

Deep Computer Vision para la obtención de métricas comerciales en sucursales

Santiago José Bueno Ortiz

Trabajo de grado para optar el título de Ingeniero Mecatrónico

Director

Cesar Hernando Valencia Niño

Doctor en Ingeniería Eléctrica

Universidad Santo Tomás, Bucaramanga

División de Ingenierías y Arquitectura

Facultad de Ingeniería Mecatrónica

2026

Contenido

Introducción.....	7
1. Deep Computer Vision para la obtención de métricas comerciales en sucursales	9
1.1 Planteamiento del problema.....	9
1.2 Justificación	9
1.3 Objetivos	11
1.3.1 Objetivo general	11
1.3.2 Objetivos específicos	11
2. Marco referencial.....	12
2.1 Marco teórico.....	12
2.2 Marco conceptual.....	13
2.3 Marco legal	15
3. Método.....	16
3.1 Establecimiento de la Infraestructura de Comunicación	16
3.2 Procesamiento y Extracción de Métricas con Deep Learning	18
3.2.1 Flujo de personas.....	19
3.2.2 Mapa de calor	20
3.3 Gestión, Registro y Estructuración de la Información	22
3.4 Implementación de Herramientas de Visualización y Despliegue.....	23
4. Resultados.....	23
5. Conclusiones.....	28
Referencias	31

Lista de figuras

Figura 1. <i>Diagrama de conexión del sistema</i>	16
Figura 2. <i>Diagrama de flujo scripts de consulta.</i>	17
Figura 3. Resultados de entrenamiento de modelo.....	18
Figura 4. <i>Detección de personas para medición de flujo</i>	20
Figura 6. <i>Mapa de cobertura y distribución de zonas en sucursal de prueba</i>	21
Figura 7. <i>Detección de personas para mapa de calor</i>	22
Figura 8. <i>Matriz de confusión de modelo para imágenes en sucursal.</i>	23
Figura 9. <i>Matriz de confusión de mapa de calor para imágenes en sucursal.</i>	24
Figura 10. <i>Matriz de confusión para flujo de personas en sucursal</i>	25
Figura 11. <i>Reporte de flujo de clientes</i>	26
Figura 12. <i>Indicador tráfico por sucursal</i>	27
Figura 13. <i>tablero tráfico y tasa de conversión global por sucursal</i>	27
Figura 14. <i>Mapa de calor en punto de venta para una semana</i>	28

Resumen

El comportamiento de clientes en sucursales constituye una métrica relevante para la toma de decisiones operativas, como la asignación de personal y la optimización de la disposición de puntos de información. En este proyecto se implementa un sistema de detección y seguimiento de personas basado en técnicas de visión por computador basada en Deep Learning, orientado a la recolección de métricas comerciales a partir de grabaciones provenientes de cámaras en sucursales de la empresa UNIDROGAS S.A.S. Se estableció una conexión entre las cámaras y un servidor centralizado encargado de ejecutar un modelo de detección entrenado bajo la arquitectura de YOLO, utilizando imágenes reales de las sucursales para mejorar la detección de clientes y la distinción entre clientes y empleados. La información generada es almacenada en el servidor y utilizada para la implementación de herramientas de visualización. Como resultado, se logró medir de forma consistente el flujo de clientes en sucursales, así como el conteo de personas en distintas zonas del establecimiento, permitiendo la generación de mapas de calor que apoyan el análisis del comportamiento espacial de los clientes y la toma de decisiones operativas.

Palabras clave: visión por computador, flujo de clientes, analítica comercial, YOLO

Abstract

Customer behavior in physical branches constitutes a relevant metric for operational decision-making, such as staff allocation and the optimization of information point layouts. This project implements a person detection and tracking system based on Deep Learning-driven computer vision techniques, aimed at collecting commercial metrics from video recordings at UNIDROGAS S.A.S. branches. A connection was established between the cameras and a centralized server responsible for executing a detection model trained under the YOLO architecture, using real images from the branches to improve customer detection and the distinction between customers and employees. The generated information is stored on the server and used for the implementation of visualization tools. As a result, it was possible to consistently measure customer flow within the branches, as well as person counts in different areas of the establishment. This allowed for the generation of heatmaps that support the analysis of customers' spatial behavior and aid in operational decision-making.

Keywords: Computer vision, Customer flow, Retail analytics, YOLO.

Glosario

Computer Vision: área de la inteligencia artificial que entrena a las computadoras para interpretar y comprender el mundo visual [1].

Deep Learning: subcampo del aprendizaje automático basado en redes neuronales artificiales con múltiples capas de procesamiento [2].

Fine-tuning: Proceso de transferencia de aprendizaje que consiste en tomar un modelo previamente entrenado en un conjunto de datos masivo y reentrenarlo ligeramente con datos nuevos y específicos [2].

Inferencia: proceso en el cual un modelo de Inteligencia Artificial ya entrenado se ejecuta sobre datos nuevos (como el video en tiempo real de las sucursales) para generar una predicción o detección [3].

Mapas de Calor: representación gráfica de datos donde los valores individuales contenidos en una matriz se representan como colores [4].

NVR (Network Video Recorder): dispositivo de red que recibe y graba flujos de video provenientes de cámaras IP [5].

Redes Neuronales Convolucionales (CNN): tipo de red neuronal de aprendizaje profundo diseñada para procesar datos estructurados en forma de rejilla [6].

Seguimiento de múltiples objetos: tarea de visión por computador que analiza videos para identificar y realizar seguimiento a objetos pertenecientes a una o más categorías [7].

YOLOv8: arquitectura de red neuronal convolucional diseñada para la detección de objetos en tiempo real [8].

Introducción

En el sector retail actual, la capacidad de transformar el comportamiento de los clientes en datos accionables es un factor que mejora significativamente la eficiencia operativa de los establecimientos. No obstante, en las sucursales de UNIDROGAS S.A.S., se identificó una carencia de herramientas automatizadas para medir el flujo y comportamiento espacial de los visitantes. La toma de decisiones sobre la asignación de personal y la disposición de puntos de información se realizaba principalmente de forma empírica, lo que generaba cuellos de botella y una distribución ineficiente del talento humano. Ante este escenario, surge la pregunta de investigación: ¿Cómo implementar un sistema que permita medir de forma precisa y automatizada el flujo de personas en las sucursales de UNIDROGAS S.A.S.?

El propósito central de esta investigación es desarrollar e implementar un sistema de monitoreo inteligente, fundamentado en visión por computador, que automatice la recolección y el análisis de métricas de flujo de clientes en las sucursales de UNIDROGAS S.A.S. El proyecto no se limita al conteo de personas, sino que proporciona una herramienta analítica capaz de segregar datos entre clientes y colaboradores. Esto permite a la dirección de la compañía transformar registros de video pasivos en información estratégica para optimizar la gestión del talento humano, mejorar la disposición del mobiliario comercial y garantizar una atención al cliente basada en la demanda real detectada en el punto de venta.

El desarrollo del proyecto fue conformado por las siguientes etapas:

- *Configuración de la infraestructura de comunicación:* la etapa consistió en el desarrollo de la arquitectura de comunicación entre las cámaras IP instaladas en las sucursales y el servidor encargado del procesamiento remoto, evaluando los

protocolos de transmisión más adecuados para asegurar un flujo de video estable y continuo.

- *Extracción de métricas de video:* se integró el modelo YOLOv8 como componente de procesamiento visual, permitiendo la ejecución de procesos de detección y tracking, con un ajuste orientado a mejorar la discriminación entre perfiles de clientes y personal interno.
- *Gestión y registro de información:* una vez procesados los datos, se realizó la sinterización de la información en el servidor. Donde se evaluaron métodos de almacenamiento que permitieran una consulta rápida para la posterior construcción de reportes y mapas de calor.
- *Implementación y Despliegue:* la fase final comprendió la puesta en marcha del sistema para su funcionamiento autónomo. Se verificó que la solución fuera escalable y sostenible para un uso constante en el entorno operativo de la empresa.

El documento se encuentra dividido en cinco capítulos, el primero presenta el planteamiento del problema y la justificación, el segundo expone el marco referencial que sustenta la investigación, el tercero describe a detalle la metodología implementada, y el cuarto evidencia los resultados obtenidos y el quinto presenta las conclusiones obtenidas y las sugerencias de mejora para futuras versiones del proyecto.

1. Deep Computer Vision para la obtención de métricas comerciales en sucursales

1.1 Planteamiento del problema

En la gestión moderna de establecimientos comerciales, como los liderados por UNIDROGAS S.A.S., la eficiencia en el servicio al cliente es la variable crítica que determina la competitividad. Actualmente, la empresa dispone de sistemas de videovigilancia convencionales; sin embargo, estos operan de manera pasiva, limitándose al registro de eventos de seguridad sin extraer valor analítico para la operación comercial. El problema con los métodos actuales radica en la incapacidad de analizar con precisión el comportamiento del consumidor, lo que deriva en una toma de decisiones basada en percepciones subjetivas y no en la demanda real detectada en el punto de venta.

Este proyecto busca desarrollar un procedimiento robusto que integre la captura de video con técnicas de visión por computador para la obtención de métricas comerciales específicas.

Partiendo de lo anterior, el estudio se conduce bajo la siguiente interrogante: ¿Cómo se puede implementar un sistema de visión por computador para automatizar la medición del flujo peatonal y optimizar la distinción entre perfiles de usuarios para mejorar la toma de decisiones operativas en las sucursales de UNIDROGAS S.A.S.?

1.2 Justificación

La implementación de este sistema de visión por computador surge de la necesidad de evolucionar la infraestructura de seguridad de UNIDROGAS S.A.S. hacia una herramienta de inteligencia operativa. Actualmente, El material de las cámaras de seguridad de las sucursales permanece como un recurso estático, útil únicamente para auditorías de

seguridad, lo que representa una pérdida de oportunidad en la recolección de métricas comerciales.

El valor técnico de este trabajo reside en la creación de un puente entre el hardware de vigilancia y algoritmos de aprendizaje profundo (Deep Learning). Al integrar la arquitectura YOLOv8, se resuelve la limitación de los métodos de conteo manuales o empíricos, los cuales son propensos al error humano y no permiten una segmentación clara entre la actividad de los colaboradores y el comportamiento de los clientes. Esta distinción es fundamental para obtener una "pureza de datos" que sirva de base para cualquier análisis de eficiencia posterior.

Desde una perspectiva de gestión, el proyecto dota a la organización de una visibilidad sin precedentes sobre la dinámica espacial de sus puntos de venta. La capacidad de generar mapas de calor y reportes de afluencia automatizados permite identificar, de manera técnica y objetiva, las zonas de mayor interés y los periodos de saturación en el servicio. Sin esta intervención, la empresa continuaría gestionando sus recursos basándose en percepciones generales, lo que dificulta la capacidad de respuesta ante cambios en los hábitos de consumo de los clientes.

Finalmente, esta solución se justifica por su escalabilidad y bajo impacto en costos de hardware, ya que aprovecha las cámaras instaladas. El proyecto no solo entrega una herramienta funcional inmediata, sino que establece un estándar tecnológico para el procesamiento de video en la nube dentro de la compañía, alineándose con las tendencias de transformación digital y garantizando que la toma de decisiones futura esté respaldada por evidencia analítica sólida y auditable.

1.3 Objetivos

1.3.1 Objetivo general

Implementar un modelo de detección y seguimiento de personas basado en visión por computador para la recolección de métricas comerciales, haciendo uso de grabaciones almacenadas en cámaras de sucursales estratégicas de la empresa UNIDROGAS S.A.S.

1.3.2 Objetivos específicos

- Analizar el estado actual, la ubicación y la cobertura de las cámaras de videovigilancia en las sucursales, con el fin de garantizar que su posición y capacidad técnica sean adecuadas para la recolección de datos requeridos por el modelo de detección.
- Desarrollar e implementar un método estable y seguro para establecer la conexión entre las cámaras de videovigilancia y servidor remoto donde se realizará el procesamiento de video.
- Implementar modelos de detección y seguimiento de personas aplicados a los videos capturados, con el fin de analizar el flujo de clientes, su permanencia en distintas zonas y la tasa de conversión de la sucursal.
- Implementar y evaluar el sistema completo de análisis de video en un entorno real, con almacenamiento de resultados y métricas en base de datos, para generar indicadores de desempeño y eficiencia en la atención de la sucursal.

2. Marco referencial

2.1 Marco teórico

La analítica comercial en entornos físicos ha evolucionado progresivamente desde métodos de observación manual y conteo empírico hacia sistemas automatizados basados en visión por computador y aprendizaje profundo (Deep Learning). Este cambio de paradigma ha permitido procesar grandes volúmenes de datos visuales de forma continua, reduciendo la dependencia de la intervención humana y habilitando la extracción de patrones de comportamiento a partir de imágenes y secuencias de video [1]

En el núcleo de estas tecnologías se encuentran las Redes Neuronales Convolucionales (Convolutional Neural Networks, CNN), un tipo de arquitectura de aprendizaje profundo diseñada para procesar datos estructurados en forma de rejilla, como las imágenes digitales. Las CNN permiten la identificación jerárquica de características visuales, desde bordes y texturas hasta formas complejas, lo que las convierte en el estándar para tareas de detección y clasificación de objetos en visión por computador [6]

Sobre esta base conceptual se desarrollan los modelos modernos de detección de objetos en tiempo real. Entre ellos, la familia de modelos YOLO (You Only Look Once) se caracteriza por realizar la detección y clasificación de objetos en un único paso de inferencia, lo que reduce significativamente la latencia del sistema. En particular, YOLOv8, desarrollado por Ultralytics, introduce mejoras en eficiencia computacional y precisión, permitiendo su uso en aplicaciones donde el procesamiento de video en tiempo real es un requisito crítico, como el monitoreo de sucursales comerciales mediante servidores remotos [8]

No obstante, la detección de personas constituye solo una parte del problema. Para la obtención de métricas de flujo y comportamiento es necesario complementar la detección

con técnicas de seguimiento de múltiples objetos (Multiple Object Tracking, MOT). Estas técnicas asignan identificadores persistentes a cada detección, lo que permite reconstruir trayectorias individuales y evitar conteos duplicados en secuencias de video. El seguimiento robusto resulta especialmente relevante en entornos de retail, donde la oclusión parcial, la variabilidad de iluminación y el tránsito simultáneo de múltiples personas pueden afectar la precisión del sistema [7].

La efectividad de los sistemas de detección y seguimiento en escenarios reales también depende de la infraestructura técnica que soporta la captura y transmisión del video. El procesamiento de flujos de video en tiempo real está condicionado por factores como la latencia de red, la estabilidad del protocolo de transmisión RTSP (Real Time Streaming Protocol) y la capacidad de cómputo disponible, ya sea en el borde (Edge Computing) o en servidores centralizados. Una transmisión inestable puede generar pérdida de fotogramas, lo que impacta negativamente la continuidad del seguimiento y la confiabilidad de las métricas obtenidas.

Finalmente, los datos derivados del seguimiento se traducen en métricas visuales de alto valor analítico, como los mapas de calor. Estas representaciones permiten identificar zonas de alta y baja densidad de presencia dentro de la sucursal, facilitando el análisis del comportamiento espacial de los clientes y transformando el registro de video pasivo en un activo estratégico para la gestión operativa y del talento humano.

2.2 Marco conceptual

Para la correcta comprensión del presente proyecto, resulta necesario definir los principales conceptos que sustentan su desarrollo. El término visión por computador, también denominada visión artificial o computer vision, se refiere a la rama de la inteligencia artificial

que emplea redes neuronales profundas para permitir que los sistemas computacionales interpreten y extraigan información de alto nivel a partir de imágenes y videos digitales. En este contexto, los modelos aprenden representaciones complejas del entorno visual que permiten detectar, clasificar y realizar seguimiento de objetos en secuencias de video.

La inferencia corresponde a la fase operativa del sistema, en la cual un modelo previamente entrenado procesa nuevos datos de entrada, como el video capturado en tiempo real por las cámaras de las sucursales, para generar predicciones sobre la localización y categoría de las personas detectadas, así como su clasificación como clientes o empleados.

La comunicación entre las cámaras IP y el servidor de procesamiento se realiza mediante el protocolo RTSP, un estándar ampliamente utilizado para la transmisión y control de flujos multimedia en tiempo real. Este protocolo permite la recepción continua de video, condición necesaria para la ejecución de algoritmos de detección y seguimiento de personas.

La precisión del sistema está directamente influenciada por la calidad del dataset utilizado durante la fase de entrenamiento. El dataset se define como el conjunto de imágenes etiquetadas que representan las condiciones reales de las sucursales, incluyendo variaciones de iluminación, disposición espacial y perspectivas de cámara. Un dataset representativo permite ajustar el modelo a las características específicas del entorno operativo de UNIDROGAS S.A.S., mejorando su desempeño durante la inferencia.

Los resultados generados por el sistema se integran en un dashboard o tablero de control, el cual actúa como una interfaz gráfica que consolida las métricas obtenidas, tales como conteos de personas, flujos de entrada y salida y mapas de calor. Este tablero facilita la interpretación de la información por parte de los tomadores de decisiones, permitiendo convertir datos técnicos en insumos operativos y estratégicos.

2.3 Marco legal

La implementación de sistemas de visión por computador en entornos comerciales en Colombia se encuentra regulada por el marco normativo de protección de datos personales, cuyo objetivo es salvaguardar los derechos fundamentales de los ciudadanos. La norma principal es la Ley 1581 de 2012, la cual establece el régimen general de protección de datos personales y define la imagen de una persona como un dato personal susceptible de tratamiento, en la medida en que permita su identificación.

Esta ley exige que el tratamiento de datos personales cumpla con principios como la finalidad, la necesidad y la proporcionalidad, lo que implica que la recolección de imágenes mediante sistemas de videovigilancia debe estar orientada a objetivos legítimos previamente informados, como la optimización operativa y la seguridad, y no exceder el alcance necesario para dichos fines [9].

De manera complementaria, la Circular Externa 002 de 2015 de la Superintendencia de Industria y Comercio establece lineamientos específicos para el uso de sistemas de videovigilancia, incluyendo la obligación de informar a los titulares de los datos mediante avisos visibles, la adopción de políticas internas de tratamiento de la información y la implementación de medidas de seguridad que garanticen la confidencialidad e integridad de las grabaciones.

Asimismo, el sistema debe respetar el derecho fundamental al habeas data, consagrado en el Artículo 15 de la Constitución Política de Colombia, el cual faculta a los ciudadanos para conocer, actualizar y rectificar la información que sobre ellos repose en bases de datos públicas o privadas. En consecuencia, la implementación del sistema propuesto debe asegurar mecanismos que permitan atender solicitudes relacionadas con el tratamiento de datos personales, garantizando el cumplimiento del marco legal vigente.

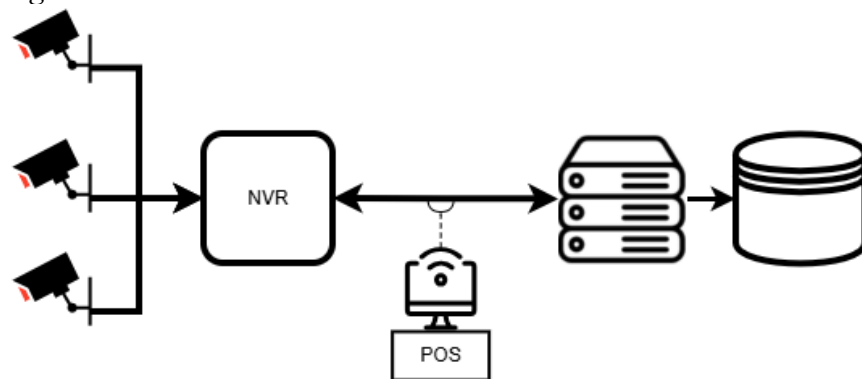
3. Método

El método se divide en cuatro fases principales que garantizan la integridad del flujo de datos y la precisión del modelo de visión por computador.

3.1 Establecimiento de la Infraestructura de Comunicación

El proceso técnico inició con la configuración de un enlace entre las cámaras de la sucursal y el servidor de procesamiento remoto, utilizando la infraestructura local existente. Para lograr la comunicación, se utilizó un equipo POS ubicado en la sucursal como puente, realizando un enrutamiento de puertos que permitiera redirigir las solicitudes externas hacia el NVR donde se centralizan las señales de video. La *Figura 1* representa la conexión establecida para comunicar el servidor con las cámaras de la sucursal.

Figura 1. *Diagrama de conexión del sistema*

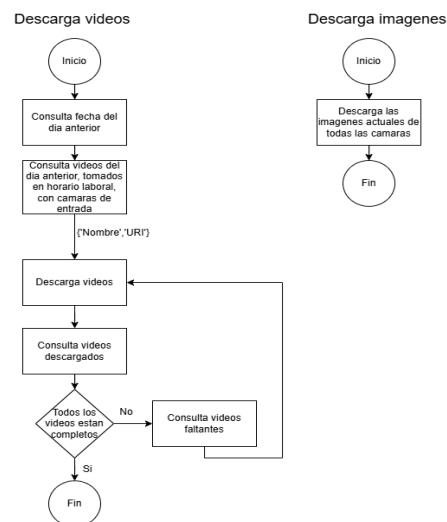


Nota: diagrama elaborado teniendo como referencia el funcionamiento del sistema.

El acceso a los datos se diseñó bajo un esquema de solicitudes HTTP, dependiendo del tipo de métrica a recolectar:

- *Medición de flujo de personas:* para asegurar la precisión en el conteo, se optó por un procesamiento asíncrono mediante solicitudes HTTP al NVR representado en la *Figura 2*. Estas solicitudes permitieron la descarga de archivos de video almacenados en fechas y horarios específicos de las cámaras que apuntan las salidas. Esta decisión técnica se tomó para mitigar el riesgo de pérdida de fotogramas (*frames*) que suele ocurrir en transmisiones en vivo por inestabilidad de red, lo cual podría derivar en conteos erróneos o inconsistentes.
- *Generación de mapas de calor:* a diferencia del conteo, la estimación de zonas concurridas no requiere la fluidez de cada fotograma; por tanto, el sistema solicita una imagen fija de cada cámara en intervalos de cinco minutos, como se muestra en la *Figura 2*. Este procedimiento permitió capturar información de múltiples cámaras de forma simultánea sin saturar el ancho de banda de la sucursal, proporcionando una muestra estadística suficiente para determinar la densidad de presencia en distintas zonas del establecimiento.

Figura 2. Diagrama de flujo scripts de consulta.



Nota: diagrama creado con base al funcionamiento de los scripts de descarga de archivos.

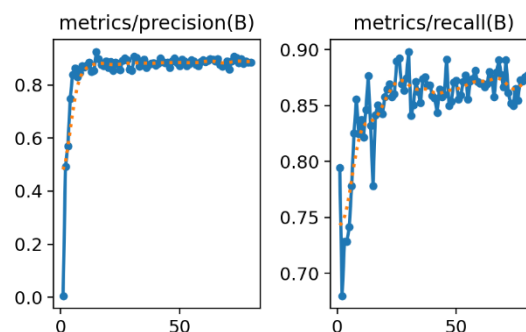
3.2 Procesamiento y Extracción de Métricas con Deep Learning

Para habilitar el procesamiento automático de las secuencias visuales provenientes de las sucursales, fue necesario disponer de un modelo de detección de personas previamente adaptado a las condiciones reales del entorno operativo. Para ello, se empleó la arquitectura YOLO, sometiendo el modelo YOLOv8n a un proceso de ajuste fino (fine-tuning) utilizando imágenes capturadas directamente en las sucursales.

Se seleccionó la variante YOLOv8n debido a su carácter liviano, lo cual resulta adecuado para reducir la carga computacional del sistema al momento de desplegar múltiples instancias del modelo en un servidor centralizado. El proceso de entrenamiento se llevó a cabo en la plataforma Google Colab mediante la librería Ultralytics.

El conjunto de datos utilizado estuvo compuesto por 1.577 imágenes de sucursales, distribuidas en 1.241 imágenes para entrenamiento, 168 para validación y 168 para prueba. Todas las imágenes contaban con anotaciones manuales de bounding boxes. Como configuración inicial, se empleó un tamaño de batch de 16 imágenes y un total de 80 épocas de entrenamiento.

Figura 3. Resultados de entrenamiento de modelo.



Nota: gráfico obtenido de los resultados de entrenamiento presentados por parte de la librería de Ultralytics.

La **Figura 3** muestra los resultados obtenidos del entrenamiento del modelo, estos indicaron que el desempeño del modelo se estabilizaba antes de completar el total de épocas definidas, sin evidenciar mejoras significativas posteriores. Dado que el objetivo del proyecto no se centra en la optimización del modelo, sino en la obtención y procesamiento de métricas a partir del flujo de video, se decidió conservar esta configuración sin realizar entrenamientos adicionales.

Una vez garantizado el flujo de datos y entrenado el modelo, se procedió a extraer las métricas necesarias, este procedimiento se estructuró bajo una lógica de tres capas. Las dos primeras capas fueron comunes para todas las métricas, mientras que la tercera capa varió según el tipo de información a obtener:

- *Detección*: localización de personas en cada fotograma del video.
- *Clasificación*: discriminación entre perfiles de "cliente" y "empleado" basada en características visuales como el uniforme corporativo.
- *Obtención de métricas específicas*: en esta capa final, el procesamiento difirió según la métrica objetivo, empleando secuencias de video continuas para el análisis de flujo de personas o imágenes capturadas en intervalos definidos para la generación de mapas de calor.

3.2.1 Flujo de personas

Para la obtención del conteo de clientes, se implementó un modelo de seguimiento (tracking) aplicado a los clientes detectados. Se definió una línea virtual mediante la ubicación de dos coordenadas en el área de entrada de la sucursal, representando el umbral que debían cruzar los clientes. Cuando una persona atravesaba dicha línea en un sentido

previamente definido, el sistema registraba el evento de cruce junto con su marca temporal. Esta información se almacenaba temporalmente en una estructura de datos para su posterior registro en la base de datos corporativa, permitiendo así el conteo preciso de entradas o salidas sin duplicación de eventos. La *Figura 4* muestra un ejemplo del modelo de detección y seguimiento en funcionamiento.

Figura 4. *Detección de personas para medición de flujo*



Nota: imagen obtenida de las cámaras de videovigilancia de la sucursal en la que se implementó el sistema.

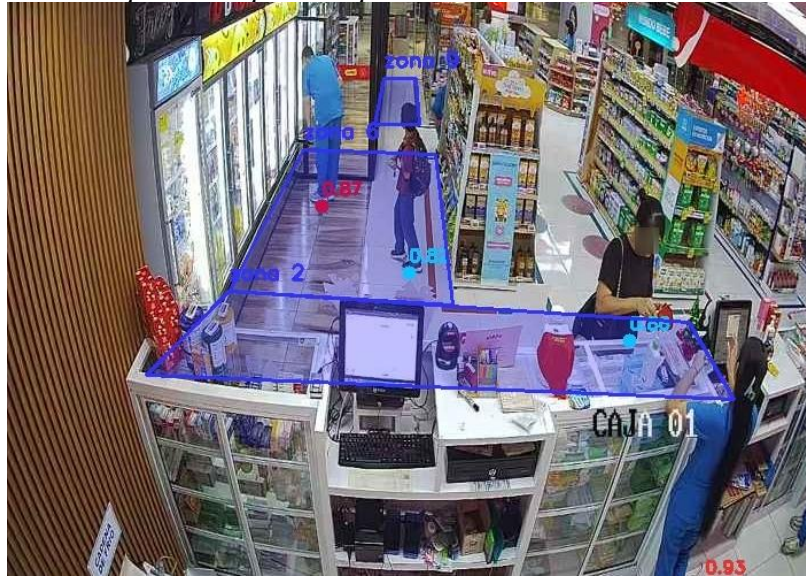
3.2.2 Mapa de calor

Para la creación de los mapas de calor, se establecieron previamente distintas zonas dentro de la sucursal, asignando cada una a la cámara que ofrecía la mejor perspectiva visual. Estas zonas se pueden observar en la *Figura 5*.

Figura 5. *Mapa de cobertura y distribución de zonas en sucursal de prueba*

Nota: diagrama creado a partir del plano de la sucursal en la que se implementó el sistema.

Al recibir una imagen de dichas cámaras, el sistema superpone las zonas definidas mediante coordenadas fijas y realiza el conteo de clientes detectados en cada una de ellas, la *Figura 6* muestra un ejemplo del funcionamiento de este sistema. A diferencia del análisis de flujo, este procedimiento no requirió seguimiento de individuos, sino conteos instantáneos por imagen. Los resultados obtenidos se almacenaron en una lista temporal para su posterior registro en la base de datos, sirviendo como insumo para el análisis del comportamiento espacial de los clientes.

Figura 6. *Detección de personas para mapa de calor*

Nota: imagen obtenida de las cámaras de videovigilancia de la sucursal en la que se implementó el sistema.

3.3 Gestión, registro y estructuración de la información

Los datos generados durante la fase de procesamiento fueron almacenados en una base de datos de manera periódica, de acuerdo con la naturaleza de cada métrica. El flujo de personas se registró mediante un script automatizado que se ejecuta una vez al día, consolidando en una base de datos centralizada toda la información correspondiente a la jornada anterior. Por su parte, los registros asociados a los mapas de calor se almacenaron con una frecuencia horaria, utilizando las imágenes más recientes capturadas por las cámaras asignadas a cada zona. Esta estrategia de almacenamiento diferenciado permitió mantener un equilibrio entre granularidad temporal, trazabilidad histórica y eficiencia en el uso de recursos de almacenamiento.

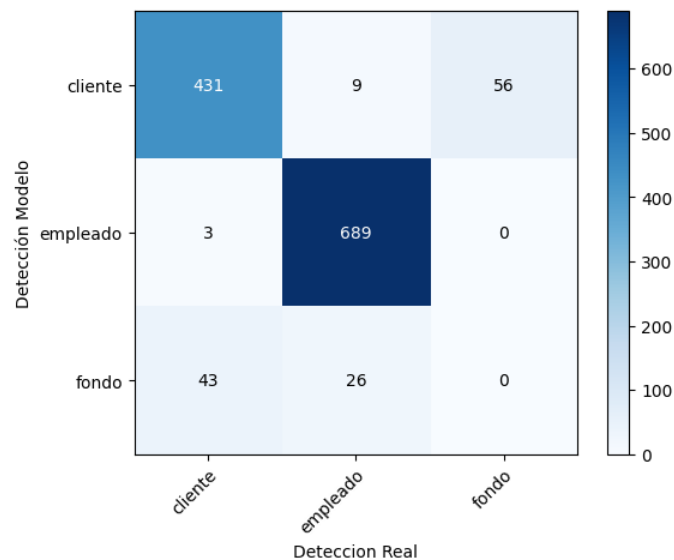
3.4 Implementación de Herramientas de Visualización y Despliegue

La fase final consiste en la interpretación de los datos técnicos en información. Se procedió al diseño y despliegue de tableros de control y mapas de calor. El despliegue se realizó de forma periódica en un entorno de producción, verificando la estabilidad del sistema bajo condiciones de operación continua y asegurando que la interfaz de visualización permita interpretar el flujo peatonal de manera sencilla y eficiente.

4. Resultados

La evaluación del modelo se realizó mediante una validación manual sobre un conjunto de 490 imágenes capturadas en sucursales, en las cuales se verificó la correcta detección de personas y la asignación de etiquetas correspondientes a los perfiles de cliente y empleado. Los resultados de dicha evaluación se demuestran en la *Figura 7*.

Figura 7. Matriz de confusión de modelo para imágenes en sucursal.

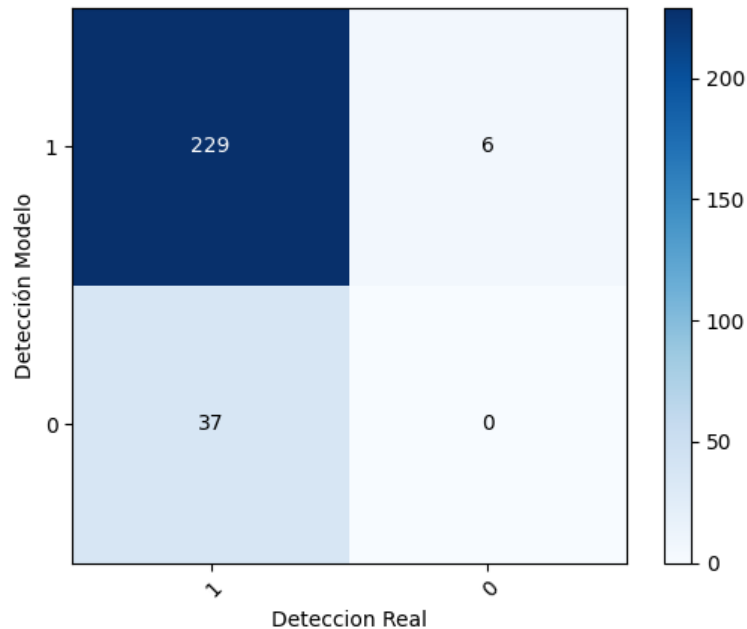


Nota: gráfico creado en base a la comparación manual entre imágenes de prueba tomadas en sucursales y los resultados del modelo para dichas imágenes

Estos resultados corroboran la eficacia del modelo usado, mostrando que al 94,18% de las detecciones se le asigna la etiqueta correcta y que logró detectar correctamente al 93,15% de las personas encontradas en las imágenes de prueba.

Además de la validación de desempeño general del modelo, se realizó una evaluación para cada una de las métricas que tiene el modelo. Para la evaluación de mapas de calor se usaron las mismas 490 imágenes usadas para la evaluación general, pero teniendo en cuenta únicamente las detecciones dentro de las zonas definidas en cada cámara, mientras que la métrica de flujo de personas se evaluó en base a un video de una hora, donde se compararon las salidas reales con las salidas registradas por el modelo, los resultados de dichas validaciones se presentan a continuación.

Figura 8. *Matriz de confusión de mapa de calor para imágenes en sucursal.*

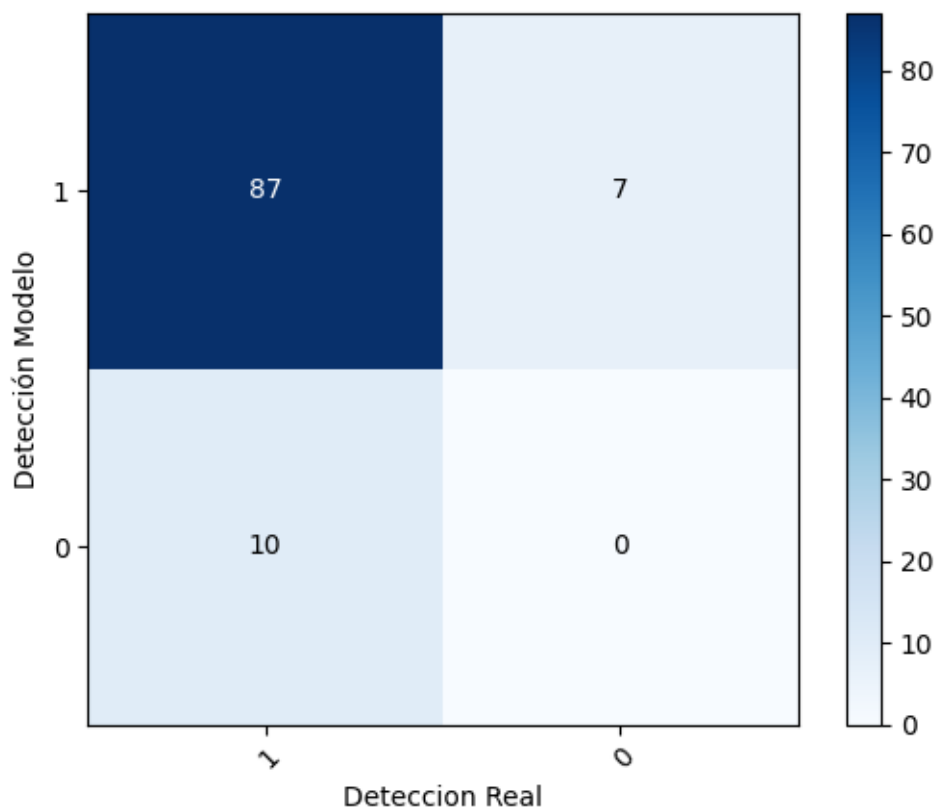


Nota: diagrama creado a partir de comparación entre imágenes tomadas en sucursales y los resultados del modelo para dichas imágenes.

Los resultados mostrados en la *Figura 8* evalúan el modelo únicamente para detectar clientes que se encuentren en zonas definidas para cada cámara en la sucursal. Al separar los resultados a zonas, se pudo observar un aumento de la precisión del modelo, siendo que el 97,45% de las detecciones corresponden correctamente a un cliente.

Finalmente, se realizó la validación del modelo de detección y seguimiento para conteo de personas, haciendo uso de una hora de video de las cámaras ubicadas en las entradas de la sucursal, se obtuvieron los siguientes resultados

Figura 9. *Matriz de confusión para flujo de personas en sucursal*



Nota: Creado a partir de la comparación entre los resultados del modelo y resultados.

reales para el conteo de personas en una hora en sucursal de prueba.

La matriz de confusión representada en la *Figura 9* muestra la eficacia del modelo para contar la cantidad de personas que salen de la sucursal, pudiendo detectar correctamente al 89,69% de las personas totales de la sucursal y siendo el 92,55% detecciones correctas.

Además, se implementaron herramientas de visualización integradas al software corporativo de la empresa, con el fin de facilitar la interpretación, análisis y consulta de la información almacenada, permitiendo una presentación clara y estructurada de los datos obtenidos.

Figura 10. Reporte de flujo de clientes

Fecha inicio: 01/01/2026
Fecha fin: 05/01/2026
Cod. Sucursal:
Nombre Sucursal:
Zona supervision:
Region:
Deposito:
Agrupacion: Hora:
Buscar

Registros 1 a 32 de 32, página 1 de 1
Ir a página: 1

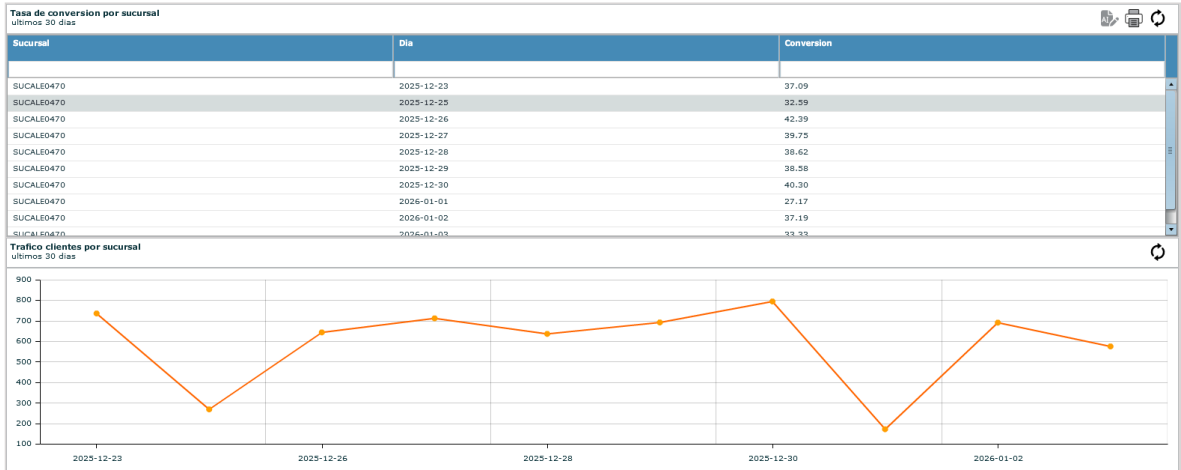
Sucursal	Zona Supervision	Region	Deposito	Fecha	Hora	Cantidad
Todos	Todos	Todos	Todos	Contiene		
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	20:00:00 - 20:59:59	21
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	19:00:00 - 19:59:59	56
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	18:00:00 - 18:59:59	73
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	17:00:00 - 17:59:59	65
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	16:00:00 - 16:59:59	57
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	15:00:00 - 15:59:59	60
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	14:00:00 - 14:59:59	34
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	13:00:00 - 13:59:59	34
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	12:00:00 - 12:59:59	65
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	11:00:00 - 11:59:59	46
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	10:00:00 - 10:59:59	46
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-03	09:00:00 - 09:59:59	19
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-02	20:00:00 - 20:59:59	51
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-02	19:00:00 - 19:59:59	44
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-02	18:00:00 - 18:59:59	91
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-02	17:00:00 - 17:59:59	98
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-02	16:00:00 - 16:59:59	82
ALEMANIA 470 MARKET	SUPERV BMANGA ZONA 5	REGION SANTANDER	DEPOSITO BUCARAMANGA	2026-01-02	15:00:00 - 15:59:59	48
						1440.00

Nota: imagen tomada de reporte de MAIAERP referente al flujo de clientes en sucursales.

La *Figura 10* muestra un reporte creado para observar la cantidad de clientes que salen de sucursales, con la posibilidad de ver la hora de cada salida de forma individual y ver la cantidad total en un periodo definido.

Para los indicadores de tráfico en la sucursal también se realizó una comparación con la cantidad de facturas generadas en el punto de venta para tener la tasa de conversión de dicha sucursal.

Figura 11. Indicador tráfico por sucursal



Nota: imagen tomada de indicador de MAIAERP referente a la tráfico en sucursales

En el indicador que se encuentra en la *Figura 11* se puede observar el histórico de salida de clientes en sucursales, con un muestreo diario, que permite observar el comportamiento en el ultimo mes para cada sucursal donde esté implementado el sistema, al momento de obtener la imagen, solo esta implementado en una sucursal, llevando un tiempo total de 10 días.

Figura 12. tablero tráfico y tasa de conversión global por sucursal



Nota: imagen tomada de indicador de MAIAERP referente a la tasa de conversión de clientes en sucursales.

En los indicadores de la **Figura 12** se muestra de forma detallada una de las métricas objetivo del proyecto, la tasa de conversión en sucursales, donde se compara el flujo de clientes de sucursales con la cantidad de facturas generadas, tiene distintas agrupaciones para poder identificar más fácilmente los factores que afectan las ventas en sucursales y poder encontrar oportunidades de mejora.

Finalmente, el grafico correspondiente al mapa de calor, mostrado en la *Figura 13*, se presentó desde Excel para usar un formato similar a uno que ya se usa en los puntos de venta, con la diferencia de que el mapa de calor antiguo se basa en las facturas en sucursal, por lo que no puede saber la cantidad total de clientes que hay en la sucursal y en que zonas se encuentra.

Figura 13. Mapa de calor en punto de venta para una semana

ZonaID	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00
01	0	0,17	0,49	0,58	0,77	0,68	0,69	0,5	0,67	0,81	0,7	0,89	0,43	0,46
02	0,05	0,15	0,25	0,54	0,5	0,47	0,36	0,53	0,62	0,65	0,51	0,4	0,21	0,21
03	0	0,05	0,06	0,03	0,02	0,05	0,05	0,17	0,17	0,15	0,19	0,12	0,07	0,04
04	0	0,01	0,08	0,09	0,18	0,17	0,07	0,12	0,2	0,1	0,13	0,11	0,05	0,03
05	0	0,01	0,03	0,04	0	0,06	0,03	0,03	0,03	0,02	0,06	0,06	0,01	0,01
06	0,02	0,1	0,04	0,12	0,3	0,05	0,09	0,11	0,12	0,13	0,14	0,12	0,17	0
07	0	0,04	0,03	0,1	0,07	0,03	0,02	0,09	0,06	0,12	0,05	0,08	0,19	0,01
08	0,01	0,01	0,04	0,02	0,05	0,05	0,02	0,06	0,03	0,08	0,04	0,03	0,03	0,03
09	0	0,01	0,06	0,1	0,19	0,1	0,07	0,15	0,12	0,08	0,07	0,07	0,08	0,12
10	0	0	0,01	0,03	0,05	0	0,03	0,11	0,06	0,08	0,09	0,02	0,08	0,05
11	0,05	0,29	0,25	0,73	0,4	0,41	0,46	0,46	0,65	0,61	0,65	0,75	0,64	0,39
12	0	0,01	0,05	0,06	0,03	0,04	0,01	0,01	0,05	0,1	0,05	0,08	0,02	0
13	0,05	0,04	0,02	0,05	0,05	0,06	0,14	0,14	0,05	0,1	0,11	0,11	0,11	0,02
14	0	0,03	0,06	0,08	0,13	0,06	0,06	0,06	0,09	0,05	0,07	0,14	0,12	0,03

Nota: diagrama creado a partir de reporte en MAIAERP referente al mapa de calor en sucursales

5. Conclusiones

El desarrollo del presente proyecto permitió demostrar la viabilidad técnica y operativa de un sistema de analítica de video basado en aprendizaje profundo para la obtención de métricas objetivas sobre el comportamiento de clientes en sucursales comerciales, dando respuesta a la necesidad identificada de contar con información confiable para la toma de decisiones operativas y estratégicas.

En concordancia con el objetivo general, se logró implementar y validar un modelo de detección y clasificación de personas capaz de diferenciar entre clientes y empleados a partir de imágenes reales capturadas en sucursales, utilizando cámaras existentes. Los resultados obtenidos evidencian un desempeño consistente del sistema, con altos valores de precisión y cobertura, lo que confirma su utilidad para la generación de métricas como el flujo de personas y la distribución espacial mediante mapas de calor.

El análisis de los datos recolectados, apoyado en métricas como precisión y sensibilidad, permitió evaluar de manera objetiva el comportamiento del modelo en un escenario real, considerando falsos positivos, falsos negativos y errores de etiquetado. La baja tasa de clasificación errónea observada respalda la capacidad del sistema para discriminar adecuadamente entre perfiles de interés, aspecto fundamental para el uso de la información en contextos comerciales.

Asimismo, la estructuración periódica de los datos en una base centralizada posibilita su posterior análisis estadístico y comparativo, permitiendo identificar patrones temporales y espaciales en el comportamiento de los clientes. La relación entre tablas de resultados y representaciones gráficas, como los mapas de calor, facilita la interpretación de la información y su comunicación a niveles gerenciales no técnicos.

Aunque se identificaron limitaciones asociadas a escenarios de alta oclusión y densidad de personas, estas no comprometen el uso general del sistema y abren oportunidades de mejora para futuras etapas del proyecto, como la optimización del modelo y la ampliación de métricas analizadas.

El desarrollo de este proyecto permitió consolidar y articular los conocimientos adquiridos a lo largo del programa de Ingeniería Mecatrónica, integrando competencias de distintos niveles de formación en un sistema funcional y aplicable a un entorno real. Desde

fundamentos básicos como la programación y el uso del lenguaje Python, hasta áreas más especializadas como el diseño, ajuste e implementación de modelos de inteligencia artificial, así como la estructuración y gestión de bases de datos, el proyecto evidenció una aplicación transversal de los contenidos de la carrera.

Adicionalmente, la integración de fuentes de video provenientes de las sucursales y su procesamiento centralizado en un servidor permitió aplicar conceptos asociados al Internet de las Cosas (IoT), particularmente en lo relacionado con la interconexión de dispositivos, el flujo continuo de datos y la automatización de procesos de análisis. Esta convergencia entre software, procesamiento de información y sistemas distribuidos refleja el enfoque interdisciplinario propio de la ingeniería mecatrónica y su orientación hacia la solución de problemáticas reales mediante tecnologías emergentes.

Referencias

- [1] R. Szeliski, Computer Vision: Algorithms and Applications (2.^a ed.), Springer Nature., 2022.
- [2] I. Goodfellow, Y. Bengio y A. Courville, Deep Learning, MIT Press, 2016.
- [3] F. Chollet, Deep Learning with Python, Manning Publications, 2018.
- [4] L. Wilkinson y M. Friendly, The History of the Cluster Heat Map, The American Statistician, 2009.
- [5] T. Nomura, Introduction to IP Video Surveillance Systems, CRC Press, 2020.
- [6] Y. LeCun, Y. Bengio y G. Hinton, Deep learning, Nature, 2015.
- [7] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri y F. Herrera, Deep learning in video multi-object tracking: A survey, Neurocomputing, 2020.
- [8] G. Jocher, A. Chaurasia y J. Qiu, «Ultralytics YOLOv8 (Versión 8.0.0) [Software].», 2023. [En línea]. Available: <https://github.com/ultralytics/ultralytics>.
- [9] Congreso de la República de Colombia, Ley 1581 de 2012: Por la cual se dictan disposiciones generales para la protección de datos personales., Diario Oficial No. 48.587, 2012.
- [10] Constitución Política de Colombia, Artículo 15 [Derecho a la Intimidad y Habeas Data]. Segunda Edición., Legis., 1991.
- [11] Superintendencia de Industria y Comercio [SIC], Circular Externa 002 de 2015: Tratamiento de datos personales en sistemas de videovigilancia., 2015.