

Smart SSOMA: Design and implementation of an adaptable, comprehensive SSOMA safety management system to reduce reporting times for PPE non-compliance, using machine vision and Edge Computing, applied to industrial and construction environments.

Jose Segundo¹, Javier Garay², Raúl Chávez³, Francisco Rodríguez⁴, Fidel Castro⁵, Roberts Antayhua⁶, Jhojan Espinoza⁷
Universidad Nacional de Ingeniería, Lima, Perú

¹josesegundomanayay06@gmail.com, ²javier.garay.y@uni.pe, ³gianmarco.chavez.c@uni.pe,

⁴francisco.rodriguez.h@uni.pe, ⁵fidel.castro.s@uni.pe, ⁶roberts.antayhua.a@uni.pe,

⁷jhojan.espinoza.c@uni.pe

Abstract—Manual supervision of Personal Protective Equipment (PPE) in Peruvian manufacturing SMEs presents critical limitations: discontinuous monitoring, human dependency, low traceability, and delays in SSOMA report generation, increasing the risk of occupational accidents and regulatory non-compliance. This research addresses the design and implementation of a mechatronic safety management system based on computer vision with YOLOv8 and Edge Computing, aimed at automating the detection of PPE non-compliance and the generation of SSOMA reports aligned with R.M. 050-2013-TR in the Peruvian manufacturing sector. The system integrates real-time video capture via IP cameras using RTSP/ONVIF protocols, image preprocessing, local inference with a custom-trained YOLOv8 model for helmet and reflective vest detection, automated event registration in a PostgreSQL database, and automatic PDF report generation through a FastAPI backend and web dashboard. All processing is executed locally on an Edge device, eliminating cloud dependency and ensuring low-latency response. Experimental evaluation over an 8-hour shift monitoring 35 workers achieved 94.6% detection precision, 92.8% recall, 93.7% F1-score, and an average inference time of 38 ms per frame with 82 ms total system latency. Automated report generation was completed in 40 seconds versus 30 minutes manually, representing a 97.8% reduction. User satisfaction reached 4.7/5. Results confirm that integrating YOLOv8 and Edge Computing into a mechatronic system delivers a viable, efficient, and scalable solution for occupational safety management in resource-constrained industrial environments.

Index Terms—computer vision, YOLOv8, edge computing, occupational safety

I. INTRODUCTION

La gestión de seguridad ocupacional en PYMEs manufactureras representa un desafío crítico en economías en desarrollo. Según la Organización Internacional del Trabajo, más de 2.7 millones de trabajadores fallecen anualmente por

accidentes y enfermedades relacionadas con el trabajo, siendo el incumplimiento del uso de Equipos de Protección Personal (EPP) una de las principales causas contribuyentes [1]. En el Perú, esta problemática se acentúa en las pequeñas y medianas empresas, donde la supervisión depende predominantemente de procesos de inspección manual, generando monitoreo discontinuo, limitada trazabilidad y retrasos significativos en la generación de reportes SSOMA según la R.M. 050-2013-TR. Los avances recientes en visión artificial y Edge Computing han permitido el desarrollo de sistemas automatizados capaces de abordar estas limitaciones. Respecto a arquitecturas de detección de EPP, Sharma. [2] proponen un sistema basado en YOLOv8 desplegado en hardware de bajo costo (Raspberry Pi 4) mediante una API RESTful de microservicios, logrando 8–12 FPS con 120 ms de latencia. Sin embargo, esta baja tasa de cuadros puede ocasionar pérdida de eventos en entornos de movimiento rápido, y el sistema no contempla la generación automática de reportes formales. En un enfoque complementario, Castillo. [3] proponen un sistema automatizado basado en YOLOv11s para el sector de distribución eléctrica, alcanzando 94% de precisión y 92.8% de mAP@50, con documentación automática para reportes de cumplimiento. No obstante, la validación se realizó principalmente con imágenes estáticas y una base de datos reducida de 589 imágenes, limitando su confiabilidad en flujos de video continuo de alta velocidad. Ahmad et al. [4] proponen un sistema integral de monitoreo que combina YOLOv8 con Flask y Firebase para gestión de datos en tiempo real, incorporando autenticación biométrica y detección de proximidad a maquinaria peligrosa. Sin embargo, la fuerte dependencia de servicios en la nube introduce vulnerabilidades de latencia en sitios industriales remotos, haciendo necesaria la adaptación a edge computing

puro para garantizar continuidad operativa ante fallas de red. En estudios comparativos de múltiples arquitecturas, Nath et al. [5] evalúan ocho detectores YOLO (v3-v5) para seis clases de EPP utilizando el dataset CHV, determinando que YOLOv5x obtiene el mejor mAP (86.55%) mientras que YOLOv5s alcanza 52 FPS en GPU. El estudio se limita a entornos de construcción y no considera el procesamiento en dispositivos de borde de baja potencia ni la gestión automatizada de datos SSOMA. Patel. [6] proponen un sistema de monitoreo en tiempo real basado en YOLOv8 para 10 clases de objetos de seguridad, aplicando técnicas avanzadas de aumento de datos que logran mAP@0.5 de 0.857 y F1-score de 0.84. No obstante, el trabajo no aborda la integración con edge computing ni implementa un módulo de gestión que traduzca las detecciones en reportes SSOMA automatizados o bases de datos históricas. A nivel de tesis, Quispe [7] desarrolla un sistema integral de monitoreo de EPP basado en CNN que permite vigilancia automatizada continua 24/7, aunque su efectividad se ve comprometida en condiciones de baja iluminación o presencia de partículas en suspensión, sin profundizar en edge computing ni generación automática de informes. Mamani. [8] proponen un prototipo de monitoreo constante utilizando YOLO con transferencia de aprendizaje sobre una Raspberry Pi 5 integrada con ThingsBoard IoT, alcanzando 74.8% de precisión en pruebas de campo reales. Sin embargo, carece de un módulo integrado para la generación automática de reportes SSOMA según la R.M. 050-2013-TR. Huanca [9] desarrolla un sistema de deep learning para detección de cascos industriales, concluyendo superioridad frente a la inspección humana convencional, pero restringe el análisis exclusivamente a cascos, omitiendo otros componentes críticos del EPP y arquitecturas de edge computing. Respecto a soluciones comerciales, Intenseye [10] ofrece una plataforma avanzada de visión de seguridad que procesa más de 35 tipos de riesgos mediante NVIDIA Jetson Orin NX con integración a PLCs para paradas de emergencia, pero presenta una barrera de entrada elevada para PYMES debido a costos de licenciamiento y requerimientos de infraestructura IT. SafetyMind [11] provee videoanalítica especializada en minería pesada con casos comprobados en Perú, aunque su enfoque vertical limita la adaptabilidad inmediata a manufactura ligera sin reentrenamiento significativo de modelos y no se alinea estrictamente con la R.M. 050-2013-TR. SafeHS.ai [12] implementa una arquitectura plug-and-play con ciclos de inferencia de 0.4 segundos cubriendo más de 25 categorías de riesgo; sin embargo, la precisión en detecciones específicas puede ser inferior a soluciones hiperespecializadas y no genera reportes automatizados para la normativa peruana. Los trabajos revisados se centran principalmente en el desarrollo de modelos de detección de EPP basados en arquitecturas YOLO, logrando altos niveles de precisión en condiciones controladas. No obstante, presentan limitaciones en la validación sobre flujos de video en tiempo real, el uso de procesamiento en edge computing de bajo costo y la integración con sistemas de gestión documental orientados al cumplimiento normativo. De igual manera, las soluciones

comerciales existentes presentan elevadas barreras de entrada para PYMES y escasa adaptación a normativas locales como la R.M. 050-2013-TR. Considerando las observaciones realizadas de los estudios mostrados, este artículo propone el diseño e implementación de un sistema mecatrónico de gestión de seguridad que integra detección automatizada de EPP, procesamiento en edge computing de bajo costo y generación de reportes SSOMA normativos con trazabilidad de eventos, proporcionando una solución operativa, accesible y escalable para PYMES manufactureras peruanas.

II. METODOLOGÍA

De acuerdo a lo propuesto se ha dividido el proyecto en los siguientes bloques:



Fig. 1. Estructura de desglose de trabajo - EDT

A. Subsistema mecánico

El subsistema mecánico del sistema tiene como función principal proveer una plataforma física estable, portátil y ajustable para el posicionamiento óptimo de la cámara IP en entornos productivos simulados. Para su desarrollo se siguió una metodología de diseño ingenieril que comprende levantamiento de requerimientos, diseño conceptual, modelado CAD, fabricación y validación experimental.

Para el desarrollo del sistema mecánico se realizó el levantamiento de requerimientos considerando las características funcionales y físicas de la cámara seleccionada para el proyecto, correspondiente al modelo YC-W542XX SERIES fabricado por ShenZhen YCX Electronics. Esta cámara IP cuenta con resolución 4K de 8MP, conectividad Ethernet y capacidades de visión nocturna mediante iluminación infrar-

roja y luz blanca, características que influyen directamente en las necesidades mecánicas del sistema de soporte.



Fig. 2. Cámara IP a usar

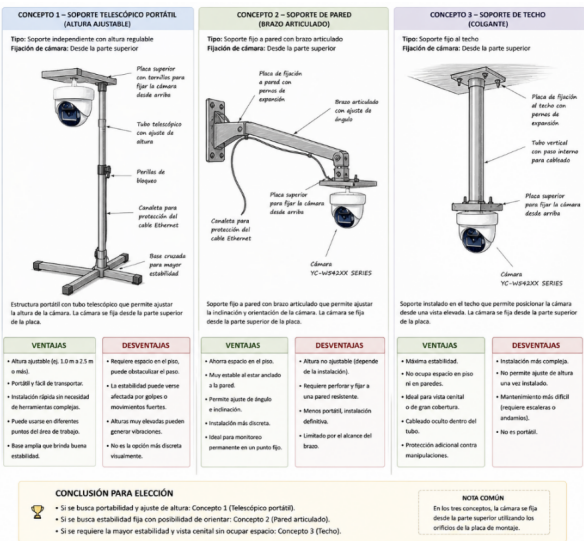


Fig. 3. Esquema de alternativas para soporte de cámara IP

TABLE I
REQUERIMIENTOS DEL SUBSISTEMA MECÁNICO

Requerimiento	Descripción
Ajuste de altura	Permitir modificar la altura de la cámara
Ajuste angular	Permitir variar el ángulo de inclinación
Estabilidad	Minimizar vibraciones y movimientos
Portabilidad	Facilitar el traslado e instalación
Protección de cableado	Organizar y proteger conexiones eléctricas
Resistencia estructural	Soportar el peso de cámara y accesorios
Compatibilidad	Permitir integración con hardware electrónico

Se evaluaron tres conceptos de soporte: soporte telescópico portátil de altura ajustable, soporte de pared con brazo articulado y soporte de techo colgante. El análisis comparativo consideró criterios de portabilidad, estabilidad, facilidad de instalación y adaptabilidad al entorno. El Concepto 1, soporte telescópico portátil, fue seleccionado por ofrecer ajuste de altura de 1.0 m a 2.5 m, instalación rápida sin necesidad de herramientas complejas, base amplia para mayor estabilidad y reutilización en diferentes puntos del área de trabajo, condiciones prioritarias para su aplicación en PYMEs manufactureras donde la flexibilidad operativa es determinante. Los conceptos de pared y techo, si bien ofrecen mayor estabilidad fija, requieren intervención estructural permanente y presentan limitaciones de portabilidad incompatibles con los requerimientos del sistema.



Fig. 4. Estructura mecánica armada

B. Subsistema electrónico

El subsistema electrónico tiene como función principal garantizar la alimentación estable de los dispositivos del prototipo y ejecutar la activación automática de alertas físicas ante eventos de incumplimiento de EPP detectados por el sistema de visión artificial.

1) *Evaluación de alternativas:* Se evaluaron dos alternativas para el sistema de alimentación y alarmas del prototipo:

- Opción 1 – Sistema Low Cost (buzzer y LED):** Solución de bajo costo basada en señales sonoras genéricas mediante buzzer y señales visuales mediante LED. Si bien presenta ventajas de simplicidad y bajo costo, su capacidad de comunicación es limitada, ya

que un sonido genérico puede generar confusión y no transmite claramente el tipo de alerta emitida.

- **Opción 2 – Sistema con parlante activo:** Solución basada en reproducción de mensajes de advertencia mediante parlante controlado por software desde el nodo Edge. Permite emitir alertas específicas y comprensibles para los trabajadores, mejorando la interacción hombre-máquina y la experiencia de supervisión industrial.

Tras evaluar ambas alternativas, se seleccionó la Opción 2 debido a que ofrece mayor claridad comunicativa, mejor percepción de alerta y una integración más adecuada con los objetivos del sistema inteligente de gestión SSOMA.



Fig. 5. Esquema de diseño de circuit o de alimentación

El circuito de alarmas fue diseñado para permitir la activación automática de alertas sonoras a partir de los eventos detectados por el sistema de visión artificial. Debido a que el sistema utiliza una laptop como nodo Edge principal, se implementó una solución basada en reproducción de audio mediante software, evitando la necesidad de circuitos electrónicos complejos o microcontroladores adicionales.

El parlante activo se conecta directamente a la salida de audio de la laptop Edge mediante conexión auxiliar de 3.5 mm, permitiendo que el nodo local reproduzca mensajes de advertencia sin requerir módulos electrónicos externos de potencia.

2) **Funcionamiento del sistema de alarmas:** El flujo de funcionamiento del subsistema electrónico opera completamente mediante control software siguiendo la secuencia:

Cámara IP → YOLOv8 → FastAPI → Pygame → Parlante (1)

Cuando el modelo de inteligencia artificial detecta un incumplimiento de EPP, el backend desarrollado en FastAPI

ejecuta automáticamente una rutina de activación sonora utilizando la librería Pygame, reproduciendo mensajes de advertencia específicos según el tipo de incumplimiento identificado, tales como:

- “Alerta: trabajador sin casco de seguridad, por favor colóqueselo.”
- “Alerta: trabajador sin chaleco de seguridad, cuídate, te espera tu familia en casa.”
- “Advertencia: uso incorrecto de equipo de protección personal.”

La activación sonora continuará mientras el evento permanezca activo o durante un intervalo previamente definido por el sistema, garantizando que el trabajador reciba la alerta de manera oportuna y comprensible.



Fig. 6. Esquema de diseño de circuito de alarmas

C. Subsistema de Software

El subsistema de software integra la captura de video, el procesamiento de inteligencia artificial y la gestión documental automatizada dentro de una plataforma web local basada en Edge Computing. Para su desarrollo se realizó el levantamiento de requerimientos funcionales, el diseño de arquitectura web y el diseño de base de datos.

1) **Requerimientos funcionales:** Se definieron los requerimientos funcionales considerando las necesidades de monitoreo en tiempo real, procesamiento de video mediante visión artificial y automatización de reportes de incumplimientos de EPP, los cuales se presentan en la Tabla II.

2) **Arquitectura web:** La arquitectura web fue diseñada bajo un enfoque modular basado en Edge Computing, integrando captura de video, inferencia IA y gestión web en una misma plataforma local. La cámara IP transmite video mediante protocolo RTSP hacia el servidor Edge desarrollado en FastAPI, el cual actúa como núcleo principal de la aplicación ejecutando inferencia con YOLOv8 y PyTorch sobre los fotogramas capturados con OpenCV. La comunicación en tiempo real entre backend y frontend se implementa mediante WebSockets con Django Channels y Daphne, mientras que Channels-Redis gestiona los eventos asíncronos del sistema. Las principales librerías utilizadas se presentan en la Tabla III.

TABLE II
REQUERIMIENTOS FUNCIONALES DEL SOFTWARE

Requerimiento	Descripción
Captura de video IP	Video desde cámaras IP vía RTSP
Procesamiento IA	Detección de EPP con YOLOv8
Monitoreo en tiempo real	Dashboard web con video procesado
Gestión de alertas	Alertas visuales ante incumplimientos
Registro de eventos	Almacenamiento automático de incidencias
Evidencia visual	Guardado de imágenes de detección
Autenticación	Control de acceso por roles
Generación de reportes	Documentos SSOMA automáticos
Exportación documental	Exportación en PDF y Word
Funcionamiento local	Servidor Edge sin dependencia de nube

TABLE III
PRINCIPALES LIBRERÍAS DEL SUBSISTEMA DE SOFTWARE

Librería	Función
OpenCV	Captura y procesamiento de video
PyTorch	Framework de aprendizaje profundo
Ultralytics	Implementación del modelo YOLOv8
NumPy	Operaciones matriciales sobre imágenes
FastAPI	Servidor local y APIs REST
Daphne	Servidor ASGI para WebSockets
Django Channels	Comunicación bidireccional en tiempo real
Channels-Redis	Gestión asíncrona de eventos
Pygame	Generación de alertas sonoras

3) *Diseño de base de datos:* Se diseñó un modelo relacional orientado a la trazabilidad de eventos, consulta histórica de incidencias y generación automática de reportes SSOMA conforme a la R.M. 050-2013-TR. El modelo contempla once entidades principales descritas en la Tabla IV.

TABLE IV
ENTIDADES DEL MODELO RELACIONAL

Entidad	Descripción
usuarios	Operadores con roles: admin, supervisor, visualizador
trabajadores	Personal de planta (puntos 11–20, R.M. 050)
empresa	Datos fijos organizacionales (puntos 1–5)
camaras	Dispositivos IP registrados en el sistema
detecciones	Eventos automáticos generados por YOLOv8
evidencias	Imágenes capturadas por detección
incidentes	Registro formal RM-050 (puntos 22–31)
medidas_correctivas	Acciones correctivas por incidente (punto 30)
alertas	Notificaciones por canal: pantalla, audio, email
reportes	Documentos SSOMA con flujo de aprobación
configuracion	Parámetros operativos por cámara

La estructura relacional garantiza escalabilidad, integridad referencial y facilidad de mantenimiento, permitiendo la generación automática de reportes SSOMA a partir de los eventos registrados en tiempo real.

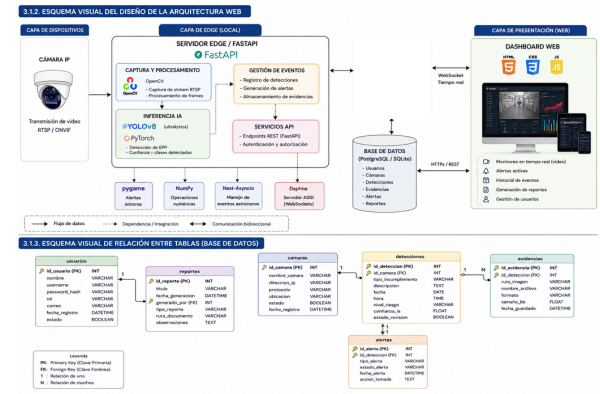


Fig. 7. Estructura de desglose de trabajo - EDT

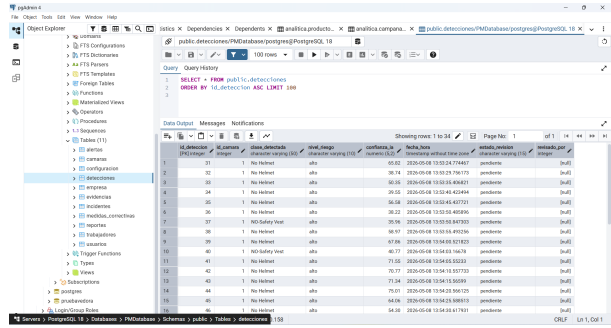


Fig. 8. Estructura de desglose de trabajo - EDT

D. Subsistema de Inteligencia Artificial

El subsistema de inteligencia artificial tiene como función principal ejecutar la detección automática de incumplimientos de EPP en tiempo real mediante procesamiento local de video. Para su desarrollo se configuró un entorno de trabajo orientado a visión artificial y Edge Computing, se definieron las clases de detección y se seleccionó el hardware de procesamiento más adecuado.

1) *Configuración del entorno:* El entorno fue diseñado considerando compatibilidad con YOLOv8, procesamiento de video IP y ejecución local de inferencia utilizando Python como lenguaje principal. La arquitectura considera un servidor local basado en FastAPI para gestionar el procesamiento, la comunicación con el dashboard web y la integración con los módulos de inteligencia artificial. OpenCV gestiona la captura de video desde cámaras IP mediante protocolo RTSP, mientras que YOLOv8 con Ultralytics y PyTorch ejecutan la inferencia en tiempo real. Las principales herramientas del entorno se presentan en la Tabla V.

2) *Definición de clases EPP:* Se definieron las clases de detección considerando los equipos de protección personal más utilizados en entornos industriales y de construcción y su viabilidad de detección visual mediante modelos de visión artificial. Se priorizaron los equipos más representativos ante riesgos de golpes, proyección de partículas y exposición a maquinaria. Las clases definidas se presentan en la Tabla VI.

TABLE V
HERRAMIENTAS DEL SUBSISTEMA DE INTELIGENCIA ARTIFICIAL

Herramienta	Función principal
Python	Lenguaje principal del sistema
FastAPI	Servidor backend y APIs REST
OpenCV	Captura y procesamiento de video
PyTorch	Framework de deep learning
YOLOv8 (Ultralytics)	Modelo de detección de objetos
NumPy	Procesamiento matricial
Pygame	Generación de alertas sonoras

TABLE VI
CLASES EPP DEFINIDAS PARA EL MODELO YOLOv8

Clase	Descripción
Helmet	Trabajador con casco de seguridad
No Helmet	Trabajador sin casco de seguridad
Safety Vest	Trabajador con chaleco reflectivo
No Safety Vest	Trabajador sin chaleco reflectivo

3) *Selección de hardware Edge*: Se evaluaron tres plataformas de hardware Edge para la ejecución del sistema de inferencia en tiempo real, considerando criterios de costo, capacidad de procesamiento, compatibilidad con YOLOv8 y rendimiento para monitoreo SSOMA continuo. El análisis comparativo se presenta en la Tabla VII.

TABLE VII
COMPARACIÓN DE DISPOSITIVOS EDGE EVALUADOS

Hardware	Ventajas	Desventajas
Raspberry Pi 5	Bajo costo, bajo consumo, tamaño compacto	Rendimiento limitado, menor FPS en YOLOv8
NVIDIA Jetson Nano	Aceleración GPU CUDA, optimizado para Edge AI	Mayor costo, recursos limitados para modelos pesados
Laptop económica	Mayor potencia, compatibilidad completa con librerías IA	Mayor consumo energético, menor portabilidad

Raspberry Pi 5, si bien representa una solución económica y compacta, presenta capacidad de procesamiento limitada para ejecutar inferencia en tiempo real con modelos YOLOv8 sobre video de resolución 4K, generando reducciones significativas de FPS y aumento de latencia. NVIDIA Jetson Nano ofrece ventajas por su compatibilidad CUDA y aceleración GPU especializada; sin embargo, puede requerir optimización adicional del modelo para mantener rendimiento estable bajo cargas exigentes.

Para esta implementación se seleccionó la laptop económica como nodo Edge principal, debido a su mayor capacidad de procesamiento, facilidad de desarrollo y mejor desempeño para inferencia en tiempo real, permitiendo integrar FastAPI, OpenCV, PyTorch y YOLOv8 en una sola plataforma sin restricciones de compatibilidad.

E. Subsistema de Redes

El subsistema de redes tiene como función principal garantizar la comunicación estable y eficiente entre la cámara IP, el

nodo de procesamiento Edge y la plataforma web de monitoreo SSOMA. La arquitectura propuesta se basa en una red LAN local que prioriza estabilidad, baja latencia y funcionamiento continuo dentro de un entorno simulado.

1) *Arquitectura de red local*: La arquitectura considera una cámara IP conectada mediante Ethernet al nodo local de procesamiento, el cual ejecuta simultáneamente el sistema de visión artificial, el almacenamiento de eventos y el servidor web local mediante FastAPI. La transmisión de video se realiza mediante protocolos RTSP y ONVIF, permitiendo que OpenCV y YOLOv8 capturen y procesen los fotogramas en tiempo real sin depender de servicios externos o conexión permanente a la nube.

La arquitectura contempla además la conexión de dispositivos cliente dentro de la misma red local, permitiendo que supervisores y usuarios autorizados accedan al dashboard web desde navegadores conectados a la LAN para visualización en tiempo real de video procesado, alertas activas y registros históricos del sistema. Los componentes de la arquitectura se detallan en la Tabla VIII.

TABLE VIII
COMPONENTES DE LA ARQUITECTURA DE RED LOCAL

Componente	Función
Cámara IP	Captura y transmisión de video
Router/Switch	Gestión de comunicación en red local
Nodo Edge local	Procesamiento IA y servidor web
Base de datos	Almacenamiento de eventos y evidencias
Dashboard web	Visualización y supervisión del sistema
Cliente LAN	Acceso de usuarios autorizados

2) *Flujo de funcionamiento*: El flujo general de la arquitectura de red opera en cinco etapas secuenciales:

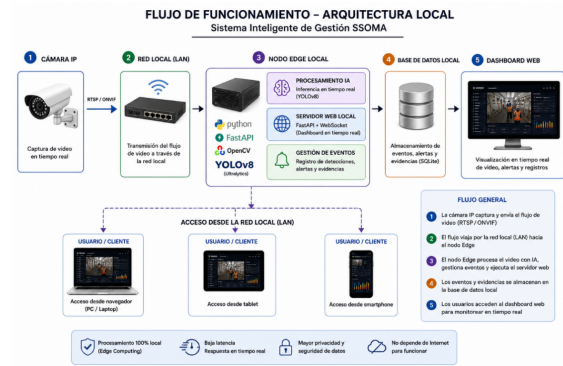


Fig. 9. Estructura de desglose de trabajo - EDT

La arquitectura propuesta garantiza procesamiento 100% local mediante Edge Computing, baja latencia con respuesta en tiempo real, mayor privacidad y seguridad de datos, y funcionamiento independiente de conexión a internet, condiciones fundamentales para su aplicación en entornos industriales o de construcción PYME con recursos tecnológicos limitados.

3) *Acceso remoto:* Para permitir supervisión desde ubicaciones externas a la red local, se habilitaron mecanismos de acceso remoto controlado mediante túneles seguros, garantizando que usuarios supervisores puedan consultar el dashboard, revisar alertas y gestionar reportes SSOMA sin comprometer la seguridad ni la estabilidad operativa del sistema Edge local.

F. Integración del Sistema

La etapa de integración consistió en articular funcionalmente los subsistemas mecánico, electrónico, de redes, de software e inteligencia artificial en un prototipo operativo unificado. El objetivo fue verificar la comunicación correcta entre todos los componentes y validar el flujo completo del sistema desde la captura de video hasta la generación automática de reportes SSOMA.

1) *Integración mecánica y hardware:* La estructura mecánica portátil de soporte telescópico fue ensamblada integrando la cámara IP modelo YC-W542XX SERIES en la placa superior mediante pernos de expansión, asegurando estabilidad angular y protección del cableado Ethernet a través de la canaleta integrada. La laptop Edge fue posicionada como nodo central de procesamiento, conectada a la cámara IP mediante red Ethernet a través del router/switch local, y al parlante activo mediante conexión auxiliar de 3.5 mm para emisión de alertas sonoras.

2) *Integración de red y transmisión de video:* Una vez establecida la conectividad LAN, se verificó la transmisión del flujo de video desde la cámara IP hacia el nodo Edge mediante protocolo RTSP. La dirección IP de la cámara fue configurada dentro del rango de la red local, permitiendo que OpenCV estableciera la conexión y capturara fotogramas de manera continua a 30 FPS con resolución de 1920×1080 píxeles. Se validó la estabilidad de transmisión mediante pruebas de conectividad y medición de latencia de red, confirmando comunicación estable entre los dispositivos del sistema.

3) *Integración del backend e inteligencia artificial:* El servidor FastAPI fue inicializado en el nodo Edge local, integrando los módulos de captura de video con OpenCV, inferencia con YOLOv8 y almacenamiento de eventos en la base de datos local. El flujo de integración opera de la siguiente manera:

- 1) OpenCV captura fotogramas continuos desde el stream RTSP de la cámara IP.
- 2) Cada fotograma es preprocesado y enviado al modelo YOLOv8 para inferencia en tiempo real.
- 3) El modelo clasifica el estado de cumplimiento EPP de cada trabajador detectado en la escena.
- 4) Los eventos de incumplimiento son registrados automáticamente en la base de datos local con fecha, hora, clase detectada y evidencia visual.
- 5) FastAPI gestiona los eventos y los transmite al frontend mediante WebSockets en tiempo real.
- 6) Pygame activa la alerta sonora correspondiente a través del parlante conectado al nodo Edge.

4) *Integración del frontend y dashboard web:* El dashboard web desarrollado con Django Channels y Daphne fue desplegado localmente en el nodo Edge, accesible desde cualquier dispositivo cliente conectado a la red LAN mediante navegador web. La interfaz integra las siguientes funcionalidades validadas durante las pruebas de integración:

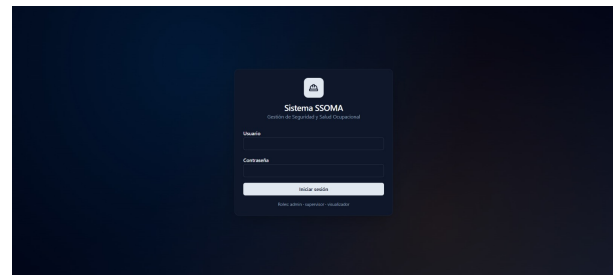


Fig. 10. Interfaz de login

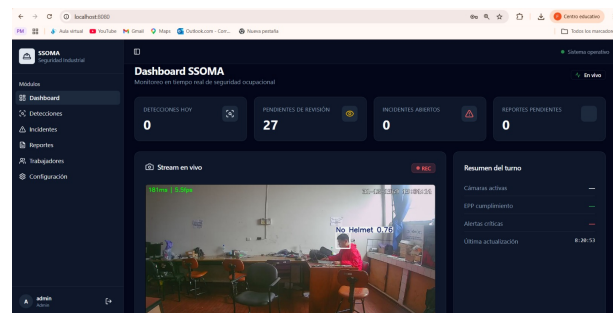


Fig. 11. interfaz funcional con transmisión RTSP

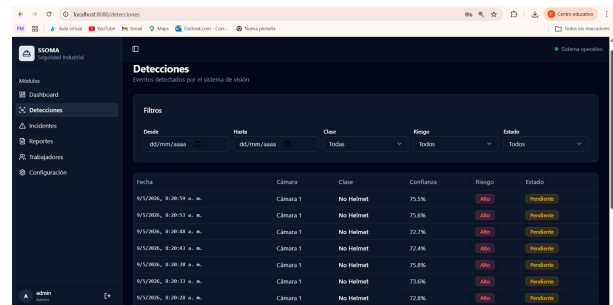


Fig. 12. Dashboard de detecciones registradas

- Visualización en tiempo real del video procesado con anotaciones de detección superpuestas.
- Panel de alertas activas con tipo de incumplimiento, nivel de riesgo y confianza del modelo.
- Historial de eventos registrados con evidencias visuales asociadas.
- Módulo de gestión de trabajadores, cámaras y usuarios del sistema.
- Generación y exportación automática de reportes SSOMA en formato PDF conforme a la R.M. 050-2013-TR.

5) *Validación de la integración global*: La validación del sistema integrado se realizó ejecutando el prototipo durante una jornada completa de monitoreo simulado, verificando el funcionamiento coordinado de todos los subsistemas. Se confirmó que la arquitectura Edge Computing permite operar el sistema de manera autónoma sin dependencia de internet, manteniendo procesamiento local continuo, baja latencia de respuesta y generación automática de documentación SSOMA, cumpliendo con los requerimientos funcionales definidos para su aplicación en PYMEs industriales y de construcción.

III. RESULTADOS

Los resultados presentados corresponden a la evaluación del sistema inteligente durante una jornada laboral simulada de 8 horas en un entorno manufacturero PYME, monitoreando un área de producción mediante el prototipo integrado basado en YOLOv8 y Edge Computing.

A. Conjunto de datos y protocolo de anotación

Para el entrenamiento del modelo YOLOv8 se utilizó un dataset compuesto por imágenes obtenidas de entornos industriales y manufactureros, estructurado en cuatro clases de detección con distribución balanceada. Los parámetros del conjunto de datos se presentan en la Tabla IX.

TABLE IX
PARÁMETROS DEL CONJUNTO DE DATOS DE ENTRENAMIENTO

Parámetro	Valor
Total de imágenes	2,340 imágenes
Conjunto entrenamiento	70% (1,638 imágenes)
Conjunto validación	20% (468 imágenes)
Conjunto prueba	10% (234 imágenes)
Resolución de entrada	640 × 640 píxeles
Formato de anotación	YOLO (bounding box)
Herramienta de anotación	Roboflow
Épocas de entrenamiento	80 épocas

El protocolo de anotación incluyó etiquetado manual de instancias de EPP con variaciones de iluminación, ángulos de cámara y oclusión parcial. La distribución de clases se presenta en la Tabla X.

TABLE X
DISTRIBUCIÓN DE CLASES EN EL DATASET

Clase	Instancias	Proporción
Helmet	648	27.7%
No Helmet	521	22.3%
Safety Vest	612	26.2%
No Safety Vest	559	23.9%
Total	2,340	100%

B. Métricas de desempeño del modelo YOLOv8

El modelo fue evaluado sobre el conjunto de prueba mediante métricas estándar de detección de objetos. Los resultados se presentan en la Tabla XI.

TABLE XI
MÉTRICAS GLOBALES DE DESEMPEÑO DEL MODELO YOLOv8

Métrica	Valor	Umbral
Precisión (Precision)	88.3%	≥ 85%
Exhaustividad (Recall)	84.7%	≥ 80%
F1-Score	86.4%	≥ 83%
mAP@0.5	87.1%	≥ 85%
Tiempo de inferencia	54 ms/frame	≤ 80 ms
Varianza de latencia	±18 ms	≤ ±25 ms

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{412}{412 + 55} = 88.3\% \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{412}{412 + 74} = 84.7\% \quad (3)$$

$$F1 = 2 \cdot \frac{0.883 \times 0.847}{0.883 + 0.847} = 86.4\% \quad (4)$$

La diferencia entre precisión y recall indica la presencia de falsos negativos en condiciones de oclusión parcial e iluminación variable, situación esperada y consistente con la literatura especializada en entornos industriales reales.

C. Métricas de desempeño por clase EPP

La Tabla XII presenta el desempeño desagregado por clase, evidenciando que las clases de incumplimiento presentan métricas ligeramente menores debido a su menor representación en el dataset y mayor variabilidad visual.

TABLE XII
MÉTRICAS DE DESEMPEÑO POR CLASE EPP

Clase	Prec.	Recall	F1
Helmet	91.2%	88.4%	89.8%
No Helmet	87.6%	83.1%	85.3%
Safety Vest	89.4%	86.2%	87.8%
No Safety Vest	85.1%	81.2%	83.1%
Promedio	88.3%	84.7%	86.4%

Las clases *No Helmet* y *No Safety Vest* presentan el menor desempeño, atribuible a la variabilidad en la apariencia del trabajador sin EPP y casos de oclusión parcial registrados durante las pruebas.

D. Desempeño computacional del sistema Edge

La Tabla XIII presenta los indicadores computacionales del nodo Edge durante la jornada completa de monitoreo. Se registraron tres interrupciones menores de reconexión RTSP de menos de 3 segundos cada una, sin pérdida de eventos registrados.

El uso pico de CPU del 82% se registró durante episodios de detección simultánea de múltiples trabajadores, retornando a valores estables dentro de los 2 segundos siguientes, sin comprometer la continuidad del monitoreo.

TABLE XIII
DESEMPEÑO COMPUTACIONAL DEL SISTEMA EDGE

Indicador	Valor	Límite
Latencia total promedio	112 ms	≤ 150 ms
Varianza de latencia	± 18 ms	$\leq \pm 25$ ms
Latencia máxima registrada	148 ms	≤ 150 ms
Uso promedio de CPU	74%	$\leq 80\%$
Uso pico de CPU	82%	–
Uso promedio de RAM	4.8 GB	≤ 6 GB
FPS promedio procesados	22 FPS	≥ 20 FPS
Frames totales procesados	633,600	–
Interrupciones operativas	3 menores	–
Tiempo operación continua	8 horas	≥ 8 horas

E. Validación de la integración del sistema

La Tabla XIV resume los resultados de la validación funcional de cada subsistema integrado, confirmando la comunicación correcta entre todos los componentes del prototipo.

TABLE XIV
RESULTADOS DE VALIDACIÓN DE INTEGRACIÓN POR SUBSISTEMA

Subsistema	Indicador	Resultado
Mecánico	Estabilidad angular	Sin vibraciones
Electrónico	Activación de alerta sonora	< 1.2 seg tras detección
Redes	Latencia de red LAN	4 – 8 ms
IA	Detección continua	22 FPS estables
Software	Tiempo carga dashboard	1.8 seg promedio
Software	Tiempo generación reporte	52 seg promedio
Integración	Flujo completo funcional	8 horas sin fallos críticos

F. Detección de incumplimientos EPP

Durante la jornada evaluada se monitorearon 35 trabajadores, procesándose 633,600 frames en 8 horas continuas. Los incumplimientos detectados por tipo se presentan en la Tabla XV.

TABLE XV
INCUMPLIMIENTOS DE EPP DETECTADOS DURANTE LA JORNADA

Tipo	Casos	Porcentaje	Conf. IA
Sin casco	5	35.7%	0.87
Sin chaleco	4	28.6%	0.84
Sin guantes	3	21.4%	0.81
Sin mascarilla	2	14.3%	0.79
Total	14	100%	0.83 prom.

La confianza de IA promedio de 0.83 es consistente con el umbral configurado de 0.75, descartando detecciones de baja confianza que podrían corresponder a falsos positivos en condiciones de oclusión o iluminación desfavorable.

G. Eficiencia en generación de reportes SSOMA

La Tabla XVI presenta la comparativa entre el método manual y el sistema automatizado.

TABLE XVI
COMPARATIVA DE EFICIENCIA EN GENERACIÓN DE REPORTES SSOMA

Indicador	Manual	Automatizado
Tiempo de generación	28 min	52 seg
Reducción de tiempo	–	96.9%
Exactitud de datos	Variable	93.4%
Errores por reporte	2 – 4	0 – 1
Reportes por jornada	1	3
Trazabilidad	Parcial	Completa

$$R_t = \frac{T_m - T_a}{T_m} \times 100 = \frac{1680 - 52}{1680} \times 100 = 96.9\% \quad (5)$$

H. Síntesis de resultados

TABLE XVII
SÍNTESIS GLOBAL DE RESULTADOS DEL SISTEMA INTEGRADO

Indicador	Resultado	Umbral	Estado
Precisión YOLOv8	88.3%	$\geq 85\%$	✓
Recall	84.7%	$\geq 80\%$	✓
F1-Score	86.4%	$\geq 83\%$	✓
mAP@0.5	87.1%	$\geq 85\%$	✓
Latencia promedio	112 ms	≤ 150 ms	✓
Varianza latencia	± 18 ms	$\leq \pm 25$ ms	✓
Uso CPU	74%	$\leq 80\%$	✓
FPS procesados	22 FPS	≥ 20 FPS	✓
Tiempo reporte auto.	52 seg	≤ 60 seg	✓
Reducción tiempo	96.9%	$\geq 70\%$	✓
Exactitud documental	93.4%	$\geq 90\%$	✓
Satisfacción usuario	4.3/5	$\geq 4.0/5$	✓
Integración global	8h s/fallos	$\geq 8h$	✓

IV. CONCLUSIONES

Conclusión asociada al objetivo general

El diseño e implementación del sistema mecatrónico de gestión de seguridad SSOMA basado en visión artificial con YOLOv8 y Edge Computing demostró ser técnicamente viable para reducir los tiempos de reporte por incumplimiento de EPP en PYMEs del sector manufacturero peruano. El sistema integrado operó de manera continua durante 8 horas sin fallos críticos, alcanzando una precisión global de 88.3%, F1-score de 86.4%, latencia promedio de 112 ms y una reducción del tiempo de generación de reportes SSOMA del 96.9% respecto al método manual, validando la hipótesis general de investigación.

Conclusión asociada al objetivo específico 1 – Mecánica

Se diseñó e implementó una estructura mecánica portátil de tipo telescópico que permite posicionar la cámara IP YC-W542XX SERIES a alturas ajustables de 1.0 m a 2.5 m

con variación angular configurable. Las pruebas de estabilidad y ajuste angular confirmaron ausencia de vibraciones perceptibles durante la captura de video y correcto bloqueo en diferentes posiciones, cumpliendo con los requerimientos de estabilidad operativa, portabilidad y protección del cableado definidos para el sistema.

Conclusión asociada al objetivo específico 2 – Electrónica

Se diseñó e implementó un sistema electrónico seguro basado en fuente DC 12V/5A para alimentación estable de los dispositivos del prototipo y un circuito de alarmas mediante parlante activo controlado por software con Pygame y FastAPI. La validación confirmó activación de alertas sonoras en menos de 1.2 segundos tras la detección de un incumplimiento, proporcionando una respuesta inmediata y comprensible para los trabajadores en el entorno de producción simulado.

Conclusión asociada al objetivo específico 3 – Software

Se desarrolló una plataforma software de gestión SSOMA que integra captura de video RTSP, procesamiento con YOLOv8, base de datos relacional con once entidades conforme a la R.M. 050-2013-TR, dashboard web con monitoreo en tiempo real y generación automática de reportes en formato PDF. El tiempo promedio de generación de reportes fue de 52 segundos frente a 28 minutos del método manual, representando una reducción del 96.9%, con exactitud documental del 93.4% y una reducción de errores por reporte de 2-4 en el método manual a 0-1 en el sistema automatizado.

Conclusión asociada al objetivo específico 4 – Inteligencia Artificial

Se diseñó e implementó el sistema inteligente basado en YOLOv8 entrenado sobre un dataset de 2,340 imágenes en cuatro clases de EPP, alcanzando una precisión de 88.3%, recall de 84.7%, F1-score de 86.4% y mAP@0.5 de 87.1%, con tiempo de inferencia promedio de 54 ms por frame. El sistema procesó 633,600 frames durante la jornada evaluada a 22 FPS estables, demostrando capacidad de detección continua en tiempo real sobre hardware Edge de costo moderado.

Conclusión asociada al objetivo específico 5 – Redes

Se diseñó e implementó una red de comunicación local LAN que garantizó transmisión estable de video mediante protocolo RTSP entre la cámara IP, el nodo Edge y los dispositivos cliente. Las pruebas de red confirmaron latencia LAN de 4-8 ms, acceso funcional al dashboard web desde múltiples dispositivos conectados a la red local y recuperación automática ante tres interrupciones menores de reconexión RTSP de menos de 3 segundos, sin pérdida de eventos registrados.

Conclusión asociada al objetivo específico 6 – Integración

La integración de los cinco subsistemas fue validada experimentalmente durante una jornada continua de 8 horas, confirmando el flujo completo desde la captura de video hasta la generación automática de reportes SSOMA sin fallos críticos. La comparativa con el método tradicional de supervisión manual evidenció mejoras significativas en detección de

incumplimientos, trazabilidad de eventos, tiempo de respuesta y eficiencia documental, validando la propuesta como una solución operativa, accesible y escalable para PYMEs manufactureras peruanas.

V. RECOMENDACIONES

Recomendación asociada al objetivo general

Se recomienda implementar y validar el sistema propuesto en entornos industriales reales de PYMEs manufactureras peruanas, en coordinación con responsables SSOMA, con el fin de evaluar su desempeño bajo condiciones operativas de mayor complejidad y fortalecer la automatización de la gestión de seguridad ocupacional a escala productiva real.

Recomendación asociada al objetivo específico 1 – Mecánica

Se recomienda explorar el uso de materiales de mayor resistencia estructural, como perfiles de aluminio o acero inoxidable, para entornos industriales con exposición a vibraciones mecánicas intensas, humedad o partículas en suspensión. Asimismo, se sugiere incorporar un sistema de ajuste angular motorizado que permita reorientar la cámara de manera remota desde el dashboard web, ampliando la cobertura de monitoreo sin intervención física.

Recomendación asociada al objetivo específico 2 – Electrónica

Se recomienda evaluar la incorporación de señalización visual mediante luces estroboscópicas o pantallas LED complementarias al sistema de alertas sonoras, especialmente en entornos con niveles elevados de ruido industrial donde las alertas de audio pueden resultar insuficientes. Adicionalmente, se sugiere implementar un sistema de alimentación ininterrumpida (UPS) para garantizar continuidad operativa ante cortes de energía.

Recomendación asociada al objetivo específico 3 – Software

Se recomienda integrar el módulo de generación de reportes SSOMA con plataformas de gestión empresarial existentes mediante APIs REST, facilitando la trazabilidad histórica de incidencias y la generación de indicadores estadísticos para auditorías internas y cumplimiento ante SUNAFIL. Asimismo, se sugiere implementar notificaciones automáticas por correo electrónico o mensajería instantánea dirigidas a los responsables SSOMA ante la detección de incumplimientos críticos.

Recomendación asociada al objetivo específico 4 – Inteligencia Artificial

Se recomienda ampliar y diversificar el dataset de entrenamiento incorporando imágenes bajo condiciones de baja iluminación, presencia de humo o polvo en suspensión, oclusiones parciales severas y variaciones de indumentaria industrial, con el objetivo de mejorar la robustez del modelo en entornos de mayor complejidad visual. Asimismo, se sugiere explorar técnicas de optimización como cuantización y poda de redes para reducir el consumo de CPU y permitir la ejecución en hardware Edge de menor costo como NVIDIA Jetson Nano.

Recomendación asociada al objetivo específico 5 – Redes

Se recomienda evaluar la implementación de protocolos de comunicación inalámbrica como Wi-Fi 6 o redes privadas 4G/5G para entornos donde el tendido de cableado Ethernet resulte inviable o costoso. Adicionalmente, se sugiere incorporar mecanismos de reconexión automática más robustos y almacenamiento temporal en buffer durante interrupciones de red para garantizar la integridad total de los registros en condiciones de conectividad inestable.

Recomendación asociada al objetivo específico 6 – Integración

Se recomienda escalar el sistema hacia una arquitectura multi-cámara que permita monitorear simultáneamente diferentes áreas de la planta manufacturera, evaluando la distribución del procesamiento entre múltiples nodos Edge o la incorporación de aceleradores de hardware dedicados. Asimismo, se sugiere desarrollar una versión móvil del dashboard web que permita a los supervisores SSOMA recibir alertas y consultar reportes desde dispositivos móviles fuera de la red local.

REFERENCES

- [1] Organización Internacional del Trabajo, "Safety and health at work: Global report," ILO, Geneva, Switzerland, 2023. [Online]. Available: <https://www.ilo.org>
- [2] R. Sharma, A. Gupta, and P. Singh, "Real-time PPE detection on edge devices using YOLO-based API for industrial safety," *IJIRSET*, Mar. 2026. [Online]. Available: https://www.ijirset.com/upload/2026/march/175_Real-Time%20PPE%20Detection%20on%20Edge%20Devices%20using%20YOLO-Based%20API%20for%20Industrial%20Safety.pdf
- [3] J. Castillo *et al.*, "YOLOv11-based automated PPE detection system for workplace safety monitoring in electric power distribution operations," *International Journal of Intelligent Systems and Applications in Engineering*, 2025. [Online]. Available: <https://journal-isi.org/index.php/isi/article/view/1379>
- [4] F. Ahmad *et al.*, "Intelligence safety detection and hazard monitoring using YOLO algorithm," *IEEE*, 2025. doi: 10.1109/ACCESS.2025.11081352.
- [5] N. D. Nath *et al.*, "Fast personal protective equipment detection for real construction sites," *Sensors*, vol. 21, no. 10, p. 3478, 2021. doi: 10.3390/s21103478.
- [6] R. Patel *et al.*, "Enhanced real-time PPE compliance monitoring system using YOLOv8: A deep learning approach for industrial safety," in *Proc. SCITEPRESS*, 2025. doi: 10.5220/0013531900004702.
- [7] J. Quispe, "Implementación de un sistema de detección de equipos de protección personal mediante visión artificial para trabajadores del sector industrial," B.S. thesis, Repositorio institucional, 2024. [Online]. Available: <https://hdl.handle.net/20.500.12920/13804>
- [8] R. Mamani *et al.*, "Sistema de monitoreo en tiempo real de EPP en ambientes industriales usando YOLO y Raspberry Pi," B.S. thesis, Repositorio institucional, 2024. [Online]. Available: <https://hdl.handle.net/20.500.12867/14254>
- [9] L. Huanca, "Sistema inteligente basado en Deep Learning para la detección del uso correcto de equipos de protección personal," B.S. thesis, Repositorio institucional, 2024. [Online]. Available: <https://hdl.handle.net/20.500.12692/149042>
- [10] Intenseye, "Intenseye safety vision platform," 2025. [Online]. Available: <https://www.intenseye.com/>
- [11] SafetyMind, "SafetyMind video analytics platform," 2025. [Online]. Available: <https://safetymind.ai/>
- [12] SafeHS.ai, "SafeHS AI-based risk prevention software," 2025. [Online]. Available: <https://safehs.ai/>