Sistema de seguridad perimetral basado en un modelo TinyML para la detección de armas en la Base de Entrenamiento de Infantería de Marina Coveñas, Sucre

Autor:

Saray Paola Benítez Retamoza [1]

Connie De Ángel Feria [2]

Asesor:

Sergio Antonio Sánchez Hernández

Alex David Morales Acosta

Corporación Universitaria Antonio José de Sucre - UAJS

Facultad de Ciencias de la Ingeniería - FACI

Ingeniería de Sistema

Sincelejo, 2026

Dedicatoria

Dedicamos este proyecto de grado a todas las personas que, de una u otra manera, han estado en cada paso de este camino e hicieron parte fundamental de este logro, especialmente a nuestras familias por el apoyo constante junto a su inspiración nos llevado a alcanzar nuestros sueños y a superar cada barrera presente en este camino.

[Autor 1]: A mis padres, José Gregorio Benítez y Yaneth Retamoza, por sus palabras de aliento, por ser mi mayor motivación para continuar incluso en los momentos más difíciles y recordarme que soy capaz de superar cualquier obstáculo. Su confianza en mí fue la fuerza que me impulsó a llegar aquí.

[Autor 2]: A mis padres, Javier De Ángel Hernández y Diana Feria Pastor, por sus innumerables sacrificios que han hecho para que pudiera alcanzar mis metas, por siempre impulsarme a superarme cada vez más y enseñarme que todo se puede, que a pesar de las dificultades siempre habrá una luz. A mi abuela, quien ya no está físicamente, pero cuya presencia siento en cada instante de mi vida.

A nuestro asesor, Sergio Sánchez, por su guía, dedicación y compromiso. Sus conocimientos y acompañamiento contribuyeron no solo al éxito de este proyecto, sino también a nuestro crecimiento personal y profesional.

Dedicamos este logro a nosotras mismas, por el esfuerzo y perseverancia que han sido prueba de nuestras fortalezas y determinación. Reconociendo cada paso, cada noche de estudio ha sido un desafío superado. Creemos en nosotras mismas y lo lejos que podemos llegar.

Agradecimientos

En primer lugar, queremos expresar nuestro agradecimiento más sincero a Dios, fuente de fortaleza, sabiduría y paz en cada paso de este camino. Su guía ha sido nuestra luz en momento de incertidumbre y su amor nos ha impulsado a seguir adelante. Este logro no hubiera sido posible sin la gracia de Dios quien estipulo fe en nuestros corazones.

Con profunda gratitud agradecemos a nuestras familias por ser pilares inquebrantables en su apoyo constante, sus sacrificios y su amor incondicional que han sostenido y nos han motivado a lo largo de este proceso. Cada palabra de aliento, esfuerzo y gestión de confianza han sido fundamental sobre la construcción de este logro al que también les pertenece.

Extendemos el agradecimiento a nuestro asesor, el profesor Sergio Sánchez, por su acompañamiento, su experiencia y su dedicación constante durante el desarrollo de este proyecto. Su orientación critica, su paciencia fue fundamental para la construcción de esta investigación. Gracias por creer en nosotras y ser una guía en este proceso académico.

Igualmente agradecemos, a los semilleros de investigación y espacios académicos brindados por la Universidad Antonio José de Sucre, por fortalecer nuestras habilidades, conocimiento y por ofrecernos la oportunidad de participación en RedCOLSI, donde adquirimos experiencia y recibimos valiosas retroalimentaciones. Agradecemos por el apoyo y herramientas necesarias en nuestra formación profesional.

Expresamos nuestro agradecemos a nuestros amigos, compañeros y cada una de las personas que extendieron su apoyo y paciencia en lo largo de este proceso. Siempre llevaremos su motivación en nuestros corazones. Estaremos eternamente agradecidas.

Resumen

La vigilancia en instalaciones de infantería de marina enfrenta importantes desafíos asociados a la dependencia de monitoreo manual, la susceptibilidad al error humano y la cobertura de extensas áreas que dificultan una supervisión continua y efectiva. Estas limitaciones reducen la capacidad de respuesta ante amenazas y evidencian la necesidad de incorporar sistemas automatizados e inteligentes que fortalezcan los mecanismos de seguridad. En este contexto, el objetivo principal de la investigación fue desarrollar un sistema de visión computacional basado en inteligencia artificial para la detección de armas en tiempo real, optimizado para su implementación en dispositivos embebidos de bajo costo.

La metodología consistió en el diseño, entrenamiento y despliegue de un modelo de detección de objetos basado en redes neuronales convolucionales, empleando la arquitectura MobileNetV2 bajo el enfoque FOMO. El entrenamiento se llevó a cabo en la plataforma Edge Impulse utilizando un conjunto de datos de 2.808 imágenes con resolución de 160×160 píxeles. Posteriormente, el modelo fue cuantizado a formato INT8 para su ejecución en la tarjeta ESP32-S3-CAM con cámara OV2640, considerando las restricciones de memoria, procesamiento y calidad de imagen propias del dispositivo.

Los resultados evidencian un desempeño adecuado del sistema, alcanzando una precisión (Precision) de 0.62 y un recall de 0.51, con capacidad de inferencia en tiempo real. En cuanto al alcance, se lograron detecciones efectivas en un rango aproximado entre 8 cm y 2 metros, siendo más confiables en distancias cortas. El desempeño del sistema se ve influenciado por factores como la iluminación, la resolución del sensor y la complejidad del entorno.

En conclusión, el estudio demuestra la viabilidad de implementar soluciones de detección de amenazas basadas en inteligencia artificial en dispositivos embebidos de bajo costo, logrando un equilibrio entre eficiencia computacional y desempeño en tiempo real. A pesar de las limitaciones del hardware y del sensor de cámara, se obtuvieron resultados satisfactorios hasta distancias de 2 metros, lo cual es relevante considerando las restricciones del sistema. No obstante, se requiere optimizar el modelo y mejorar las condiciones de captura para aumentar la robustez y precisión en escenarios operativos reales.

Palabras Claves: Detección de armas, MobileNet, Sistemas embebidos, TinyML, FOMO, Vision por computador.

Abstract

Surveillance in naval infantry facilities faces significant challenges associated with reliance on manual monitoring, susceptibility to human error, and the need to cover extensive areas that hinder continuous and effective supervision. These limitations reduce the ability to respond promptly to potential threats and highlight the need for automated and intelligent systems to strengthen security mechanisms. In this context, the main objective of this research was to develop an artificial intelligence-based computer vision system for real-time weapon detection, optimized for deployment on low-cost embedded devices.

The methodology involved the design, training, and deployment of an object detection model based on convolutional neural networks, using the MobileNetV2 architecture under the FOMO (Faster Objects, More Objects) approach. Training was conducted on the Edge Impulse platform using a dataset of 2,808 images with a resolution of 160×160 pixels. The model was subsequently quantized to INT8 format to enable efficient execution on the ESP32-S3-CAM board equipped with an OV2640 camera, considering the device's limitations in memory, processing capacity, and image quality.

The results demonstrate an adequate system performance, achieving a precision of 0.62 and a recall of 0.51, with real-time inference capability. In terms of detection range, the system was able to effectively identify objects within an approximate distance of 8 cm to 2 meters, with higher reliability at shorter distances. System performance is influenced by factors such as lighting conditions, sensor resolution, and environmental complexity.

In conclusion, this study demonstrates the feasibility of implementing AI-based threat detection solutions on low-cost embedded devices, achieving a balance between computational efficiency and real-time performance. Despite hardware and camera

limitations, satisfactory results were obtained at distances of up to 2 meters, which is significant given the system constraints. However, further optimization of the model and improvements in image acquisition conditions are necessary to enhance robustness and accuracy in real-world operational scenarios.

Keywords: Weapon detection, MobileNet, Embedded systems, TinyML, FOMO, Computer vision.

Tabla de Contenido

Lista de tablas

Tabla de figuras

Listado de abreviaciones

IA – Inteligencia Artificial

CNN – Redes Neuronales Convolucionales

RNN – Single Neuronales Recurrentes

P – Precisión

R – Recall – Tiempo de recuperación

IoT – Internet de las Cosas

YOLO - You Only Look Once

FOMO – Faster Objects, More Objects

MobileNet – Familia de modelos CNN

ESP32-S3 – Microcontrolador de Espressif Systems serie S3

FPS – Frames Per Second

Edge IA – Edge Artificial Intelligence

Introducción

En el contexto geopolítico actual, las bases militares se han consolidado como objetivos estratégicos de alto valor, lo que las convierte en blancos recurrentes de ataques en diferentes regiones del mundo. Esta situación se evidencia en eventos recientes, como las tensiones entre Estados Unidos e Irán, donde el Critical Threats Project (2026) reporta múltiples operaciones ofensivas contra infraestructuras militares. De manera similar, Reuters (2026) documenta ataques en países del Golfo, incluyendo Kuwait e Irak, con víctimas mortales y numerosos heridos. Asimismo, en marzo del mismo año, una base militar en Nigeria fue atacada por grupos insurgentes, evidenciando la recurrencia y diversidad de estas amenazas en escenarios internacionales (Reuters; Associated Press, 2026).

En el contexto colombiano, esta problemática adquiere una dimensión particular debido al uso creciente de tecnologías accesibles como drones armados con explosivos por parte de grupos armados ilegales. Según Perspectivas de Seguridad en Colombia (2025), las bases militares, estaciones de policía y centros de formación se han convertido en objetivos frecuentes. Esta tendencia se confirma con datos del Ejército Nacional Colombiano, que reportó la neutralización de 8.872 ataques en 2025, de los cuales 982 fueron interceptados mediante tecnologías de bloqueo de señal. Adicionalmente, entre abril de 2024 y diciembre de 2025 se registraron 393 ataques con drones, impulsados por la facilidad de adquisición de estos dispositivos en mercados internacionales (Agudelo, 2025).

Este panorama evidencia que los sistemas de seguridad tradicionales, basados en monitoreo manual, vigilancia estática y sensores convencionales, presentan limitaciones significativas frente a amenazas modernas. La dependencia de la supervisión humana, la susceptibilidad al error operativo, la fatiga del personal y la necesidad de cubrir extensas áreas reducen la efectividad de la vigilancia y aumentan los tiempos de respuesta ante eventos

críticos (Beltrán Escobar, 2024). En consecuencia, surge un interés creciente por el desarrollo de sistemas automatizados capaces de realizar detección temprana de amenazas mediante tecnologías avanzadas, superando las limitaciones de los enfoques tradicionales (Ángel Rojo, 2025).

En este sentido, la comunidad científica ha explorado diversas soluciones basadas en inteligencia artificial, particularmente en el campo de la visión computacional y la detección de objetos. Modelos avanzados como YOLO (en sus diferentes versiones), Faster R-CNN, SSD (Single Shot Detector) y RetinaNet han demostrado altos niveles de precisión en entornos controlados; sin embargo, presentan elevados requerimientos computacionales, de memoria y, en muchos casos, dependencia de infraestructura en la nube o hardware especializado (GPU/TPU) para su ejecución eficiente.

De manera complementaria, en los últimos años se ha consolidado el análisis de imágenes y video mediante agentes de inteligencia artificial basados en modelos multimodales, especialmente aquellos sustentados en arquitecturas tipo Transformer. Estos modelos, como Vision Transformers (ViT) (Dosovitskiy et al., 2021), DETR (Carion et al., 2020) y enfoques multimodales como CLIP (Radford et al., 2021), permiten integrar información visual y textual para tareas más complejas como reconocimiento contextual, descripción automática de escenas, seguimiento de objetos y análisis semántico en tiempo real. A diferencia de los enfoques tradicionales de detección, estos sistemas no solo identifican objetos, sino que también comprenden relaciones espaciales y temporales dentro de la escena, lo que amplía significativamente sus aplicaciones en entornos dinámicos. No obstante, estos enfoques también presentan limitaciones en términos de costo computacional y consumo energético cuando se consideran implementaciones en dispositivos de borde.

Esta situación limita su aplicabilidad en entornos operativos reales, especialmente en escenarios militares donde la conectividad a internet puede ser intermitente o inexistente, y donde se requieren sistemas autónomos con baja latencia. En este contexto, persiste una brecha significativa entre el alto desempeño de los modelos de inteligencia artificial en entornos controlados y su implementación eficiente en escenarios reales con restricciones de hardware, energía y conectividad.

Como alternativa, han emergido enfoques dentro del paradigma de Edge AI y TinyML, que permiten ejecutar modelos ligeros directamente en microcontroladores, reduciendo la dependencia de la nube y mejorando la capacidad de respuesta en tiempo real. Diversos trabajos respaldan la viabilidad de estas soluciones. Por ejemplo, Pete Warden y Daniel Situnayake (2019) demuestran la implementación de modelos de aprendizaje automático en microcontroladores con recursos limitados, evidenciando la factibilidad de la inferencia local con bajo consumo energético.

Asimismo, Song Han et al. (2016) proponen técnicas de compresión de redes neuronales como pruning y cuantización, permitiendo reducir significativamente el tamaño de los modelos sin afectar considerablemente su precisión, lo cual facilita su despliegue en hardware embebido.

En el ámbito de la visión por computador eficiente, Andrew Howard et al. (2017) y Mark Sandler et al. (2018) desarrollan arquitecturas como MobileNet y MobileNetV2, diseñadas específicamente para dispositivos con recursos limitados, logrando un balance adecuado entre precisión y eficiencia computacional.

Adicionalmente, Francesco Conti et al. (2020) presentan arquitecturas orientadas a la ejecución de redes neuronales en sistemas embebidos de ultra bajo consumo, mientras

que Davide Rossi et al. (2021) evidencian la viabilidad del uso de plataformas paralelas de bajo consumo para aplicaciones de inteligencia artificial en el borde. Estos trabajos demuestran que es posible trasladar capacidades avanzadas de análisis de imágenes hacia dispositivos embebidos con restricciones energéticas.

El presente trabajo se justifica por la necesidad de fortalecer los sistemas de seguridad perimetral mediante soluciones tecnológicas accesibles, autónomas y escalables, que reduzcan la dependencia del monitoreo humano y mejoren la capacidad de respuesta ante amenazas emergentes. Asimismo, contribuye al ámbito académico mediante la exploración de modelos eficientes de inteligencia artificial aplicados a sistemas embebidos, y al sector defensa mediante el desarrollo de herramientas aplicables en escenarios operativos reales.

El alcance de esta investigación se centra en la detección de armas en secuencias de video en tiempo real, mediante el uso de modelos de inteligencia artificial optimizados bajo el paradigma TinyML, implementados en dispositivos embebidos de bajo consumo para aplicaciones de vigilancia perimetral.

En respuesta a esta problemática, el objetivo principal de esta investigación es desarrollar una herramienta de bajo costo para la vigilancia de entornos militares mediante el uso de modelos de inteligencia artificial ligeros basados en TinyML, orientados a la detección de armas en tiempo real. La principal contribución de este trabajo radica en la integración de modelos optimizados de visión computacional con hardware de bajo costo, permitiendo su implementación en escenarios con limitaciones de infraestructura y conectividad.

1. Planteamiento del Problema

1.1. Descripción del problema

En el contexto actual de seguridad, las bases militares y centros de formación de la Fuerza Pública se han consolidado como objetivos estratégicos de alto valor para actores armados, lo que ha incrementado la frecuencia y sofisticación de los ataques. En Colombia, el fortalecimiento de grupos armados ilegales y la incorporación de tecnologías de fácil acceso como drones con explosivos y artefactos no convencionales han ampliado su capacidad ofensiva. Esta realidad se evidencia en múltiples eventos recientes, como el atentado en 2025 contra la base militar de Puerto Jordán (Arauca), que dejó un militar fallecido y siete heridos (Parrado, 2025), y los ataques con drones en el Cauca, incluyendo la represa La Salvajina y la vereda El Amparo, con víctimas mortales y heridos (Defensoría del Pueblo, 2025).

A estos hechos se suman casos de alto impacto en escuelas y centros de formación**,** que evidencian la vulnerabilidad de estos entornos. En 2019, el atentado con carro bomba contra la Escuela de Cadetes de Policía General Santander en Bogotá dejó 22 personas fallecidas y más de 80 heridas (Naciones Unidas, 2019; Corte Suprema de Justicia, 2024). De manera similar, en 2025, la Escuela Militar de Aviación “Marco Fidel Suárez” en Cali fue objeto de un ataque con explosivos, causando al menos seis fallecidos y más de 50 heridos (Fuerza Aeroespacial Colombiana, 2025). Estos eventos demuestran que no solo las bases operativas, sino también los espacios de formación, son blancos recurrentes, incrementando el riesgo para personal en entrenamiento y afectando directamente la capacidad institucional.

Desde una perspectiva causal, esta problemática responde a varios factores interrelacionados. En primer lugar, la disponibilidad de tecnologías accesibles facilita su uso con fines ofensivos. En segundo lugar, los sistemas de seguridad tradicionales continúan dependiendo en gran medida de la vigilancia manual, el monitoreo estático y sensores

convencionales, los cuales presentan limitaciones frente a amenazas dinámicas. Asimismo, la cobertura de extensas áreas, la existencia de puntos ciegos, la distancia entre puestos de control y la sobrecarga de funciones del personal de seguridad incrementan la probabilidad de error humano y reducen la capacidad de detección temprana (Beltrán, 2025; Diana Marchena comunicación persona, 2025).

Como consecuencia de estas condiciones, se evidencia una brecha significativa entre los estándares doctrinales de seguridad como los establecidos en el MFE 3-37, que demandan sistemas integrados, sincronizados y proactivos (Ejército Nacional de Colombia, 2017), y la realidad operativa de muchas unidades. Esta discrepancia se manifiesta en una limitada capacidad de anticipación frente a amenazas, un incremento en el riesgo para el personal militar y en formación, así como en la vulnerabilidad de infraestructuras críticas y posibles afectaciones a las operaciones. Un ejemplo representativo de esta situación es la Base de Entrenamiento de Infantería de Marina de Coveñas, cuya complejidad operativa y amplia extensión territorial dificultan la implementación de una vigilancia continua y efectiva mediante esquemas tradicionales.

Frente a esta problemática, se han implementado soluciones como sistemas de bloqueo de señal, sensores especializados y plataformas de vigilancia avanzadas. En paralelo, la investigación científica ha desarrollado modelos de visión computacional basados en inteligencia artificial para la detección de amenazas. Sin embargo, muchas de estas soluciones presentan limitaciones, como su dependencia de infraestructura en la nube, altos costos y requerimientos computacionales elevados, lo que dificulta su implementación en entornos con restricciones de conectividad y recursos.

En este contexto, surge la necesidad de desarrollar soluciones tecnológicas autónomas, eficientes y de bajo costo que permitan fortalecer los sistemas de seguridad perimetral. En

respuesta a esta necesidad, la presente investigación propone el desarrollo de un sistema basado en modelos ligeros de visión computacional bajo el enfoque de *TinyML*, orientado a la detección de armas en tiempo real en dispositivos embebidos. Esta propuesta busca reducir la dependencia del monitoreo humano, cerrar brechas de cobertura, mejorar la detección temprana de amenazas y contribuir a la protección tanto de bases militares como de centros de formación en escenarios operativos reales.

1.2. Formulación del Problema

A partir de la problemática descrita nos planteamos abordar la siguiente pregunta de investigación.

¿Cómo el desarrollo de un sistema de seguridad perimetral basado en un modelo TinyML para la detección de armas puede mejorar la seguridad en la Base de Entrenamiento de Infantería de Marina Coveñas, Sucre?

2. Justificación

Los avances en tecnologías emergentes asociadas a la Industria 4.0 han impulsado el desarrollo de soluciones orientadas a optimizar los sistemas de vigilancia en diversos entornos. Tecnologías como la inteligencia artificial, el Internet de las Cosas (IoT), la computación en la nube y en el borde (*cloud* y *edge computing*), el análisis de datos masivos (*Big Data*) y los sistemas ciberfísicos han permitido una integración más eficiente entre el mundo físico y digital, facilitando el monitoreo inteligente y la automatización de procesos en escenarios críticos.

No obstante, los sistemas tradicionales de videovigilancia aún presentan limitaciones significativas, principalmente debido a su alta dependencia del monitoreo humano continuo, lo que incrementa la probabilidad de error, reduce la eficiencia operativa y dificulta la respuesta oportuna ante eventos de riesgo. Esta situación resulta especialmente crítica en entornos donde se requiere supervisión constante y cobertura de grandes áreas, como en instalaciones militares y de infraestructura estratégica.

En respuesta a estas limitaciones, los modelos de visión computacional basados en aprendizaje profundo han demostrado ser herramientas eficaces para el análisis automatizado de información visual en tiempo real. En particular, las redes neuronales convolucionales han permitido mejorar la precisión en la detección de objetos y reducir las falsas alarmas, optimizando la identificación de eventos relevantes (Zhang, Li, & Liu, 2023; Ghosh et al., 2022). Asimismo, el desarrollo de arquitecturas ligeras ha dado lugar al enfoque *TinyML*, que posibilita la ejecución de modelos de inteligencia artificial directamente en dispositivos embebidos, disminuyendo la latencia, reduciendo la dependencia de la nube y permitiendo respuestas en tiempo real en entornos con recursos limitados (LatticeWork, 2022).

En este contexto, la presente investigación se justifica en la necesidad de desarrollar soluciones tecnológicas que integren estas capacidades en escenarios reales de seguridad. Específicamente, se propone el diseño de un sistema de seguridad perimetral basado en modelos *TinyML* para la detección de armas, orientado a su implementación en la Base de Entrenamiento de Infantería de Marina de Coveñas, Sucre. Esta propuesta busca mejorar los procesos de monitoreo, reducir la dependencia de la supervisión humana y fortalecer la capacidad de detección temprana de amenazas mediante el uso de dispositivos de bajo costo y alto rendimiento.

Finalmente, los resultados de esta investigación aportan tanto en el ámbito aplicado como académico. Desde el punto de vista práctico, contribuyen al fortalecimiento de los sistemas de seguridad en instalaciones estratégicas, mejorando la eficiencia en la detección de amenazas. Desde el ámbito investigativo, promueven el desarrollo y la aplicación de modelos ligeros de inteligencia artificial en sistemas embebidos, ampliando las posibilidades de implementación de soluciones basadas en *Edge AI* en contextos reales.

3. Objetivos

3.1. Objetivo General

Desarrollar un sistema de seguridad perimetral basado en un modelo TinyML para la detección de armas en la Base de Entrenamiento de Infantería de Marina Coveñas, Sucre.

3.2. Objetivos Específicos:

- Identificar los requerimientos técnicos, tecnológicos y físicos necesarios para la construcción de un sistema de detección de armas en la Base de Entrenamiento de Infantería de Marina en Coveñas, Sucre.
- Codificar un modelo de inteligencia artificial basado en redes neuronales convolucionales para la detección automática de armas en tiempo real, optimizado para dispositivos embebidos bajo el enfoque TinyML.
- Validar el desempeño del modelo mediante métricas de evaluación en escenarios relevantes, con el fin de determinar su eficiencia en la detección de armas en entornos militares de protección.

4. Marco Teórico

4.1. Estado del Arte

En los últimos años, la videovigilancia ha evolucionado significativamente gracias a la incorporación de tecnologías emergentes, particularmente aquellas basadas en inteligencia artificial y visión computacional. Diversos estudios han abordado el desarrollo de sistemas inteligentes capaces de superar las limitaciones de la vigilancia tradicional, caracterizada por su alta dependencia del monitoreo humano y su limitada capacidad de respuesta en tiempo real.

En este contexto, la detección automática de objetos especialmente de armas y el procesamiento eficiente de información visual se han consolidado como líneas de investigación clave, orientadas a mejorar la precisión del monitoreo, reducir errores operativos y fortalecer los mecanismos de seguridad. A continuación, se presentan algunos antecedentes desde el ámbito internacional, nacional y local relevantes que evidencian los avances y tendencias en esta área.

Plumerai & Espressif (2023) desarrollaron un sistema de detección de personas ejecutado directamente en el microcontrolador ESP32-S3, mediante una red neuronal liviana optimizada para dispositivos de bajo consumo. El modelo fue capaz de operar en tiempo real con una velocidad aproximada de 3.3 FPS y un uso de memoria cercano a 166 KB. Además, presentaron demostraciones funcionales que evidenciaron la detección en vivo sobre el dispositivo, validando la capacidad del ESP32-S3 para implementar soluciones de vigilancia autónoma en aplicaciones de seguridad perimetral. Su aporte radicó en demostrar la viabilidad de la visión artificial en el borde en escenarios que requieren monitoreo continuo y respuesta inmediata.

Shalby, Pavan y Roveri (2024) propusieron StreamTinyNet, una arquitectura de TinyML para el análisis de video en tiempo real que utilizó información espacio-temporal de múltiples fotogramas, superando enfoques basados en imágenes individuales. Para su validación, emplearon datasets públicos e implementaron el modelo en la placa Arduino Nicla Vision, demostrando su funcionamiento en dispositivos con recursos limitados. En cuanto a los resultados, alcanzaron precisiones cercanas al 85 %–90 %, junto con una reducción significativa en el consumo de memoria y energía. Su aporte radicó en evidenciar experimentalmente la viabilidad del procesamiento en el borde en dispositivos de bajo consumo.

Bochkovskiy, Wang y Liao (2021) desarrollaron YOLOv7, una arquitectura avanzada para la detección de objetos en tiempo real que incorporó nuevas estrategias de entrenamiento y optimizaciones estructurales. El modelo alcanzó un rendimiento aproximado de 56.8 % mAP en el benchmark COCO, superando a otros detectores en tareas de alta exigencia y manteniendo baja latencia en la inferencia. Sin embargo, sus experimentos se realizaron en entornos de alto rendimiento basados en GPUs, lo que implicó un elevado costo computacional y limitó su implementación directa en dispositivos embebidos.

Boyle, Moosmann, Baumann, Heo y Magno (2024) presentaron DSORT-MCU, un método de detección de objetos pequeños optimizado para microcontroladores dentro del enfoque de Edge AI. Este método introdujo una técnica de segmentación basada en la subdivisión de la imagen en regiones y la combinación de predicciones en zonas superpuestas, mejorando el rendimiento de modelos como FOMO y TinyissimoYOLO. Para su validación, realizaron experimentos en el microcontrolador GAP9, donde lograron reducir el error medio de conteo de 12.9 a 6.2 MAE en el dataset CARPK. Además, alcanzaron velocidades de 33.4 FPS con FOMO y 2.8 FPS con TinyissimoYOLOv1.3 en imágenes completas, con tiempos de

inferencia de 7.31 ms y 16.2 ms por región, respectivamente. Estos resultados demostraron la viabilidad de implementar detección de objetos en dispositivos Edge con microcontroladores.

Muñoz, Santaquiteria, Deniz y Bueno (2025) desarrollaron un sistema de dos etapas para la detección de armas de fuego ocultas basado en termografía y aprendizaje profundo. El método detectó posibles armas a nivel de fotograma y posteriormente verificó su asociación con la persona identificada, lo que permitió reducir falsos positivos. Asimismo, propusieron un algoritmo ligero optimizado para dispositivos embebidos de baja capacidad, facilitando su implementación en sistemas portátiles. Sus resultados experimentales alcanzaron un mAP50-95 de 64,52 %, superando métodos previos y mejorando la fiabilidad del sistema al reducir falsos negativos.

Oñate Miranda (2020) implementó nodos inteligentes para la detección de armas dentro de una red de videovigilancia utilizando técnicas de visión artificial. El sistema fue desarrollado en Python mediante OpenCV y clasificadores HAAR, y se desplegó sobre una Raspberry Pi 3 Model B+. La arquitectura integró un módulo de captura de imágenes que analizó las características de los objetos para identificar posibles armas, enviando alertas mediante tecnología GSM junto con la imagen capturada. Los resultados mostraron tiempos aproximados de 6,5 segundos para la detección y envío de la alerta, y cerca de 10 segundos para el almacenamiento de la información, evidenciando un desempeño adecuado para aplicaciones de monitoreo en tiempo casi real.

López Andrés (2024) desarrolló un sistema de videovigilancia basado en el microcontrolador ESP32-S3, el cual permitió la captura, procesamiento y transmisión de imágenes en tiempo real ante la detección de movimiento. El sistema incluyó una interfaz gráfica que facilitó su control y configuración, así como la posibilidad de procesar las imágenes tanto en el microcontrolador como en un computador, según el modo de operación.

Los resultados evidenciaron la detección de movimiento y la transmisión de video en tiempo real, validando su aplicabilidad en sistemas de vigilancia de bajo costo.

En síntesis, los estudios analizados evidenciaron dos enfoques principales en el desarrollo de sistemas de videovigilancia inteligente. Por un lado, se identificaron modelos de alto rendimiento, como YOLO en sus diferentes versiones, Faster R-CNN, SSD (Single Shot Detector), RetinaNet y arquitecturas basadas en *transformers*, los cuales han alcanzado altos niveles de precisión en tareas de detección de objetos; sin embargo, presentan elevados requerimientos computacionales y una fuerte dependencia de infraestructuras basadas en GPU.

Por otro lado, los enfoques basados en Edge AI y TinyML demostraron la capacidad de operar en dispositivos embebidos con recursos limitados, alcanzando resultados significativos como 3.3 FPS con un uso de memoria de 166 KB, precisiones entre el 85 % y el 90 %, y velocidades de hasta 33.4 FPS. No obstante, estos enfoques aún presentan limitaciones en términos de estabilidad y desempeño en escenarios complejos. En este contexto, se evidencia una brecha entre la eficiencia computacional y la precisión en entornos reales, por lo que la presente investigación se orienta al desarrollo de un sistema basado en TinyML que permita equilibrar estas variables, optimizando la detección de armas en condiciones reales de operación.

4.2. Marco conceptual

El presente proyecto se fundamenta en tecnologías de inteligencia artificial, aprendizaje automático y visión por computador aplicadas a la detección automática de objetos en sistemas de vigilancia. A continuación, se describen los conceptos teóricos esenciales para comprender las etapas de desarrollo del modelo de detección implementado en dispositivos embebidos, específicamente mediante el uso de cámaras y el microcontrolador ESP32-S3, orientado al reconocimiento de objetos de interés en entornos de seguridad.

4.2.1. Inteligencia Artificial

La inteligencia artificial (IA) es una rama de la computación orientada al desarrollo de sistemas capaces de realizar tareas como percepción visual, clasificación, reconocimiento de patrones y toma de decisiones mediante algoritmos entrenados con datos. Estos sistemas buscan simular ciertas capacidades cognitivas humanas para analizar el entorno y generar respuestas autónomas adaptadas a diferentes situaciones (Russell & Norvig, 2021). En este proyecto, la IA se utiliza como eje fundamental para que los modelos procesen imágenes capturadas por la cámara y, con base en ello, identifiquen armas que puedan representar situaciones de riesgo, permitiendo así el análisis automatizado de escenas.

4.2.2. Machine Learning

El machine Learning es un campo de la inteligencia artificial que permite que los sistemas aprendan patrones a partir de datos y mejoren su rendimiento en tareas específicas sin necesidad de una programación explícita. Este enfoque utiliza modelos estadísticos y algoritmos capaces de identificar relaciones, clasificar información o realizar predicciones mediante procesos iterativos de entrenamiento (MIT Sloan School of Management, 2024). En este proyecto, el aprendizaje supervisado se emplea para entrenar modelos con imágenes etiquetadas relacionadas con armas, permitiendo que el sistema reconozca elementos

relevantes dentro del contenido visual y genere resultados en tiempo real en la tarjeta ESP32-S3

4.2.3. Deep Learning

El aprendizaje profundo (Deep Learning) es una subárea del aprendizaje automático que emplea redes neuronales con múltiples capas para extraer representaciones complejas desde datos visuales, textuales o auditivos. Este enfoque permite que los modelos identifiquen características relevantes sin intervención manual, lo que incrementa su precisión en tareas como reconocimiento de objetos y clasificación de escenas (IBM, 2023). En este proyecto, el Deep Learning resulta esencial debido a que el modelo entrenado en Edge Impulse, basado en redes neuronales profundas como MobileNet, permite procesar imágenes capturadas por cámaras embebidas y realizar la detección de objetos en tiempo real dentro del sistema de vigilancia propuesto.

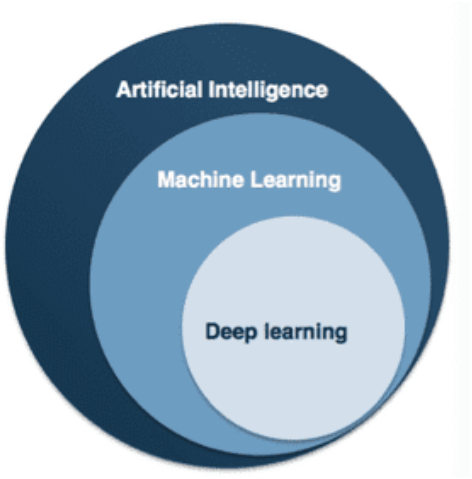


Figura 1. *Relación de los conceptos de inteligencia artificial, aprendizaje de máquina y aprendizaje profundo.*

Nota. Diagrama de Deep Learning. AWS (2020).

4.2.4. Redes Neuronales

Las redes neuronales artificiales son “arquitecturas computacionales basadas en unidades de procesamiento interconectadas (neuronas artificiales) que, mediante algoritmos

de aprendizaje automático, adaptan sus parámetros internos para mapear relaciones no lineales en conjuntos de datos complejos" (Zhang et al., 2023, p. 45). Estas redes, organizadas en capas que transforman progresivamente la información, permiten el aprendizaje representacional jerárquico y han demostrado un desempeño sobresaliente en áreas como la visión por computadora, el reconocimiento de patrones y el análisis de datos visuales. En este proyecto, las redes neuronales constituyen la base del procesamiento del modelo embebido para la detección automática de objetos a partir de imágenes capturadas por cámaras.

4.2.5. Edge AI

La inteligencia artificial en el borde (Edge AI) se refiere a la integración de técnicas de inteligencia artificial con la computación en el borde, permitiendo el procesamiento de datos y la toma de decisiones directamente en dispositivos cercanos a la fuente de información (Jewani & Abimannan, 2023). Este enfoque reduce la latencia, optimiza el uso del ancho de banda y mejora la privacidad de los datos al evitar su transmisión a la nube. Además, permite la operación autónoma de dispositivos en tiempo real, lo que resulta especialmente útil en aplicaciones como los sistemas de vigilancia y seguridad. En este proyecto, Edge AI permite ejecutar el modelo de detección de armas directamente en el sistema embebido, facilitando la respuesta inmediata ante posibles amenazas.

4.2.6. TinyML

TinyML es un enfoque de la inteligencia artificial que permite la implementación de modelos de aprendizaje automático en dispositivos embebidos con recursos limitados, como microcontroladores, mediante técnicas de optimización que reducen el tamaño, el consumo energético y la complejidad computacional de los modelos (Banbury et al., 2020). Este paradigma posibilita la ejecución de inferencias directamente en el dispositivo, sin depender de la nube, lo que favorece aplicaciones en tiempo real y de bajo consumo. Además, TinyML

facilita el desarrollo de sistemas inteligentes autónomos, especialmente en entornos donde se requiere eficiencia energética y operación continua. En este proyecto, TinyML permite desplegar el modelo de detección de armas en dispositivos embebidos, garantizando procesamiento local y respuesta inmediata ante posibles amenazas.

4.2.7. MobileNetV2

MobileNetV2 es un modelo ligero de CNN con menos parámetros y un tamaño de entrada de 224 x 224. La arquitectura MobileNetV1 utilizo conceptos de convoluciones separables en profundidad que aplica un solo filtro a cada canal de entrada, y las convoluciones puntuales apuntan a combinar la salida. MobileNet se compone de dos bloques, unos con stride = [1 1] y otro con stride = [2 2], lo que reduce el tamaño de la entrada, este modelo se ha utilizado ampliamente en el análisis de imágenes médicas (Chandola et al., 2021).

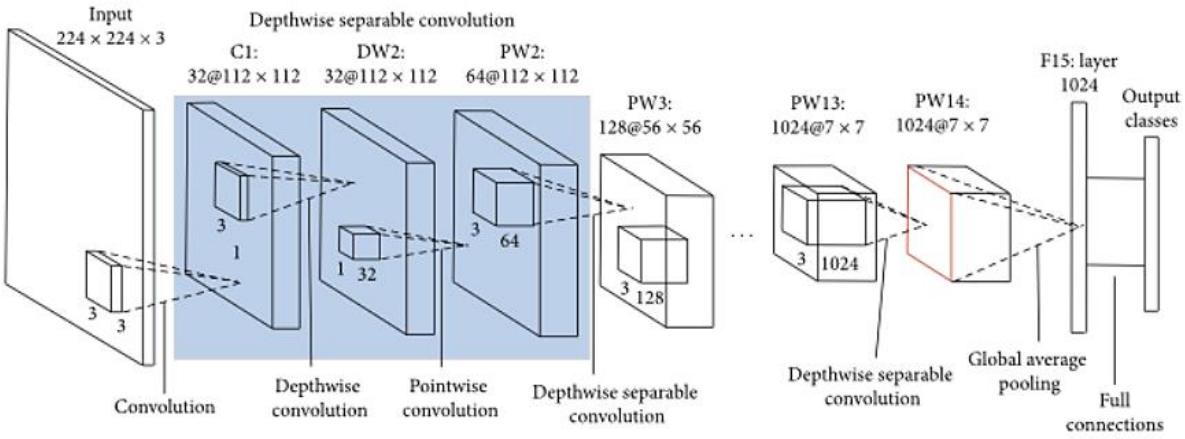


Figura 2: Comprensión de arquitectura MobileNet

Nota. Adaptado de Chandola (2021).

4.2.8. ESP32-S3 WROOM

El ESP32-S3-WROOM con cámara OV5640 es un módulo compacto de bajo costo desarrollado por Espressif Systems, el cual integra un microcontrolador Xtensa LX7 de doble

núcleo a 240 MHz, conectividad Wifi 2.4 GHz, Bluetooth LE 5.0 y hasta 8 MB de PSRAM para el procesamiento de imágenes en tiempo real. Su cámara OV5640, basada en un sensor CMOS de 5 megapíxeles, permite capturar imágenes en resoluciones de hasta 2592×1944 y soporta enfoque automático en versiones avanzadas (Espressif Systems, 2023). Gracias a su eficiencia energética y a interfaces como DVP, SPI e I2C, el módulo resulta ideal para aplicaciones embebidas de visión artificial, reconocimiento facial y vigilancia, lo cual coincide con la implementación realizada en este proyecto para la detección de armas mediante modelos optimizados desde Edge Impulse.

Figura 3 Tarjeta ESP-32S3 WROOM.

Nota. Imagen de referencia de Espressif Systems (2023).

4.2.9. Google Colab

Google Colab es una plataforma basada en notebooks Jupyter que permite ejecutar código Python en la nube con acceso gratuito a recursos como GPU y CPU. Esto facilita el entrenamiento de modelos de aprendizaje automático sin necesidad de hardware local especializado, además de permitir la colaboración en tiempo real y el empleo de bibliotecas como TensorFlow, PyTorch y OpenCV (Google, 2023). En este proyecto, Colab se utilizó para cargar datasets, realizar ajustes del modelo y procesar imágenes, optimizando el flujo de entrenamiento antes del despliegue final.

4.2.10. Edge Impulse

Edge Impulse es una plataforma de aprendizaje automático orientada a dispositivos "Edge AI", diseñada para crear, entrenar y desplegar modelos optimizados para microcontroladores y sistemas embebidos con recursos limitados. De esta forma, permite trabajar con datos visuales, sensoriales y acústicos integrando herramientas de anotación, entrenamiento, cuantización y exportación en formatos compatibles con dispositivos como el ESP32-S3 (Edge Impulse, 2025). En el marco de este proyecto, Edge Impulse fue utilizada para entrenar el modelo de detección de armas y exportarlo en una versión liviana apta para ejecutarse localmente en el microcontrolador.

4.2.11. Arduino IDE

El Arduino IDE es un entorno de desarrollo integrado de código abierto utilizado para programar placas basadas en microcontroladores como la tarjeta ESP32-S3, permitiendo escribir, compilar y cargar programas en lenguaje C/C++ (Arduino, 2024). En este proyecto, se empleó para cargar y ejecutar el modelo TinyML de detección de armas en el microcontrolador, así como para monitorear en el puerto serial el funcionamiento y los resultados del sistema.

4.2.12. Métricas de desempeño

Las métricas de desempeño son indicadores que permiten evaluar el comportamiento de un sistema y su capacidad para cumplir con su propósito. En los sistemas de detección, estas métricas ayudan a medir la eficiencia del modelo al identificar objetos, patrones o eventos. En general, son herramientas fundamentales para analizar la precisión y la efectividad de un modelo.

En sistemas de detección se emplean las siguientes.

- **Precisión:** Es una de las métricas utilizada para medir con exactitud las predicciones positivas realizadas por un modelo en sistemas de detección.

$$Presicion = \frac{TP}{TP + FP}$$

En donde,

TP (True Positives): Casos positivos correctamente clasificados.

FN (False Negatives): Casos negativos que fueron clasificados incorrectamente reales como positivos.

Valores altos de precisión indican que la mayoría de las predicciones positivas realizadas por el modelo son correctas.

- **Tiempo de recuperación (Recall):** También conocida como sensibilidad, es un parámetro que calcula que tan bien un modelo puede identificar correctamente los casos positivos reales.

$$Recall = \frac{TP}{TP + FN}$$

En donde,

TP (True Positives): Casos positivos correctamente clasificados.

FN (False Negatives): Casos positivos reales que el modelo no detectó.

Valores altos del Recall indican que el modelo detecta la mayoría de los casos reales.

- **mAP50 (Mean Average Precision at 50%):** Es un parámetro empleado para medir el rendimiento de modelos de detección de objetos. El mAP50 evalúa la precisión promedio de un modelo al realizar predicciones, y su valor está basado en un umbral del 50%. Esta métrica permite establecer cuánta coincidencia existe entre las predicciones y las anotaciones reales de los objetos (Solawetz, 2026).

- **mAP50:95:** Es un parámetro utilizando para medir el rendimiento de los modelos de detección de objetos. Esta métrica es una extensión del mAP y evalúa la precisión del modelo en función de un rango de umbrales de Intersección sobre la Unión (IoU), que varía del 50% al 95%. De acuerdo, el mAP50:95 analiza cómo el modelo realiza predicciones en diferentes niveles de solapamiento entre las predicciones y las anotaciones reales (Lan1tan, 2024).

5. Metodología de la Investigación

5.1 Método de investigación

La presente investigación se desarrolló bajo un enfoque cuantitativo, de tipo descriptivo y de carácter aplicado, con un diseño no experimental de corte transversal. El enfoque cuantitativo se fundamenta en la medición objetiva de variables y el análisis estadístico de los resultados, permitiendo evaluar el desempeño del modelo mediante indicadores como la precisión (Precision) y el recall, los cuales facilitan la interpretación numérica de la eficiencia del sistema. Este enfoque se sustenta en los planteamientos de Roberto Hernández Sampieri (2014), quien establece que los estudios cuantitativos permiten medir fenómenos y probar comportamientos mediante datos observables y cuantificables.

El diseño no experimental se justifica en que las variables no son manipuladas deliberadamente, sino que se observan en su contexto natural durante la implementación del sistema, tal como lo plantea Fred N. Kerlinger (1986), quien define este tipo de estudios como aquellos donde los fenómenos se analizan sin intervención directa del investigador. Asimismo, el carácter transversal implica que la recolección de datos se realizó en un único momento del tiempo, permitiendo evaluar el comportamiento del modelo en condiciones

específicas de operación. El tipo descriptivo se orienta a caracterizar el desempeño del sistema en términos de métricas de evaluación, sin establecer relaciones causales, en concordancia con lo expuesto por César Augusto Bernal (2010). Finalmente, el carácter aplicado responde a la necesidad de desarrollar una solución tecnológica concreta, orientada a resolver un problema real en el contexto de la seguridad en la Base de Entrenamiento de Infantería de Marina de Coveñas.

Metodológicamente, el estudio se estructuró en torno al diseño, desarrollo e implementación de un sistema de detección de objetos orientado a entornos embebidos. En este proceso, se empleó un conjunto de aproximadamente 2808 imágenes, las cuales fueron previamente etiquetadas para identificar las clases de interés, permitiendo el entrenamiento supervisado del modelo. Con el fin de garantizar su ejecución en dispositivos de bajo consumo, las imágenes fueron procesadas a una resolución de 160×160 píxeles, optimizando así el balance entre precisión y eficiencia computacional.

Posteriormente, el sistema fue evaluado mediante métricas cuantitativas, con el objetivo de determinar su desempeño en escenarios simulados. Esta evaluación permitió analizar su eficiencia y comportamiento bajo condiciones controladas de operación, proporcionando evidencia sobre su viabilidad para aplicaciones en entornos con restricciones de hardware y energía.

5.2 Procedimiento de la Investigación

El proyecto consta de 3 objetivos diseñados en 6 actividades para abordar y desarrollar los objetivos propuestos de la investigación. A continuación, describimos cada una de ellas garantizando un avance coherente en el final de nuestro resultado.

- ***Objetivo 1: Identificar los requerimientos técnicos, tecnológicos y físicos necesarios para la construcción de un sistema inteligente de detección de armas para evaluar las limitaciones de la vigilancia tradicional en la Base de Entrenamiento de Infantería de Marina en Coveñas, Sucre.***
    - **Actividad 1:** Se realizó una revisión sistemática del estado del arte y un proceso de levantamiento de información orientado a identificar tecnologías emergentes aplicadas a la detección de armas en entornos de vigilancia perimetral. Esta actividad incluyó el análisis de modelos de aprendizaje profundo (deep learning), arquitecturas de visión por computador y enfoques basados en inteligencia artificial, evaluando su desempeño, requerimientos computacionales y aplicabilidad en contextos reales. Como resultado, se establecieron criterios técnicos para la selección de tecnologías adecuadas al sistema propuesto.
    - **Actividad 2:** Se identificaron, analizaron y especificaron los requerimientos tecnológicos, computacionales y electrónicos necesarios para la implementación del sistema, incluyendo sistema operativo, capacidades de procesamiento, consumo de memoria, dispositivos de captura (cámaras y sensores) y plataformas de hardware embebido (microcontroladores y tarjetas de desarrollo). Esta actividad contempló la evaluación y selección de componentes con base en criterios de eficiencia, costo y rendimiento, con el propósito de optimizar la ejecución de modelos de inteligencia artificial en dispositivos de borde (edge computing) integrados a sistemas de videovigilancia.

Objetivo 2: Codificar un modelo de inteligencia artificial basado en redes neuronales convolucionales para la detección automática de armas en tiempo real, optimizado para dispositivos embebidos bajo el enfoque TinyML.

- **Actividad 3:** Se llevó a cabo la recopilación, depuración, normalización y preparación de los conjuntos de datos utilizados para el entrenamiento del modelo de detección de objetos. El dataset se construyó a partir de fuentes abiertas, incluyendo plataformas como Roboflow y Edge Impulse, así como datos propios. Posteriormente, se realizaron procesos de anotación manual y semiautomática de las imágenes, identificando las clases de interés (pistol y knife), con el propósito de garantizar la calidad del conjunto de datos y su idoneidad para el entrenamiento supervisado del modelo.
- **Actividad 4:** Se seleccionó la arquitectura de aprendizaje profundo que sirvió como base para el desarrollo del sistema de detección, priorizando modelos de redes neuronales convolucionales ligeras adecuados para entornos de cómputo restringido. A partir de esta selección, se diseñó y entrenó un modelo orientado a la detección automática de objetos en flujos de video capturados por cámaras. El proceso de entrenamiento incorporó técnicas de ajuste de hiperparámetros y regularización, con el fin de optimizar el rendimiento del modelo y garantizar su compatibilidad para el despliegue en dispositivos embebidos, específicamente en el microcontrolador ESP32-S3.

Objetivo 3:· Validar el desempeño del modelo mediante métricas de evaluación en escenarios relevantes, con el fin de determinar su eficiencia en la detección de armas en entornos militares de protección.

- **Actividad 5:** Se evaluaron las arquitecturas previamente entrenadas en entornos controlados, con el propósito de medir su desempeño inicial. El modelo de detección de armas fue desplegado en la tarjeta ESP32-S3 y sometido a pruebas utilizando conjuntos de imágenes con presencia de armas. Se analizaron métricas de evaluación como precisión (precision) y sensibilidad (recall), permitiendo cuantificar la capacidad del modelo para identificar correctamente los objetos de interés y establecer una línea base de rendimiento antes de su validación en condiciones reales.
- **Actividad 6:** Se realizó la validación experimental del sistema en condiciones operativas, empleando imágenes capturadas en tiempo real mediante la cámara integrada de la ESP32-S3. Durante esta fase, se analizaron variables asociadas al desempeño del sistema, tales como la estabilidad ante variaciones de iluminación, el tiempo de inferencia, las métricas de precisión y la robustez frente a cambios en la distancia y el ángulo de captura. Estos resultados permitieron evaluar la eficiencia y confiabilidad del modelo en escenarios cercanos a su aplicación real.

6. Resultados y Discusión

Como resultado de esta investigación, se desarrolló una herramienta de bajo costo orientada a la detección automática de armas, con potencial de implementación en entornos militares, evidenciando su viabilidad para aplicaciones de vigilancia inteligente en tiempo

real. A continuación, se describen los principales componentes y características del sistema propuesto.

La herramienta desarrollada se estructuró sobre una arquitectura de hardware embebido basada en la tarjeta ESP32-S3, la cual integra un procesador de doble núcleo Tensilica Xtensa LX7 de 32 bits, con una frecuencia de operación de hasta 240 MHz. Este dispositivo dispone de conectividad inalámbrica WiFi de 2.4 GHz y Bluetooth, así como soporte para hasta 8 MB de memoria PSRAM y 16 MB de memoria Flash, lo que lo convierte en una plataforma adecuada para aplicaciones de inteligencia artificial en el borde (edge AI).

Adicionalmente, se empleó un módulo de cámara compatible tipo OV2640, capaz de capturar imágenes con una resolución máxima de 2 megapíxeles (UXGA: 1600×1200 píxeles), con soporte para múltiples formatos de salida como JPEG, RGB565 y YUV. Este módulo incorpora control automático de exposición, balance de blancos y ajuste de ganancia, lo que permite mejorar la calidad de imagen en diferentes condiciones de iluminación. Asimismo, su interfaz DVP (Digital Video Port) facilita la integración directa con el microcontrolador, posibilitando la adquisición de imágenes en tiempo real desde el entorno operativo para su posterior procesamiento en el sistema embebido.

A nivel de software, el sistema se soportó en entornos de desarrollo orientados a TinyML, empleando herramientas como Edge Impulse para el diseño, entrenamiento y despliegue del modelo, así como entornos complementarios como Google Colab y Roboflow para la gestión, preprocesamiento y aumentación del conjunto de datos. El sistema operativo subyacente corresponde al entorno de ejecución del SDK del ESP32, el cual permite la integración eficiente de modelos optimizados de aprendizaje automático en dispositivos con recursos limitados.

A continuación, se presenta el diagrama de bloques del sistema propuesto.

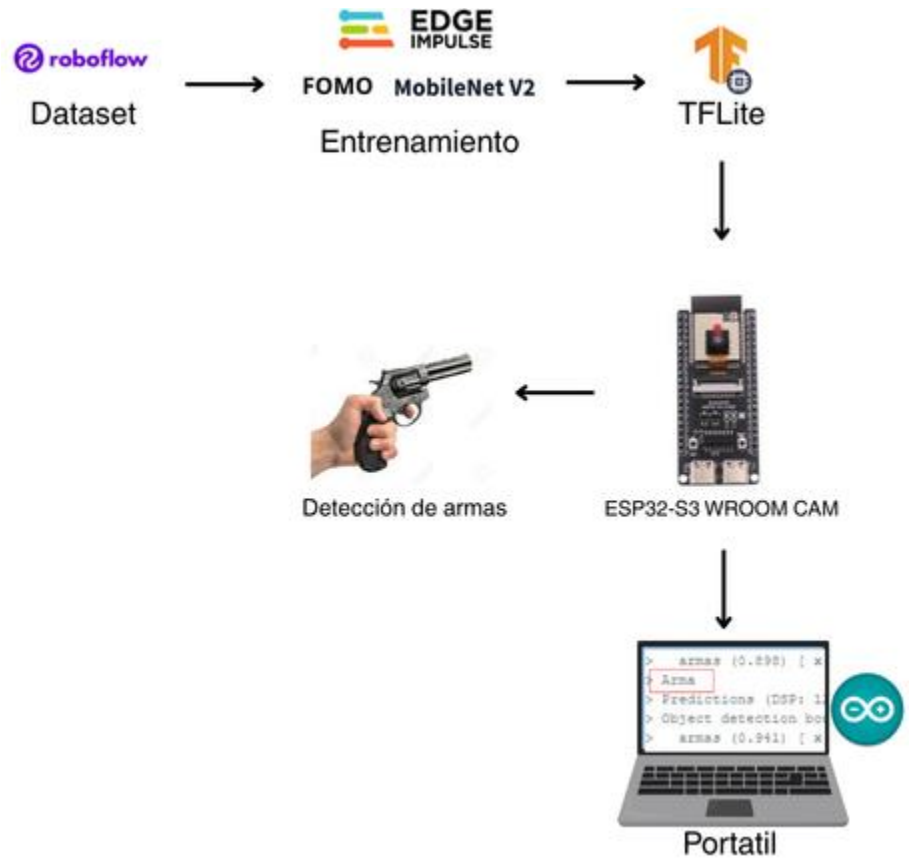


Figura 4. *Diagrama de bloques del sistema.*

El diagrama de bloques representa el flujo general del sistema de detección de armas basado en inteligencia artificial, implementado en dispositivos embebidos. En una primera etapa, las imágenes fueron recolectadas y anotadas para la construcción del *dataset*, empleando la plataforma Roboflow. Posteriormente, este conjunto de datos fue importado y gestionado en Edge Impulse, donde se desarrollaron las etapas de procesamiento, entrenamiento y validación del modelo.

Durante la fase de entrenamiento, se configuró un modelo de detección basado en la arquitectura FOMO (*Faster Objects, More Objects*) combinada con MobileNet, optimizada

para su ejecución en dispositivos con recursos computacionales limitados. Este enfoque permite la identificación eficiente de objetos mediante mapas de calor (*heatmaps*) y probabilidades asociadas a cada clase, reduciendo significativamente la carga computacional frente a métodos tradicionales.

El modelo se configuró con imágenes de entrada de 160×160 píxeles, 150 épocas de entrenamiento y ajuste automático de la tasa de aprendizaje (*learning rate*). Adicionalmente, se aplicaron estrategias de aumentación de datos y ajuste de hiperparámetros dentro del entorno de Edge Impulse, con el objetivo de mejorar la capacidad de generalización del modelo. El sistema fue entrenado para la detección de dos clases principales (“knife” y “pistol”), incorporando además una clase adicional denominada “background”. El conjunto de datos estuvo conformado por 2.808 imágenes anotadas, distribuidas en 70% para entrenamiento, 20% para validación y 10% para pruebas.

Una vez completada la fase de entrenamiento, el modelo optimizado fue exportado desde Edge Impulse en formato TensorFlow Lite (TFLite) y desplegado en la tarjeta ESP32-S3 WROOM. Este dispositivo, en conjunto con su módulo de cámara, permitió la captura y el procesamiento de imágenes en tiempo real. En esta etapa, el sistema ejecutó la detección de armas, identificando objetos de interés como pistolas y cuchillos y generando como salida etiquetas de clase con sus respectivos niveles de confianza.

A continuación, se presentan los resultados obtenidos del modelo desarrollado en Edge Impulse para su implementación en el sistema embebido.

Tabla 1. *Modelo de detección Edge Impulse*

Modelo	**Precision (P)**	**Recall (R)**	**F1-Score**
MobileNetV2 0.1	0.63%	0.51%	55.8%

En la Tabla 1 se presenta el desempeño del modelo entrenado en Edge Impulse, basado en la arquitectura FOMO (*Faster Objects, More Objects*) con MobileNetV2. Los resultados evidencian un desempeño moderado, con una precisión de 0.63 y un recall de 0.51 para las clases distintas del fondo (*non-background*), alcanzando un F1-Score de 0.558.

A pesar de que estas métricas son inferiores frente a modelos más complejos, el modelo fue seleccionado debido a su bajo costo computacional, capacidad de cuantización y compatibilidad con dispositivos embebidos. Estas características permitieron su implementación en la tarjeta ESP32-S3, logrando la ejecución de inferencias en tiempo real bajo restricciones de memoria y procesamiento.

Con el fin de contextualizar el desempeño obtenido, se realizó una comparación con modelos entrenados en otras plataformas.

Tabla 2. *Comparación de Modelos entrenados en plataformas*

Modelo	**Precision (P)**	**Recall (R)**	**mAPp@50**	**mAP@50-95**	**F1-Score**
MobileNetv3_7.5, (Colab)	90.96%	88.80%	92.3%	90.1%	97.2%
RF-DETR (Roboflow)	95.8%	90.0%	95.8%	80.2%	93.0%

Los resultados de la Tabla 2, evidencian que los modelos entrenados en entornos con mayor capacidad computacional presentan métricas significativamente superiores en términos de precisión, recall y mAP. Sin embargo, estos modelos requieren mayores recursos de memoria y procesamiento, lo que impide su implementación en dispositivos embebidos como la ESP32-S3.

En particular, el modelo basado en MobileNetV3 entrenado en Google Colab y el modelo RF-DETR entrenado en Roboflow demostraron un alto desempeño en tareas de detección de objetos. No obstante, su complejidad computacional limita su aplicabilidad en escenarios de *edge computing* con restricciones de hardware.

En contraste, el modelo desarrollado en Edge Impulse, aunque presenta métricas más bajas, logró un balance adecuado entre desempeño y eficiencia computacional, permitiendo su despliegue en un sistema embebido y la ejecución de detecciones en tiempo real.

Durante la ejecución del sistema en condiciones operativas, se analizó el comportamiento de las detecciones a partir de la información generada por el modelo. Cuando no se identificaron objetos de interés, el sistema reportó la ausencia de armas (“Sin arma”); en caso contrario, se generaron las coordenadas espaciales aproximadas del objeto detectado junto con su nivel de confianza.

El análisis de estas salidas permitió observar el comportamiento dinámico de los objetos en la escena. En particular, la variación progresiva de las coordenadas espaciales evidenció el desplazamiento de los objetos dentro del campo visual. Asimismo, la variación en la activación del mapa de calor permitió inferir cambios en la proximidad del objeto respecto a la cámara.

Finalmente, el sistema logró detectar armas en tiempo real, operando como un clasificador por fotograma (presencia/ausencia de arma). No obstante, se evidenciaron variaciones en los niveles de confianza entre detecciones consecutivas, especialmente en relación con la clase *background*, lo que puede generar inestabilidad en los resultados. En este sentido, se identifica como línea de mejora la incorporación de estrategias de análisis temporal, tales como suavizado de predicciones o integración de múltiples fotogramas, con el fin de aumentar la robustez del sistema y reducir la tasa de falsos positivos.

Experimento 1: Comportamiento del sistema en un dispositivo embebido.

Los resultados obtenidos durante las fases de entrenamiento y validación del modelo de detección de armas permitieron diseñar un experimento orientado a evaluar la viabilidad de implementación y el desempeño del sistema en un dispositivo embebido, específicamente en la tarjeta ESP32-S3. El objetivo principal consistió en analizar el comportamiento operativo del modelo en condiciones cercanas a su aplicación real, considerando las limitaciones propias del hardware, tales como memoria, capacidad de almacenamiento y estabilidad de la inferencia en tiempo real.

Las pruebas se llevaron a cabo en un entorno controlado, manteniendo condiciones relativamente constantes en términos de iluminación, fondo, ángulo de captura y distancia del objeto. En este contexto, la ESP32-S3 demostró ser capaz de ejecutar el modelo de manera eficiente, logrando realizar detecciones en tiempo real en un rango aproximado entre 8 cm y 2 metro. No obstante, durante la ejecución se evidenciaron variaciones en los niveles de confianza entre fotogramas consecutivos, generando comportamientos intermitentes en la detección.

A partir de estos resultados, se implementaron estrategias de post-procesamiento orientadas a mejorar la estabilidad del sistema. Entre ellas, se incluyó el uso de un umbral mínimo de confianza de 0.65, el filtrado de detecciones espurias según su activación espacial y un mecanismo de persistencia temporal basado en la consistencia de detecciones entre múltiples fotogramas consecutivos. Estas estrategias permitieron reducir el efecto de "parpadeo" en las predicciones, logrando una respuesta más estable sin afectar significativamente el rendimiento en tiempo real. Asimismo, se observó que el uso de umbrales más estrictos (por ejemplo, 0.70) contribuye a disminuir los falsos positivos, aunque puede reducir la sensibilidad del sistema en condiciones adversas.

En cuanto al comportamiento espacial de las detecciones, se identificó que la estabilidad en las regiones activadas se relaciona con la arquitectura FOMO, la cual divide la imagen en una cuadrícula fija y genera mapas de activación en lugar de *bounding boxes* tradicionales. Este comportamiento facilita la ejecución en hardware limitado, aunque introduce ciertas restricciones en la precisión espacial de la localización.

Por otra parte, se determinó que factores externos, particularmente las condiciones de iluminación, tienen un impacto significativo en el desempeño del sistema. En escenarios de baja iluminación o con presencia de luz intensa directa, se evidenciaron pérdidas de información visual que afectan la capacidad del modelo para identificar correctamente los objetos de interés.

En términos generales, la ESP32-S3 demostró ser una plataforma viable para la implementación de modelos de detección de objetos basados en técnicas de TinyML, logrando un equilibrio adecuado entre precisión, eficiencia computacional y facilidad de despliegue. Los resultados obtenidos evidencian que el desempeño del sistema no depende

únicamente del modelo de inteligencia artificial, sino también de las condiciones del entorno y de las estrategias de procesamiento aplicadas.

A continuación, se presenta evidencia experimental del funcionamiento del sistema de detección de armas en tiempo real sobre la ESP32-S3, donde se observa la identificación de objetos con niveles de confianza superiores a 0.80.

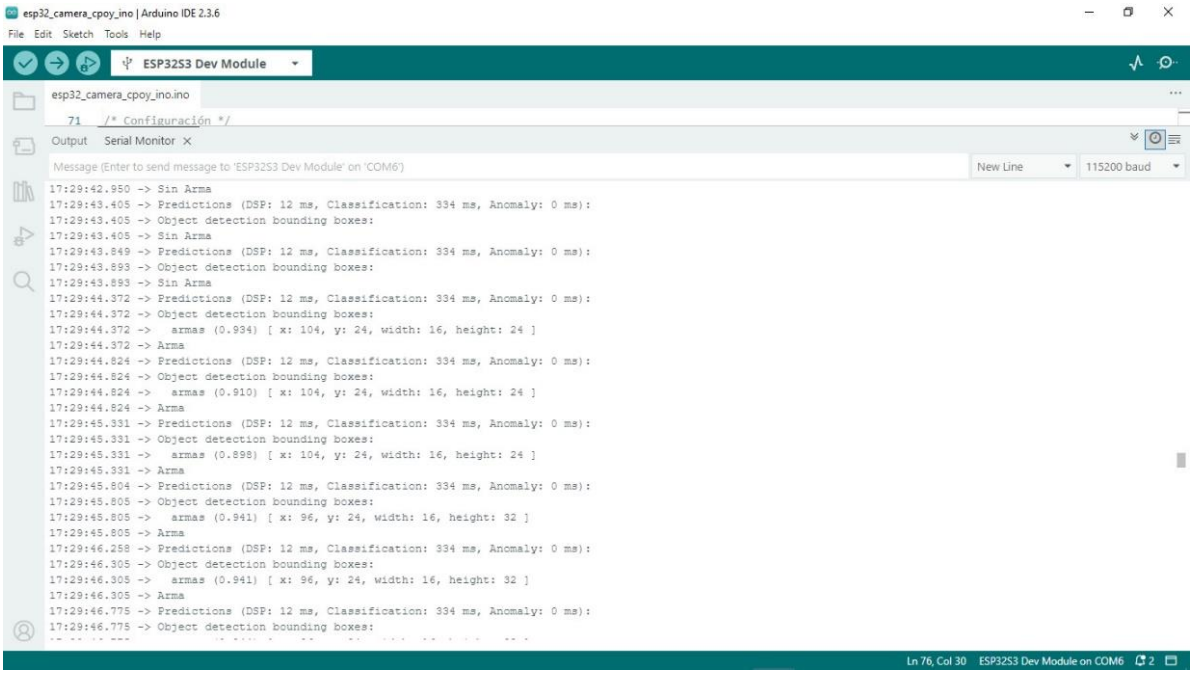
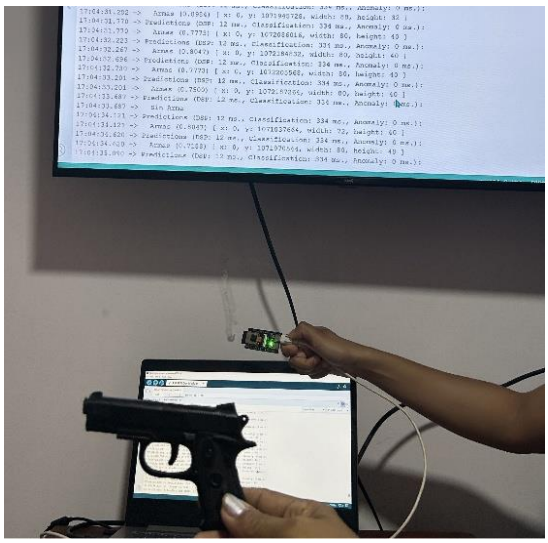


Figura 5. *Prueba del modelo en entorno de desarrollo (Arduino).*



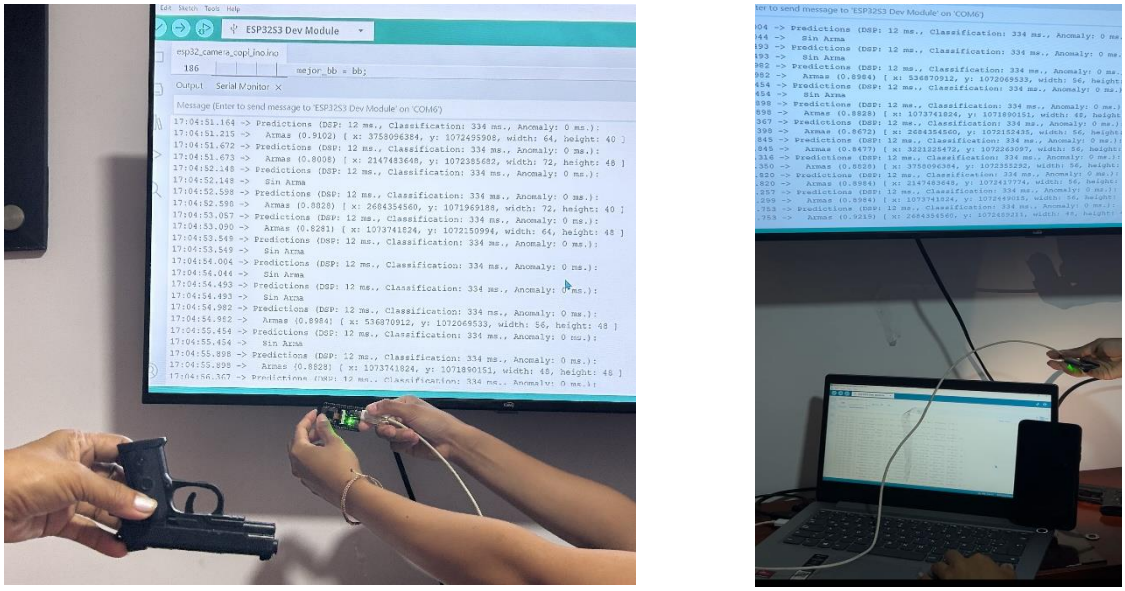


Figura 6 *Ejecución del modelo en condiciones de prueba.*

7. Conclusiones

El desarrollo del sistema de seguridad perimetral basado en técnicas de TinyML permitió demostrar la viabilidad de implementar modelos de detección de armas en tiempo real sobre dispositivos embebidos de bajo costo, específicamente en la tarjeta ESP32-S3. El modelo propuesto, basado en la arquitectura FOMO (*Faster Objects, More Objects*) combinada con MobileNetV2 y desarrollado en Edge Impulse, alcanzó métricas de desempeño moderadas (Precisión: 0.62, Recall: 0.51 y F1-Score: 0.558), logrando operar en un rango de detección entre 8 cm y 1 metro bajo condiciones adecuadas de iluminación.

Un hallazgo relevante de esta investigación es la validación del compromiso inherente entre desempeño y eficiencia computacional en sistemas de *edge AI*. Si bien existen modelos con mayor precisión en entornos de alto rendimiento, estos no son viables en dispositivos embebidos debido a sus elevados requerimientos de memoria y procesamiento. En este sentido, el modelo implementado representó un equilibrio funcional que permitió su ejecución en tiempo real, cumpliendo con las restricciones del hardware sin comprometer completamente la capacidad de detección.

Asimismo, los resultados evidencian que el desempeño del sistema no depende exclusivamente de la arquitectura del modelo, sino también de factores externos como las condiciones de iluminación, la distancia y el ángulo de captura, así como de la calidad y diversidad del conjunto de datos. En particular, se identificaron limitaciones relacionadas con la variabilidad en los niveles de confianza entre detecciones consecutivas, lo que puede generar inestabilidad en la salida del sistema. Esto pone de manifiesto la necesidad de complementar los modelos de detección con estrategias adicionales de procesamiento que incrementen la robustez en escenarios reales.

Desde una perspectiva aplicada, este trabajo aporta evidencia significativa sobre la implementación de inteligencia artificial en sistemas embebidos para aplicaciones de seguridad perimetral, demostrando que es posible desarrollar soluciones funcionales con recursos limitados. Sin embargo, también resalta las restricciones actuales del enfoque TinyML en tareas de detección complejas, especialmente cuando se requiere alta precisión y confiabilidad en entornos dinámicos.

8. Trabajos futuros

A partir de los resultados obtenidos y las limitaciones identificadas, se plantean las siguientes líneas de trabajo futuro:

- ***Mejoramiento del conjunto de datos**:* Ampliar el *dataset* incorporando mayor diversidad de escenarios, condiciones de iluminación, ángulos de captura y tipos de armas, así como técnicas avanzadas de aumentación de datos, con el fin de mejorar la capacidad de generalización del modelo.
- ***Optimización del modelo:*** Explorar arquitecturas más eficientes y técnicas de compresión de modelos, tales como *quantization-aware training*, *pruning* y *knowledge distillation*, que permitan incrementar el desempeño sin afectar significativamente el consumo de recursos.
- ***Análisis temporal de detecciones**:* Implementar enfoques basados en secuencias de imágenes (por ejemplo, filtrado temporal, ventanas deslizantes o modelos ligeros recurrentes) para reducir la inestabilidad entre fotogramas y disminuir la tasa de falsos positivos.
- ***Evaluación en múltiples plataformas embebidas**:* Extender la validación del sistema a otros dispositivos con capacidades de cómputo superiores (como ESP32-S3 con

aceleradores, placas con NPU o SBCs), con el fin de analizar el comportamiento del modelo en diferentes configuraciones de hardware.

- ***Integración con sistemas de alerta****:* Incorporar mecanismos de notificación en tiempo real (IoT), como envío de alertas a aplicaciones móviles o sistemas centrales de monitoreo, que permitan una respuesta oportuna ante la detección de amenazas.
- ***Validación en entornos reales****:* Realizar pruebas en escenarios operativos reales no controlados, evaluando el desempeño del sistema frente a condiciones adversas, ruido visual y múltiples objetos en escena.

9. Referencias bibliográficas

Agudelo, J. (2025). *Perspectivas de seguridad en Colombia: El desafío de los drones*. Ejército Nacional de Colombia.

Ángel Rojo, J. (2025). *Sistemas automatizados para la detección temprana de amenazas*. Revista de Tecnologías de Seguridad, 12(2), 45-60.

Arduino. (2024). *Arduino IDE Documentation*. https://www.arduino.cc/en/software

Banbury, C., Reddi, V. J., Lam, M., Fu, W., Beyne, W., Ali, A., ... & Janapa Reddi, V. (2020). *Benchmarking TinyML Systems: Challenges and Direction*. arXiv preprint arXiv:2003.04818.

Beltrán, M. (2025). *Factores críticos en la vigilancia militar perimetral*. (Comunicación personal, 15 de febrero de 2025).

Beltrán Escobar, M. (2024). *Limitaciones de la vigilancia convencional en infraestructuras críticas*. Editorial Seguridad Integral.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2021). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. arXiv. https://doi.org/10.48550/arXiv.2207.02696

Boyle, J., Moosmann, J., Baumann, M., Heo, J., & Magno, M. (2024). *DSORT-MCU: Small object detection for microcontrollers in Edge AI*. IEEE Journal on Emerging and Selected Topics in Circuits and Systems.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). *End-to-End Object Detection with Transformers*. In European Conference on Computer Vision (pp. 213-229). Springer.

Chandola, V., Banerjee, A., & Kumar, V. (2021). *Anomaly detection: A survey*. ACM Computing Surveys (CSUR).

Conti, F., Garofalo, A., Rossi, D., Pullini, A., & Benini, L. (2020). *A 1.3 TOPS/W @ 32nm ASIP for Neural Network inference in ultra-low-power embedded systems*. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.

Corte Suprema de Justicia. (2024). *Sentencia sobre el atentado a la Escuela General Santander*. Sala de Casación Penal.

Critical Threats Project. (2026). *Iran-US Tensions: Report on military infrastructure operations*. AEI.

Defensoría del Pueblo. (2025). *Alerta temprana sobre el uso de artefactos no convencionales en el Cauca*. Informe No. 045-25.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. ICLR.

Edge Impulse. (2025). *Edge Impulse Documentation: Embedded machine learning*. https://docs.edgeimpulse.com/

Ejército Nacional de Colombia. (2017). *Manual de Fundamentos del Ejército (MFE 3-37): Protección*.

Espressif Systems. (2023). *ESP32-S3 Series Datasheet v1.6*. https://www.espressif.com/

Fuerza Aeroespacial Colombiana. (2025). *Informe oficial sobre el ataque a la Escuela Militar de Aviación Marco Fidel Suárez*.

Ghosh, S., Das, N., Das, I., & Rakshit, S. (2022). *Deep Learning for Object Detection: A Comprehensive Review.* Journal of Intelligent Systems.

Google. (2023). *Google Colaboratory Frequently Asked Questions*. https://research.google.com/colaboratory/faq.html

Han, S., Mao, H., & Dally, W. J. (2016). *Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding*. ICLR.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv preprint arXiv:1704.04861.

IBM. (2023). *What is Deep Learning?*. IBM Education.

Jewani, S., & Abimannan, S. (2023). *Edge AI: Fundamentals and Implementation*. Springer Nature.

LatticeWork. (2022). *Edge Computing and TinyML: The Future of Security*. Tech Report.

López Andrés, J. (2024). *Implementación de sistemas de videovigilancia de bajo costo con ESP32*. Universidad Nacional.

Marchena, D. (2025). *Perspectivas de seguridad en la Base de Coveñas*. (Comunicación personal, 12 de marzo de 2025).

MIT Sloan School of Management. (2024). *Machine Learning Explained*. Management Insights.

Muñoz, E., Santaquiteria, J., Deniz, O., & Bueno, G. (2025). *Detección de armas ocultas mediante termografía y aprendizaje profundo*. Revista de Visión Artificial.

Naciones Unidas. (2019). *Informe sobre el atentado contra la Escuela de Policía General Santander*. Oficina de Derechos Humanos.

Oñate Miranda, J. (2020). *Nodos inteligentes para detección de armas mediante visión artificial y Raspberry Pi*. (Tesis de grado).

Parrado, A. (2025). *Atentado en Puerto Jordán: Vulnerabilidad de las bases militares*. Diario El Tiempo.

Perspectivas de Seguridad en Colombia. (2025). *Informe anual sobre amenazas asimétricas*.

Plumerai & Espressif. (2023). *Real-time person detection on ESP32-S3*. Plumerai Technical Blog.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. In International Conference on Machine Learning (pp. 8748-8763). PMLR.

Reuters; Associated Press. (2026). *Security threats in Gulf Countries and Africa*. Reuters World News.

Rossi, D., Conti, F., Garofalo, A., & Pullini, A. (2021). *Parallel Ultra-Low-Power Platforms for Next Generation Edge AI*. IEEE.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4510-4520).

Shalby, A., Pavan, S., & Roveri, M. (2024). *StreamTinyNet: Real-time video analysis on the edge*. TinyML Research.

Warden, P., & Situnayake, D. (2019). *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. O'Reilly Media.

Zhang, Y., Li, S., & Liu, X. (2023). *Deep Learning in Computer Vision: Principles and Applications*. Academic Press.

ANEXOS

A continuación, se presenta el enlace al repositorio GitHub, donde encontraran el código fuente del sistema desarrollado, incluyendo el desarrollo del modelo de detección y su despliegue en el dispositivo embebido.

https://github.com/connieferia/Modelo_deteccion_armas