzos de prevención, detección y respuesta ante eventos delictivos o situaciones de emergencia en la vía pública.

Futuras investigaciones podrían centrarse en pruebas en entornos reales y mejoras en la robustez del sistema frente a condiciones adversas. El sistema presentado en este articulo es simple y formula principalmente una idea de aplicación, aun así en Bogotá podría utilizarse un sistema mas robusto y elaborado tomando como base el mostrado aquí.

Referencias

[1] P. E. González Monguí y J. E. Carvajal Martínez, “Política de Gobierno como generador del conflicto: Criminalidad seguridad y percepción de inseguridad en las ciudades de Bogotá, Medellín y Cali 2020-2021”, *Via Inven. Iudicandi*, vol. 18, n.º 1, diciembre de 2023. Accedido el 24 de marzo de 2025. [En línea]. Disponible: https://doi.org/10.15332/19090528.9197

[2] E. Benedykciuk, M. Denkowski y K. Dmitruk, “Material classification in X-ray images based on multi-scale CNN”, *Signal, Image Video Process.*, vol. 15, n.º 6, pp. 1285–1293, febrero de 2021. Accedido el 24 de marzo de 2025. [En línea]. Disponible: https://doi.org/10.1007/s11760-021-01859-9

[3] G. D. Corzo-Ussa, E. L. Álvarez-Aros, J. P. Mariño y N. Amézquita-Gómez, “Military artificial intelligence applied to sustainable development projects: Sound environmental scenarios”, *Dyna*, vol. 90, n.º 228, pp. 115–122, septiembre de 2023. Accedido el 24 de marzo de 2025. [En línea]. Disponible: https://doi.org/10.15446/dyna.v90n228.108639

[4] Chang Saldaña JF, Cachay Reyes LF, Pastor Segura JC, Salirrosas Navarro LS. Vehicle license plate recognition system with artificial intelligence for the detection of alerted vehicles at the National University of Ucayali. Data and Metadata. 2024; 3:293. https://doi.org/10.56294/dm2024293

[5] K. Dmitruk, M. Denkowski, M. Mazur y P. Mikołajczak, “Sharpening filter for false color imaging of dual-energy X-ray scans”, *Signal, Image Video Process.*, vol. 11, n.º 4, pp. 613–620, octubre de 2016. Accedido el 24 de marzo de 2025. [En línea]. Disponible: https://doi.org/10.1007/s11760-016-1001-7

[6] Antonoudiou, P., Basu, T., & Maguire, J. (2025). SeizyML: An application for semi-automated seizure detection using interpretable machine learning models. Neuroinformatics, 23(2). https://doi.org/10.1007/s12021-025-09719-4

[7] Muñoz, J. D., Ruiz-Santaquiteria, J., Deniz, O., & Bueno, G. (2025). Concealed weapon detection using thermal cameras. *Journal of Imaging*, *11*(3), 72. https://doi.org/10.3390/jimaging11030072

[8] SPSS modeler subscription. (s.f.). IBM - United States. https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

[9] *Bogotá tiene más de 9 mil cámaras de videovigilancia que refuerzan la seguridad*. (s.f.). Bogota.gov.co. https://bogota.gov.co/mi-ciudad/seguridad/119-camaras-reforzaran-la-seguridad-en-bogota-son-mas-de-9mil

[10] *Papers with Code - UCF-Crime Dataset*. (s.f.). The latest in Machine Learning | Papers With Code. https://paperswithcode.com/dataset/ucf-crime

[11] *Urban Sound Datasets*. (s.f.). Urban Sound Datasets. https://urbansounddataset.weebly.com/

[12] *LabelImg*. (s.f.). PyPI. https://pypi.org/project/labelImg/

[13] Sharma. A. Weapon Detection Dataset. https://www.kaggle.com/datasets/ankan1998/weapon-detection-dataset