ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.603

Artificial Intelligence-Based Application for Monitoring Safety Equipment in a Construction Sector Company

Leonardo Salinas Paullo [1]; Kevin Torrejon Mundaca [1]; Fernando Sierra-Liñan [1]
[1]Universidad Tecnológica del Perú, U19219126@utp.edu.pe, U19206947@utp.edu.pe, C22334@utp.edu.pe

Abstract* - This research focuses on the problem of occupational accidents due to the improper use of personal protective equipment (PPE) in the construction industry. The implementation of a detection model based on Artificial Intelligence is necessary to increase the efficiency and accuracy of monitoring these implements, as well as to reduce the incidence rate. During the development and training process of the model, 5,886 high-definition (HD) JPG images were collected in a dataset on the RoboFlow platform. For processing, these images were scaled to 640 x 640 pixels and were related to working environments in this sector, as well as different climatic conditions, spaces, and focuses. A comparison of the models was carried out. YOLOv8 presented a mAP of 93.38%, in contrast to SSD, which reached 34.22%, and Faster R-CNN, which achieved 41.99% on average during the first 50 training epochs. Subsequently, for the final review of the training, 400 epochs of the YOLOv8 model were completed, resulting in a mAP of 93.22%, a recall of 89.67%, and an accuracy of 91.18%. With tuning and training finalized, the model was used for the development of a web system, which was subsequently hosted on a cloud server to facilitate access. This tool promotes compliance with safety regulations during the execution of daily tasks in this sector.

***Keywords:* Artificial intelligence, Monitoring, Supervision, Safety Equipment, Security.**

Aplicación Basada en Inteligencia Artificial para el Monitoreo de Equipos de Seguridad en una Empresa del Sector Construcción

Leonardo Salinas Paullo [1]; Kevin Torrejon Mundaca [1]; Fernando Sierra-Liñan [1]
[1]Universidad Tecnológica del Perú, U19219126@utp.edu.pe, U19206947@utp.edu.pe, C22334@utp.edu.pe

Resumen - Esta investigación se centra en la problemática de los accidentes laborales por el uso inadecuado de implementos de seguridad personal (EPP) en la industria de construcción. La implementación de un modelo de detección basado en Inteligencia Artificial resulta necesaria para aumentar la eficacia y la precisión de la supervisión de estos implementos, además de reducir la tasa de incidencias. Durante el proceso de desarrollo y entrenamiento del modelo se recopilaron en un conjunto de datos en la plataforma RoboFlow, en su totalidad se procesaron 5886 imágenes JPG en alta definición (HD), que para su procesamiento se escalaron a 640 x 640 píxeles con, estas relacionadas con los entornos laborales de este sector, así como diferentes condiciones climáticas, espacio y enfoque. Se llevó a cabo una comparativa de los modelos. YOLOv8 presentó una mAP del 93.38% en contraste con SSD, que llegó a alcanzar el 34.22% y Faster R-CNN, el 41.99% en promedio para las primeras 50 épocas de entrenamiento. Posteriormente, para la revisión final del entrenamiento, se alcanzaron las 400 épocas del modelo YOLOv8, brindando una resultante de mAP de 93.22%, un recall de 89.67% y una precisión de 91.18%. Con la afinación y entrenamiento, el modelo se usó para el desarrollo de un sistema web el cual fue posteriormente alojado en un servidor en la nube para su acceso, impulsando a través de esta herramienta el cumplimiento de las normativas de seguridad durante la ejecución de labores cotidianas en este sector.

Palabras clave: Inteligencia artificial, Monitoreo, Supervisión, Equipos de seguridad, seguridad.

I. INTRODUCCIÓN

Una de las industrias que ha impulsado de manera significativa el desarrollo de una sociedad moderna es la construcción [1]. La cual se ha visto optimizada por medio de los avances tecnológicos, cambios científicos y diversas ramas de la ingeniería, los cuales brindan mejores herramientas. No obstante, la intervención humana resulta ser preponderante en este rubro, puesto que es necesario para la gestión organizativa y sistemática [2]. Por ende, la seguridad de los operadores adquiere una atención medular. Ello debido a la naturaleza de las labores que realizan, ya sea con el uso de herramientas y materiales pesados, filosos, o de composición (química o física) que pueda ser dañina para la salud [3]. Esto al realizar instalaciones en áreas elevadas o de difícil acceso [4].

Por otro lado, en el contexto internacional, podríamos señalar que la Organización Internacional del Trabajo [3] sostiene que este sector presenta un índice de accidentes laborales por encima del promedio, lo que incentiva a las empresas a implementar mecanismos de aseguramiento mucho más minuciosos y específicos para los trabajadores [10]. Sin embargo, ninguno de estos métodos resulta ser infalible al estar presente la arbitrariedad y la baja consistencia en estos procesos [3]. Las causas de incidencias en proyectos u obras construcción se atribuyen a la escasa atención de las empresas constructoras en materia de salud y seguridad, alta rotación de personal, seguido de errores de los trabajadores vinculados con la falta de conocimiento o comprensión de instrucciones en el uso de implementos de seguridad personal (EPP) [5]. En consecuencia, el emplear equipamiento de seguridad reduce el riesgo de accidentes laborales, siempre que se mantenga su correcto funcionamiento durante el uso como a través de programas de mantenimiento. Por ello, es esencial identificar y priorizar adecuadamente los equipos que requieren mayor atención [6].

Estas dificultades resaltan la necesidad de explorar alternativas que permitan una supervisión más efectiva, por ello se considera el uso de la inteligencia artificial como un medio que permita garantizar la atenuación del error humano, así mismo reducir la tasa de incidencias y potenciar el cumplimiento normativo mediante la estructuración de un modelo de procesamiento de imágenes en tiempo real. Por otro lado, precisión de las características que esta herramienta posee a la hora de identificar patrones y correlaciones en grupos de datos complejos (datos como tráfico, ruido, consumo de energía, movimiento, etc.), pueden aportar capacidades computacionales más avanzadas en ecosistemas digitales [7]. De esta manera posibilita automatizar y tomar de decisiones basada en información recopilada [8]. Por lo tanto, su integración puede optimizar los procesos de este sector, al proporcionarle un constante entrenamiento que se acople al aprendizaje automático de esta tecnología con “assets” digitales específicos de imágenes que ejemplifiquen los ambientes laborales donde se harán uso de los implementos de seguridad. Sumado a ello se considera otro factor que ha posicionado a la IA dentro de diversos campos; nos referimos a la introducción del Deep Learning, que impactó positivamente la capacidad de procesamiento de computadores modernos. Este subcampo del aprendizaje automático emplea redes neuronales artificiales que captan

patrones informativos en imágenes que evitan ciertas deficiencias que sí presenta la vista humana [7][9]; este añadido aporta una nueva línea de conocimiento a la supervisión de seguridad ocupacional y demostrar la efectividad del uso de IA en ambientes industriales. Empleando como base modelos que conformen el reconocimiento de equipos y condiciones de seguridad específicos [10].

Específicamente este estudio tiene como objetivo suministrar modelos de IA desde una biblioteca de imágenes procesadas bajo etiquetas y clasificaciones según su utilidad para obtener un resultado de entrenamiento bajo algoritmos de precisión y productividad que se puedan visualizar a través de una aplicación web para detectar el uso inadecuado de EPP en las actividades cotidianas de este sector. La estructura restante de la investigación se divide de la siguiente manera. La sección II se basa en las bases teóricas que aportan conocimiento sobre los conceptos, modelos aplicados y antecedentes que aporten un preámbulo sobre el tema tratado. La sección III se centra en detallar la estructura de etapas y el método utilizado en este estudio. La sección IV muestra el proceso de entrenamiento de imágenes e incluye una explicación detallada de resultados por modelo utilizado, así como el despliegue de la aplicación. La Sección V presenta la discusión de resultados en contraste con los antecedentes e investigaciones afines. Finalmente, la sección VI presenta conclusiones y sugerencias para trabajos futuros.

II. BASES TEÓRICAS

A. Inteligencia Artificial

La inteligencia artificial como tecnología emergente en los últimos años se ha cimentado en todo ámbito profesional, industrial, científico, entre otros [11]. Esto se debe a las capacidades que le brindan a la humanidad de alcanzar, automatizando u omitiendo procedimientos complejos y repetitivos, o generando aprendizaje gracias al desarrollo de Internet durante esta última década [12]. Lo que impulsó la investigación innovadora en IA y ha sumado un alcance significativo en la potencia computacional que puede emplear, como en las redes neuronales o aprendizaje profundo, que la dirigieron hacia un periodo de crecimiento continuo [13]. En el caso de esta investigación la flexibilidad que nos brinda para adaptar sus características a servicio del sector de construcción ha sido crucial para evidenciar la necesidad hacer uso de esta herramienta en este ámbito, aspectos de la IA como el aprendizaje profundo, el aprendizaje automático y el procesamiento del lenguaje natural han facilitado su uso en diversos aspectos de la construcción [10], como la supervisión de la integridad del trabajador, la predicción de riesgos, la estimación de costes, la optimización de procesos, etc.

Como ejemplos claros de ello, los autores de [13] referenciaron estudios especializados en el uso de estas redes neuronales para predecir la resistencia de material noble como el hormigón, y sobre la implementación de sistemas inteligentes neuronales y difusos para predecir el rendimiento en la ingeniería estructural de edificaciones, lo que abarca una visión revisada de aspectos relacionados con la preservación de la efectividad y calidad de una obra de construcción. Sin embargo, en este trabajo nos enfocamos en utilizar la inteligencia artificial para desarrollar un sistema que permita salvaguardar la seguridad de los trabajadores mediante el monitoreo de los equipos de seguridad.

B. Monitoreo del equipo de seguridad

Los equipos de seguridad personal son implementos que están comúnmente presentes en la industria de construcción, dadas las amplias y complicadas circunstancias en las obras, diversas tareas de estos proyectos dependen del apoyo humano para llevarlas a cabo, lo que hace inevitable el trabajo manual pesado. Los autores [5] resaltan lo mencionado en OSHA 3151-12R, respecto a que los EPP se usan para minimizar la exposición a una variedad de peligros. Por tanto, también ha sido factor central en sus regulaciones como uno de estos intentos por reducir los riesgos, potenciando el control de su uso y adecuación al operario de construcción. En la referencia [14] se toma en cuenta que ante factores de peligro presentes en esta industria que causan los MAH (Peligros de accidente grave), los equipos de protección personal son considerados primordialmente para reducir el riesgo de que se susciten, por tanto, es importante que estos sean confiables, ya que cada fallo que impida cumplir su funcionalidad implica una probabilidad mayor de catástrofe laboral.

C. YoloV8

Este modelo se centra en enfocar la localización de objetos y categorizar cada uno de ellos durante los trabajos de detección instantáneamente. Siendo una versión que envuelve una serie de mejoras significativas comparadas a sus predecesores (Yolov7, Yolov5, etc.), YOLOv8 centraliza su utilidad en seccionar la imagen mediante la reversión de regiones, las cuales realizan detecciones propias haciendo uso de un mapeado de clasificación y un cuadro de delimitaciones, donde se puede obtener al mismo tiempo el cuadro de detección real del objeto [15]. Su arquitectura basada en IA perfeccionada para tareas de identificación en simultáneo de diferentes activos, esta consta de 3 secciones, Backbone(columna), Neck (cuello) y Head (cabeza) que conforman estrategias de sobre muestreo y submuestreo para alternar la cantidad de información importante por ser procesada [16], y de esta manera brindar mayor fidelidad y rapidez, que resultan especialmente útiles para la construcción de aplicaciones que requieren resultados en un lapso corto de tiempo.

D. Faster R-CNN

Este modelo se constituye bajo redes neuronales convolucionales que incluyen también el uso de Region Proposal Network (RPN) para gestionar e indicar ciertas regiones de la imagen en proceso que posean información o activos de interés, reduciendo considerablemente el tiempo de procesamiento [17]. Luego se envía esta información a la red base para su clasificación y delimitar con bounding boxes en una imagen. Los resultados finales de la detección de objetos se generan al fusionar las cajas propuestas por la RPN con las salidas de la red encargada de la regresión y clasificación. A su vez, se usan los mapeos obtenidos de las

imágenes para aportar información espacial y semántica clave que permite a la RPN predecir las regiones de interés (RoIs) [29].

E. SSD (Single Shot Multibox Detector)

El modelo SSD basa su utilidad en emplear múltiples capas convolucionales de diferentes escalas para detectar objetos de distintos tamaños dentro de una imagen. Estas trabajan de manera simultánea, adecuado para detectar objetivos pequeños, objetivos superpuestos y objetivos múltiples [18], permitiendo que el modelo sea altamente escalable y capaz de manejar diferentes escalas eliminando la necesidad de un sistema separado de propuestas de regiones, lo que aumenta su eficacia y facilidad para integrar a aplicaciones que sean desarrolladas y desplegadas en equipos computacionales no tan potentes [19].

F. Métricas:

- *Precisión:*

La métrica mide la precisión del modelo al relacionar predicciones positivas correctas con el total generado, reflejando su confiabilidad contundente [20]:

$$precisión = \frac{TP}{TP + FP}$$

TP = Verdaderos positivos.
FP = Falsos positivos.

- *Recall:*

Evalúa la precisión del modelo al comparar las predicciones positivas correctas con el total de casos positivos reales [21][23], y usa el TP y FP, al igual que la precisión, para su análisis:

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

TP = Verdaderos positivos.
FP = Falsos positivos.
FN = Falsos Negativos.

- *Mean average precision (mAP):*

Mide la calidad del modelo de acuerdo con su precisión promedio en varios niveles de recall [22], brindando una métrica única para determinar su desempeño en la clasificación de resultados [23]:

$$mAP = \frac{\Sigma[P(Ri) \cdot \Delta Ri]}{n}$$

n = cantidad de instancias

P(Ri) = precisión en nivel de recall

ΔRi = cambio de recall Ri-1 a Ri

G. Fine-tuning completo (Full fine-tuning)

Este es el método de ajuste fino seguido en el entrenamiento del modelo elegido para la integración con la aplicación, se forma desde una base de preentrenado del modelo para posteriormente entrenarse todo el modelo dirigido a cumplir un objetivo específico. Ofreciendo mayor escalabilidad ante cambios o adiciones encontradas en el camino, pero tomando más tiempo y recursos de los equipos empleados para el entrenamiento.

H. Antecedentes relevantes

En el ámbito internacional tenemos el caso desarrollado en la referencia [24] acerca de un vehículo autónomo que evita obstáculos mediante la combinación de inteligencia artificial y sistemas de imagen estereoscópica, profundizan en el procesado de imágenes mediante segmentación semántica basada en inteligencia artificial nos permite extraer información sobre el terreno, obstáculos y fondo. Esto propone un uso innovador que integra detección de profundidad pasiva y segmentación semántica, potenciado con el uso de un módulo de cámara estéreo y el kit NVIDIA Jetson Nano, permitiendo al vehículo hacer inferencias precisas sobre su entorno y evitar obstáculos en tiempo real en contraste a lo que un sistema manual puede llegar a detectar con supervisión humana.

En materia de la ingeniería civil, la referencia [25] nos Describe la creación de una red neuronal convolucional (CNN) diseñada para la detección en tiempo real de equipos de protección personal (EPP) en entornos de construcción, basada en el algoritmo SSD-Mobile Net. Este método usó un conjunto de 3200 de capturas manuales de sistemas CCTV en el ambiente laboral e Imágenes públicas recopiladas a través de tecnología de web scraping. Como resultado del 70% de imágenes que se asignaron al grupo de entrenamiento, 339 grupo de validación y 153 al de prueba donde se alcanzó un 36% de efectividad en promedio evidenciando la sinergia de las funciones de clasificación, localización y regularización, lo que indicó que el modelo fue entrenado correctamente para monitorear estos implementos.

Otra aplicación interesante de la inteligencia artificial se revisó por los autores [26] respecto la clasificación de vehículos mediante un sistema de vigilancia y monitoreo de características como marcas, modelos, tipo, entre otros. Con el fin de potenciar su implementación en las denominadas ciudades inteligentes, para el beneficio de las investigaciones policiales y la automatización del cobro de peajes según el tipo de vehículo. Este trabajo se basó en el uso de sensores implementados en entornos CCTV para el análisis de imágenes entrenadas por modelos 3D de vehículos que sirvan como entrada de conocimiento para analizar las imágenes en 2D, en el marco de un algoritmo de CNN y Deep learning para establecer una red automatizada de detección.

Respecto al ámbito nacional, la utilidad del procesamiento de imágenes a través de modelos 3D realizado por [27] en el monitoreo de uno de los nevados más importantes del país como es el Huascarán, en conjunto con datos meteorológicos se revisaron datos obtenidos durante el año 2023 que genero un total de 58 lecturas que bajo la plataforma de Ubidots y la implementación de la tecnología AR-IoT implementada para obtener conocimiento de las condiciones atmosféricas y la toma de

decisiones respecto a los recursos que puedan ser aprovechados en esta área nevada.

En el sector construcción, [28] detalla el proceso de entrenamiento de modelos de Deep Learning para detectar en vivo el uso de equipos de seguridad, basándose en los parámetros que la OSHA (Administración de Seguridad y Salud Ocupacional de los EE. UU.) para exigir a las áreas de contratación y recursos humanos de las constructoras mejores mecanismos de supervisión para reducir el margen de incidencias. En este estudio se utilizaron modelos basados en YOLO, además de una sola red neuronal de procesamiento en CNN con tecnologías de clasificación (VGG-16, ResNet-50 y Xception) con las cuales se analizaron 1500 imágenes pertenecientes a un dataset que recopilo data digital de escenarios recurrentes en estos ambientes laborales.

Los autores [30] especializaron su estudio en la detección de objetos, primordialmente cascos de seguridad, para perfeccionar el análisis de un conjunto de 3000 assets digitales en GDUT-HWD, incluyendo características variadas como iluminación, escenario, enfoque, uso, etc. Asimismo, se empleó el uso de Reverse Progressive Attention, RPA, tecnología que se especializa en mejorar la detección de objetos multi escala, esta revisión se propaga desde las capas superiores hacia las inferiores, mejorando la detección de objetos pequeños. Lo que resultó en un método de una sola etapa que tuvo como base CNN y SSD para alcanzar un mAP del 83.89% con un tamaño de imagen de 512 × 512 en el conjunto de datos propuesto.

III. MÉTODO

La investigación se llevó a cabo siguiendo las etapas presentadas en la Fig. 1.

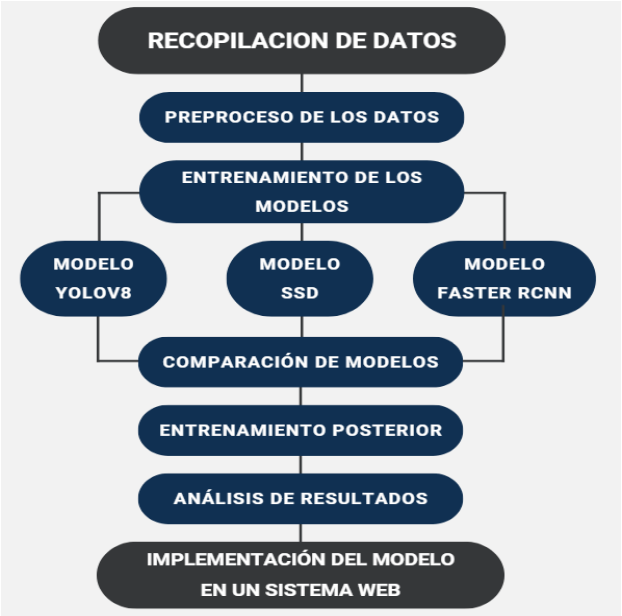


Fig. 1 Etapas de investigación

La colección de datos se obtuvo de diversas fuentes en RoboFlow, donde se aplicaron etapas de preprocesamiento, incluyendo el redimensionamiento de las imágenes a 640x640, ajuste automático de orientación, contraste, brillo, y desenfoque, entre otras técnicas, para simular diversas condiciones de entrenamiento. Posteriormente, el entrenamiento de los tres modelos se realiza en un equipo que incluye una GPU NVIDIA RTX 4070 Super, un procesador AMD Ryzen 7 5700X y 16 GB de memoria RAM DDR4 a 3200 MHz, garantizando así la capacidad computacional necesaria para manejar tareas intensivas de procesamiento y aprendizaje profundo, durante 50 épocas, uno utilizando la arquitectura YOLOv8, otro con Faster R-CNN y el otro con SSD (Single Shot Multibox Detector), con el objetivo de comparar sus resultados y seleccionar el modelo óptimo. Luego, se llevó a cabo un entrenamiento adicional de 400 épocas para mejorar aún más el desempeño del modelo seleccionado. Finalmente, se analizó el rendimiento del modelo con este ajuste y se implementó en un sistema web desarrollado específicamente para este propósito.

IV. RESULTADOS

A. *Recopilación de datos*

Para el aprendizaje del modelo de monitoreo se recopilaron 5886 imágenes referentes a actividades de la industria de construcción. Donde se tomaron en cuenta factores como la cantidad de trabajadores, tipos de proyectos (urbano o industrial), iluminación variable de acuerdo con el horario del día y los implementos de seguridad personal (EPP) que se utilizan en cada toma. Esto con la finalidad de ofrecer diversas variaciones que puedan ser analizadas por las técnicas de IA y brindar resultados más precisos ante cualquier escenario planteado durante el monitoreo en las obras. Dentro de la plataforma RoboFlow se cargaron las imágenes para su respectiva etiquetación (Persona, Casco, Chaleco, Guante, No-Casco, No-Chaleco, No-Guante), con el fin de que sean clasificados automáticamente en 3 carpetas: entrenamiento (training), validación (valid) y prueba (test) en la estructura del modelo.

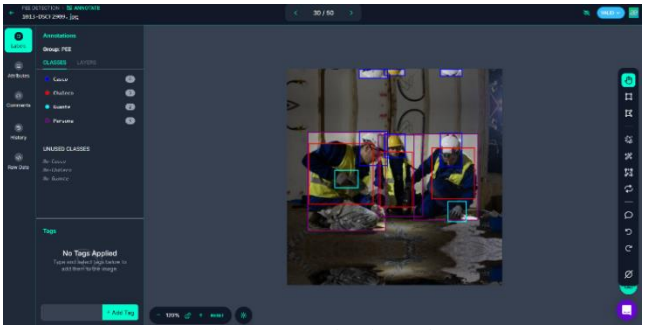


Fig. 2 Proceso de etiquetado



Fig. 3 Carpetas de dataset Robo Flow para entrenamiento IA

B. *Preprocesamiento de datos*

Las imágenes utilizadas en este estudio tenían un formato original JPG en alta definición (HD), con resoluciones que variaban en promedio entre 1280 × 720 y

2048 × 1080 píxeles. Estas imágenes se sometieron a un proceso de preprocesamiento utilizando las herramientas proporcionadas por Roboflow, aplicando las siguientes técnicas: Ajuste automático de la orientación de la imagen y escalado uniforme a 640 × 640 píxeles, utilizando el método "Stretch". Además, se realizó un aumento de datos (data augmentation) duplicando las imágenes originales con la aplicación de las siguientes transformaciones: Inversión horizontal y vertical, rotación de 90°, rotación aleatoria entre -15° y +15° para añadir variabilidad angular, ajuste de brillo entre -20% a +20%, a juste de exposición entre -10% y +10% para simular variaciones en las condiciones de iluminación, desenfoque con un radio de hasta 2.5 píxeles para simular distorsión leve, e incorporación de hasta un 1.01% de ruido en los píxeles de la imagen para simular interferencias visuales.

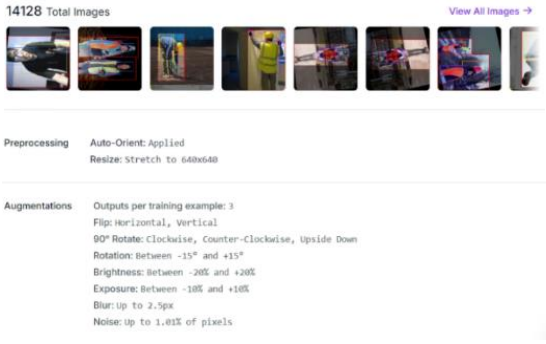


Fig. 4 Dataset con imágenes preprocesados

De esta forma, se logró obtener un conjunto de datos compuesto por 14,128 imágenes después de aplicar el preprocesamiento y las técnicas de aumento de datos. A continuación, se representa en la Tabla I que detalla la cantidad de clases que existen en las carpetas correspondientes.

TABLA I
CONJUNTO DE CLASES

Clase	Cantidad de Clases			
	Train	Valid	Test	TOTAL
Casco	9220	914	455	**10589**
Chaleco	10709	1028	518	**12255**
Guante	7380	736	361	**8477**
No-Casco	2796	237	113	**3146**
No-Chaleco	2019	168	82	**2269**
No-Guante	3140	282	135	**3557**
Persona	11701	1116	549	**13366**
TOTAL	**46965**	**4481**	**2213**	**53659**

C. Entrenamiento de los modelos

1) *YoloV8:* Es una de las últimas versiones del modelo de detección de objetos YOLO, que conserva la arquitectura básica de sus predecesores, pero introduce mejoras clave, como una nueva red neuronal que combina Feature Pyramid Network (FPN) y Path Aggregation Network (PAN) con el objetivo de mejorar la detección de objetos en diversas escalas y resoluciones, además de una herramienta de etiquetado que simplifica el proceso de anotación mediante funciones como el etiquetado automático y atajos personalizables [31]. La FPN mejora la detección al reducir la resolución de la imagen mientras incrementa los canales de características, y la PAN agrega características de distintos niveles mediante conexiones residuales, posibilitando una detección más precisa de objetos de diferentes tamaños y formas [32].

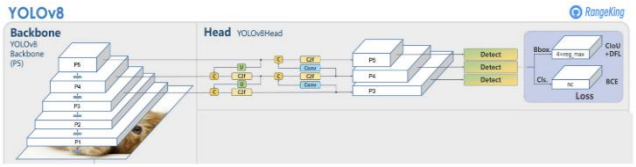


Fig. 5 Arquitectura YoloV8

El modelo YOLOv8 fue entrenado durante 50 épocas, obteniendo como resultado una curva de aprendizaje en ascenso, tal como se ilustra en la Fig. 6. Se evidenció un progreso significativo en el desempeño del modelo en las épocas iniciales; además, a partir de la época 30, se observó que llegó a superar un mAP del 0.90, lo cual indica un óptimo aprendizaje del modelo.

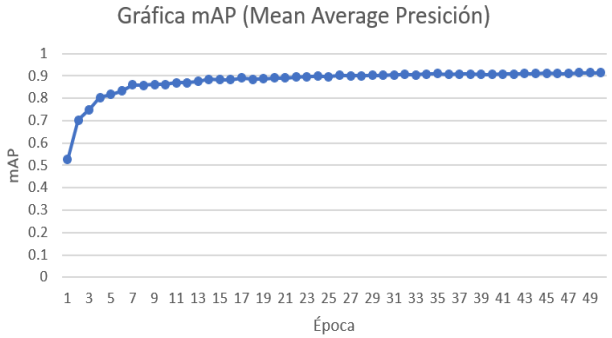


Fig. 6 Curva de aprendizaje de YOLOv8

2) Faster R-CNN: Utiliza una red de propuestas de regiones (RPN) como alternativa al algoritmo de búsqueda selectiva, lo que mejora la velocidad de detección [33]. El proceso central inicia con la recabación de características mediante una red neuronal convolucional (CNN), como VGG16 o ZF, que genera mapas de características de las imágenes [34]. A partir de estos mapas, la RPN selecciona regiones candidatas que podrían contener objetos de diferentes tamaños [33].

Estas regiones se transforman en vectores de características mediante ROI Pooling, lo que permite clasificar los objetos presentes en ellas y ajustar las dimensiones de las regiones seleccionadas para mejorar la precisión [34].

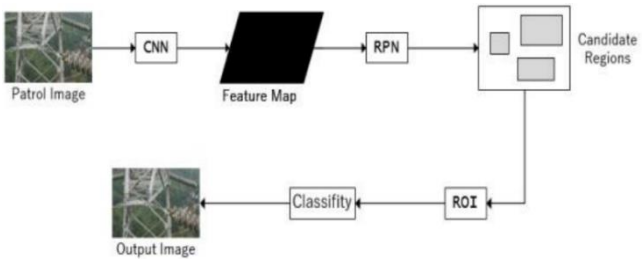


Fig. 7 Arquitectura Faster RCNN

El modelo Faster RCNN fue entrenado durante 50 épocas, obteniendo como resultado una curva de aprendizaje en ascenso, tal como se ilustra en la Fig. 8. Se evidenció un progreso significativo en el desempeño del modelo; no obstante, a partir de la época 10, se observó una estabilización en su rendimiento, alcanzando un límite cercano a 0.45 sin experimentar mejoras adicionales.

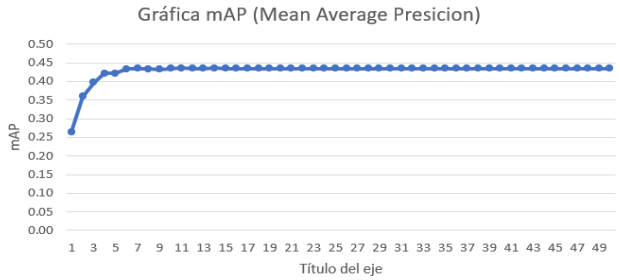


Fig. 8 Curva de aprendizaje de Faster RCNN

3) SSD (Single Shot Multibox Detector): La arquitectura SSD se basa en la combinación de la idea de cajas de anclaje, como en las RPN, y mapas de características multiescala, como en la FPN, con el objetivo de lograr una detección rápida y de alta calidad. Esta arquitectura incluye una pirámide de características superpuesta a una red convolucional base (por ejemplo, VGG16), seguida por un proceso de supresión no máxima (NMS) para generar los resultados finales [35]. En la pirámide de características, cada capa cumple una función específica al detectar objetos en diferentes escalas, generando directamente los resultados de detección desde los mapas de características en varios niveles. Aunque los mapas de características de niveles bajos contienen información esencial de ubicación, carecen de suficiente contexto semántico, pudiendo causar pérdida de información de objetos pequeños al pasar por la red troncal. En SSD, solo una capa de bajo nivel, conv4_3 en VGG16, se asigna para la detección de objetos pequeños, lo que puede limitar la precisión en este caso [36].

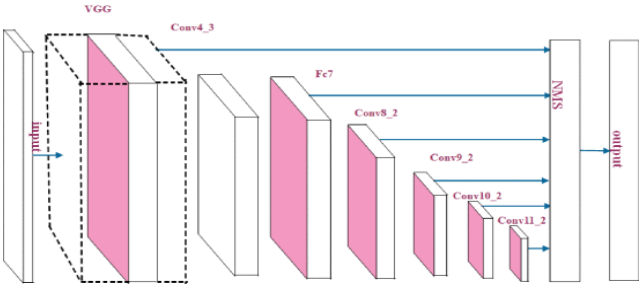


Fig. 9 Arquitectura SSD

El modelo SSD fue entrenado durante 50 épocas, obteniendo como resultado una curva de aprendizaje en ascenso, tal como se ilustra en la Fig. 10. Se evidenció un progreso significativo en el desempeño del modelo; no obstante, a partir de la época 11, se observó una estabilización en su rendimiento, alcanzando un límite cercano a 0.35 sin experimentar mejoras adicionales.

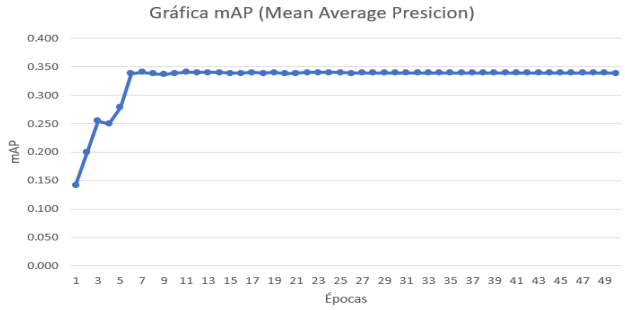


Fig. 10 Curva de aprendizaje de SSD

4) Comparación de modelos: Se realizó una comparación entre tres modelos de detección de objetos: Faster R-CNN, SSD (Single Shot Multibox Detector) y YOLOv8. El análisis se llevó a cabo utilizando los resultados obtenidos en la época 50 como punto de referencia. Los parámetros evaluados incluyeron la precisión, el recall y el mAP (mean Average Precision). La precisión evalúa la exactitud de las predicciones positivas del modelo en una imagen. En la Fig. 11 presenta las curvas de precisión de los tres modelos.

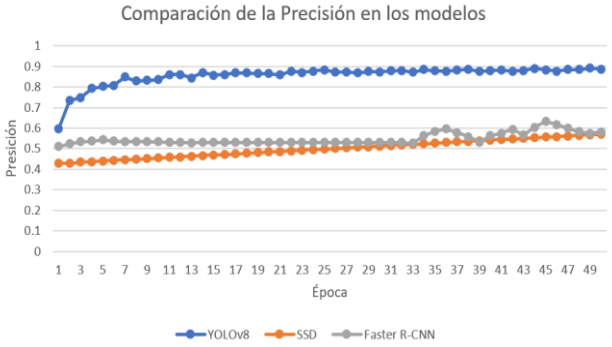


Fig. 11 Comparación entre las curvas de precisión de los modelos

Se observa que el modelo YOLOv8 alcanza una precisión del 88.63%, lo que evidencia un rendimiento superior respecto a los otros modelos analizados.

El recall, o sensibilidad, evalúa la capacidad de un modelo para detectar correctamente todas las instancias positivas en un conjunto de datos. En la Fig. 12 se muestran las curvas de recall de los tres modelos.

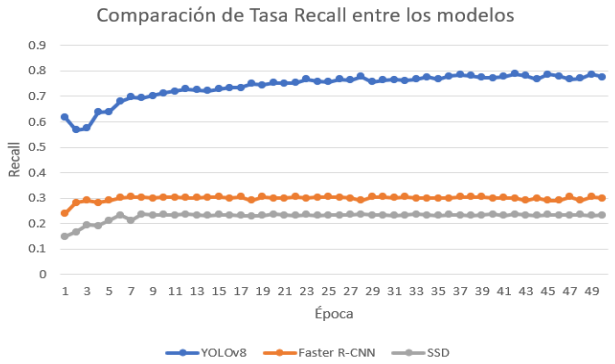


Fig. 12 Comparación entre las curvas de aprendizaje de los modelos

El gráfico demuestra que el modelo YOLOv8 alcanza un recall del 77.56%, lo que demuestra un desempeño superior en comparación con los demás modelos evaluados.

El mAP (Precisión Promedio Media) es una métrica empleada para medir el rendimiento de los modelos de detección de objetos, calculando la precisión promedio de cada clase y luego promediándolas. Un mayor mAP indica un mejor desempeño global. Las curvas de los principales promedios de precisión se visualizan en la Fig. 13.

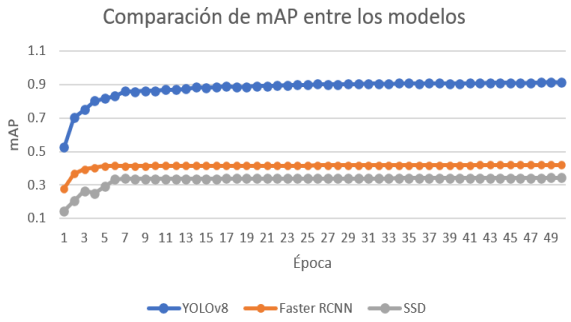


Fig. 13 Comparación entre las curvas de los principales promedios de precisión de los modelos

En el gráfico se observa que el modelo YOLOv8 alcanza una mAP del 91.38%, lo que demuestra un desempeño superior en comparación con los demás modelos evaluados.

5) Entrenamiento posterior: De acuerdo con la comparación de modelos, se determina que YOLOv8 es el más adecuado para abordar este problema, superando el rendimiento de los otros modelos evaluados. Por ello, se optimizó el modelo YOLOv8 mediante un entrenamiento de 400 épocas. Como se ilustra en la Fig. 14, a partir de la época 200 se evidencia una ligera, pero consistente mejora en los indicadores de rendimiento del modelo, incluyendo el mAP, el recall y la precisión, lo que refuerza su idoneidad para la tarea planteada.

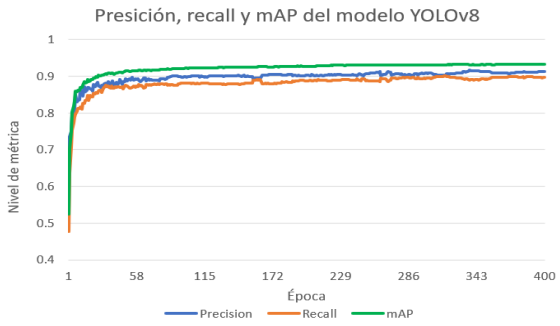


Fig. 14 Precisión, recall y mAP del modelo YOLOv8 en 400 épocas

6) Análisis de resultados: La evaluación del modelo, basada en la matriz de confusión normalizada presentada, refleja un desempeño sólido en la mayoría de las clases, evidenciando un modelo bien entrenado para la tarea de detección de PPE. Es importante destacar que las clases "Casco" y "Persona" muestran niveles sobresalientes de precisión, con valores de 0.94 y 0.96, respectivamente, en la diagonal principal. Esto pone de manifiesto la habilidad del modelo para identificar estas categorías clave con una baja tasa de errores. Por otro lado, en las clases "No-Casco" y "Chaleco", el modelo mantiene un desempeño consistente, alcanzando valores normalizados de 0.93 y 0.94, respectivamente. Esto evidencia su efectividad en diferenciar entre categorías positivas y negativas asociadas a la presencia de elementos de seguridad. Sin embargo, en clases como "No-Guante" y "Guante", se observa un desempeño ligeramente menor, con valores de 0.84 y 0.89. Esto indica que estas categorías aún presentan oportunidades para mejorar su clasificación, posiblemente debido a características menos distintivas o a una menor cantidad de datos de entrenamiento en estas clases.

En general, los resultados obtenidos demuestran que el modelo está bien optimizado para las categorías principales y es capaz de realizar predicciones confiables en la mayoría de las clases evaluadas. Los márgenes de error observados son relativamente bajos, lo que refuerza la aplicabilidad práctica del modelo en entornos reales y su capacidad para abordar tareas complejas de detección con un alto nivel de confianza. El tiempo total para el entrenamiento y afinamiento de Yolo fueron 12 horas donde el margen de mejora llego a alcanzar un 3% incluyendo todas las clases.

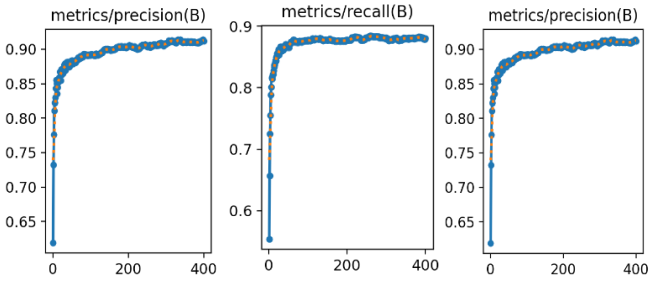


Fig. 15 Matriz de confusión del modelo YOLOv8

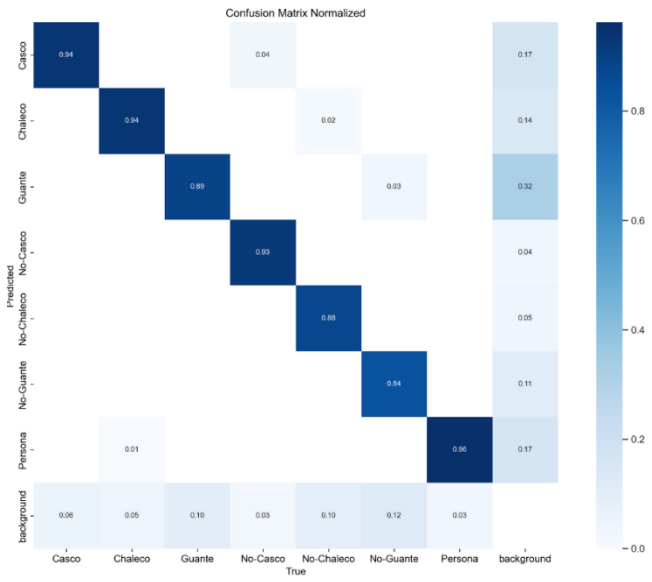


Fig. 16 Matriz de confusión del modelo YOLOv8

Las métricas que se consiguieron al finalizar el último entrenamiento del modelo YOLOv8, están presentados en la Tabla II.

TABLA II
RESULTADOS DE LAS MÉTRICAS EVALUADAS DE YOLOV8

Métrica Evaluada	Resultado
mAP	93.22%
Precisión	91.18%
Recall	89.67%

El modelo YOLOv8 desarrollado en esta investigación alcanza un mAP del 93.22%, lo que indica una precisión sobresaliente en la detección de objetos. Este resultado refleja que el modelo tiene una capacidad excepcional para identificar correctamente los objetos en las imágenes, minimizando las falsas alarmas y garantizando predicciones más confiables en una amplia variedad de escenarios.

Asimismo, la precisión lograda es del 91.18%, lo que demuestra que el modelo tiene un rendimiento sobresaliente al clasificar correctamente los objetos detectados, con una mínima cantidad de falsos positivos. Un nivel de precisión tan alto es fundamental en aplicaciones donde la fiabilidad y la exactitud en la clasificación de objetos son esenciales, como en ambientes industriales o de seguridad

Además, el recall alcanzado es del 89.67%, lo que refleja la capacidad del modelo para identificar el 89.67% de los objetos presentes en las imágenes. Esto demuestra una mejora significativa en la capacidad del modelo para identificar la mayoría de los objetos, reduciendo considerablemente la probabilidad de omitir objetos relevantes y destacando su eficacia en tareas exigentes de detección.

7) Implementación del modelo: Se desarrolló un sistema inteligente basado en el modelo YOLOv8 para la detección de equipos de protección personal (EPP) utilizados por los operadores. Este sistema cuenta con una interfaz web que permite tres modalidades de detección: a través de cámara web, carga de imágenes y carga de videos. En la sección izquierda de la interfaz se muestra el conteo de equipos de seguridad detectados, organizados por categoría, mientras que en la sección derecha se visualizan las detecciones superpuestas sobre la imagen, video o la transmisión de la cámara web. El desarrollo del sistema se realizó empleando Python para el backend, utilizando Flask como framework principal para facilitar la comunicación entre el frontend y el backend. Esto permitió la transferencia eficiente de los resultados del modelo hacia la interfaz web. En el frontend, se emplearon tecnologías como JavaScript, HTML y CSS para asegurar una experiencia de usuario intuitiva y funcional. Finalmente, el sistema fue implementado en un servidor en la nube, permitiendo el acceso remoto desde cualquier ubicación con conexión a internet. La Fig. 16 ilustra la interfaz del sistema con un ejemplo de prueba realizado.



Fig. 17 Sistema web para la detección de PPE

V. DISCUSIÓN DE RESULTADOS

Tras analizar los resultados obtenidos, se demostró que el modelo YOLOv8 presentó una mAP (Mean Average Precision) del 93.22%, superior en comparación con las alternativas evaluadas, como SSD (Single Shot Multibox Detector) que llegó alcanzar el 57.02% y Faster R-CNN, el 58.04%, durante las primeras 50 épocas de entrenamiento. Debido a ello, como el análisis de la curva de aprendizaje reveló una tendencia distintiva en el modelo YOLOv8, se realizó la extensión del entrenamiento a lo largo de 350 épocas adicionales para explorar su potencial, dando un total de 400 épocas.

En cuanto a las métricas obtenidas por el modelo YOLOv8, la curva de precisión alcanza un notable 91.18%, lo que indica la capacidad del modelo para identificar con precisión los objetos clasificados como positivos, minimizando los falsos positivos y asegurando una alta fiabilidad en la clasificación. Asimismo, la curva de recall

registra un 89.67%, lo que refleja la eficacia del modelo para detectar la mayoría de los objetos relevantes en las imágenes, reduciendo los falsos negativos y garantizando que pocos objetos de interés queden sin detectar. Por último, el mAP (Mean Average Precision) se sitúa en un sobresaliente 93.22%, lo que sintetiza el equilibrio entre precisión y recall, reflejando la habilidad del modelo para identificar y categorizar objetos de manera eficaz, manteniendo una consistencia notable en su rendimiento general. Los resultados obtenidos corroboran el destacado desempeño del modelo en tareas de detección de objetos, posicionándolo como una herramienta sólida y confiable.

En cuanto a investigaciones previas, algunos trabajos se centraron exclusivamente en la detección de casco, si lo utiliza o no. Por ejemplo, en la referencia [30], se reportó un mAP de 83.89%, inferior al obtenido en esta investigación como se citó anteriormente. Lo cual evidencia el avance significativo alcanzado por nuestra propuesta en términos de precisión y consistencia. Por otro lado, además de mejorar el desempeño en la detección de cascos, hemos ampliado el alcance al incluir clases adicionales como "Chaleco", "No-Chaleco", "Guante" y "No-Guante", logrando una detección integral de equipos fundamentales de protección personal (PPE).

Otros estudios previos, como el presentado en [28], se enfocaron en la detección de PPE con modelos basados en YOLO y clasificadores adicionales, logrando un máximo de 72.3% de mAP. Sin embargo, nuestra investigación no solo evidencia una mejora considerable en el desempeño, alcanzando un mAP de 93.22% con YOLOv8, sino que también amplía el alcance al incluir la detección de una mayor variedad de elementos de PPE, como "Guante", y su respectiva clase negativa ("No-Guante").

Investigaciones previas, como el descrito en [37], han logrado avances significativos en la detección de PPE utilizando modelos de última generación como YOLOv9-e, alcanzando una precisión del 70.9%. No obstante, nuestra investigación evidencia un progreso notable, con el modelo YOLOv8 superando esta cifra al lograr un mAP de 93.22%, acompañado de una precisión de 91.18% y un recall de 89.67%. Estos resultados no solo destacan la efectividad de nuestro enfoque, sino que también demuestran una capacidad superior para detectar una variedad más amplia de elementos de PPE, como Casco, Chaleco y Guante, junto con sus respectivas clases negativas.

Por último, resaltar que en relacion costo-beneficio se obtuvo un margen promedio del 15% menos de inversión y en promedio de precisión 18.5% más a comparación de las investigaciones incluidas en la comparativa de resultados.

VI. CONCLUSIÓN Y TRABAJOS FUTUROS

En la revisión bibliográfica realizada, se constató que los modelos de redes neuronales convolucionales son los más empleados en la detección de objetos, como es el caso de los equipos de protección personal, utilizando imágenes. Se seleccionaron diversos conjuntos de datos a través de la plataforma Roboflow, los cuales fueron analizados y preprocesados para su uso en la construcción de tres modelos: Faster R-CNN, SSD (Single Shot Multibox Detector) y YOLOv8. Tras el entrenamiento, el modelo YOLOv8 demostró ser el más eficaz, por lo que se procedió a afinarlo durante 350 épocas, dando un total de 400 épocas, alcanzando un mAP de 93.22%, un recall de 89.67% y una precisión de 91.18%. Con el modelo entrenado, se desarrolló un sistema web que implementa el modelo, el cual fue posteriormente alojado en un servidor en la nube para su acceso remoto.

Se identificó una menor cantidad de imágenes disponibles para algunas clases del conjunto de datos, lo que influyó en el aprendizaje del modelo. Sin embargo, el modelo evidenció un desempeño robusto y consistente en la mayoría de las categorías, lo que destaca su capacidad general para la detección precisa de objetos. En el caso específico de las clases "Guante" y "No-Guante", aunque presentaron valores ligeramente inferiores, el desempeño obtenido sigue siendo satisfactorio, con métricas que reflejan un nivel considerable de precisión. En un futuro estudio nos enfocaremos en equilibrar la cantidad de datos entre las clases utilizadas para fortalecer la taza de resultados para potenciar el rendimiento del modelo. Si bien el modelo tiene la capacidad de detectar entradas y salidas en el enfoque, en las clases como “persona” se planea agregar la enumeración de entidades para mejor orden, además de implementar técnicas como Pruning y Quantization que nos permitan para reducir la demanda de requisitos de Hardware y Software del modelo, mejorando su desempeño en dispositivos con recursos más limitados.

REFERENCIAS

[1] E. Forcael, M. Medina, A. Opazo-Vega, F. Moreno, y G. Pincheira, "Additive manufacturing in the construction industry," *Automation in Construction*, vol. 169, Art. no. 105888, Jan. 2025. [En línea]. Disponible en: https://doi.org/10.1016/j.autcon.2024.105888

[2] R. Marchiori, S. Song, y J. Moon, "Developing heat stress training assessments: A training-driven methodology approach to enhance safety in the construction industry," *Journal of Safety Research*, vol. 92, pp. 262–271, Feb. 2025. [En línea]. Disponible en: https://doi.org/10.1016/j.jsr.2024.11.027

[3] Organización Internacional del Trabajo, Seguridad y salud en la construcción, 2022. [En línea]. Disponible: https://www.ilo.org/es/resource/otro/seguridad-y-salud-en-la-construcción-edición-revisada

[4] O. Adeyemo y J. Smallwood, "Impact of Occupational Health and Safety Legislation on Performance Improvement in the Nigerian Construction Industry," Journal of Construction Engineering and Management, 2017. [En línea]. Disponible en: https://doi.org/10.1016/j.proeng.2017.08.008

[5] S. Rasouli, Y. Alipouri, y S. Chamanzad, “Smart Personal Protective Equipment (PPE) for construction safety: A literature review,” Safety

Science, vol. 170. [En línea]. Disponible en: https://doi.org/10.1179/2049396714Y.0000000107
[6] M. Imam, K. Baïna, Y. Tabii, E. M. Ressami, Y. Adlaoui, S. Boufousse, I. Benzakour, y E. H. Abdelwahed, "Integrating real-time pose estimation and PPE detection with cutting-edge deep learning for enhanced safety and rescue operations in the mining industry," *Neurocomputing*, vol. 618, Art. no. 129080, Feb. 2025. [En línea]. Disponible en: https://doi.org/10.1016/j.neucom.2024.129080
[7] O. Karaman, A. Alhudhaif y K. Polat, "Development of smart camera systems based on artificial intelligence network for social distance detection to fight against COVID-19," Computers in Biology and Medicine, vol. 133, pp. 104365, 2021. [En línea]. Disponible: https://doi.org/10.1016/j.asoc.2021.107610
[8] S. Abirami, M. Pethuraj, M. Uthayakumar, and P. Chitra, "A systematic survey on big data and artificial intelligence algorithms for intelligent transportation system," Case Studies on Transport Policy, vol. 17, Art. no. 101247, Sep. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.cstp.2024.101247
[9] M. R. Jong y A. J. de Groof, "Advancement of Artificial Intelligence Systems for Surveillance Endoscopy of Barrett's Esophagus," Digestive and Liver Disease, vol. 56, no. 7, pp. 1126-1130, julio 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.dld.2023.11.038
[10] J. Chin, K.-A. A. Cheon, S. Park, W.-y. Park, and C. Choi, "Efficiency Enhancement in Real-Time Detection and Intervention of Problematic Behaviors in Individuals With Developmental Disabilities Using Artificial Intelligence (AI) CCTV System," Journal of the American Academy of Child & Adolescent Psychiatry, vol. 63, no. 10, Supplement, p. S206, Oct. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.jaac.2024.08.169
[11] T. G. Mai, M. Nguyen, A. Ghobakhlou, W. Q. Yan, B. Chhun, and H. Nguyen, "Decoding a decade: The evolution of artificial intelligence in security, communication, and maintenance within the construction industry," Automation in Construction, vol. 165. [En línea]. Disponible en: https://doi.org/10.1016/j.autcon.2024.105522
[12] M. Sujkowski, J. Kozuba, P. Uchroński, A. Banaś, P. Pulit y L. Gryżewska, "Artificial Intelligence Systems for Supporting Video Surveillance Operators at International Airport," Transportation Research Procedia, vol. 74, pp. 1284-1291, 2023. [En línea]. Disponible en: https://doi.org/10.1016/j.trpro.2023.11.273
[13] K. Li, T. Shan, H. Wu, Z. Zou, R. Huang, R. Chang, y A. Shrestha, "Technology opportunity discovery linking artificial intelligence and construction technologies: A graph convolution network-based approach" [En línea]. Disponible en: https://doi.org/10.1016/j.engappai.2024.109401
[14] R. G. Pirbalouti, B. Behnam, and M. Karimi Dehkordi, "A risk-based approach to identify safety-critical equipment in process industries," Results in Engineering, vol. 20, Art. no. 101448, [En línea]. Disponible en: https://doi.org/10.1016/j.rineng.2023.101448
[15] X. Xu, Q. Li, Z. Du, H. Rong, T. Wu, S. Wang, S. Li, and J. Zheng, "Recognition of concrete imperfections in underwater pile foundation based on binocular vision and YOLOv8," *KSCE Journal of Civil Engineering*, in press, pre-proof, p. 100075, Oct. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.kscej.2024.100075
[16] D. Wang, Y. Zhang, R. Zhang, G. Nie, and W. Wang, "Detection and assessment of post-earthquake functional building ceiling damage based on improved YOLOv8," *Journal of Building Engineering*, vol. 98, p. 111315, Dec. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.jobe.2024.111315
[17] H. Sun, A. Yue, W. Wu, and H. Yang, "Enhanced marine fish small sample image recognition with RVFL in Faster R-CNN model," *Aquaculture*, vol. 595, Part 1, p. 741516, Jan. 2025. [En línea]. Disponible en: https://doi.org/10.1016/j.aquaculture.2024.741516
[18] J. Cai, Y. Makita, Y. Zheng, S. Takahashi, W. Hao, y Y. Nakatoh, "Single shot multibox detector for honeybee detection," Computers and Electrical Engineering, vol. 104, parte B, p. 108465, Dec. 2022. [En línea]. Disponible en: https://www.sciencedirect.com/science/article/abs/pii/S0045790622006802
[19] W. Qian, Z. Zhu, C. Zhu, W. Luo, y Y. Zhu, "Efficient deployment of Single Shot Multibox Detector network on FPGAs," *Integration*, vol. 99, p. 102255, Nov. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.vlsi.2024.102255
[20] C. B. Nwaneto, C. Yinka-Banjo, and O. Ugot, "An object detection solution for early detection of taro leaf blight disease in the West African sub-region," *Franklin Open*, vol. 10, no. 1, Mar. 2025. [En línea] Disponible en: https://doi.org/10.1016/j.fraope.2024.100197
[21] K. Vinoth and P. Sasikumar, "VINO_EffiFedAV: VINO with efficient federated learning through selective client updates for real-time autonomous vehicle object detection," Results in Engineering, vol. 25, Mar. 2025, Art. no. 103700. [En línea]. Disponible en: https://doi.org/10.1016/j.rineng.2024.103700
[22] F. Martinelli, F. Mercaldo, and A. Santone, "Object Detection and Localisation in Thermal Images by means of UAV/Drone," *Procedia Computer Science*, vol. 225, pp. 2234-2243, 2023. [En línea]. Disponible en: https://doi.org/10.1016/j.procs.2023.10.214
[23] C. Zhang, X. Chen, P. Liu, B. He, W. Li, and T. Song, "Automated detection and segmentation of tunnel defects and objects using YOLOv8-CM," *Tunnelling and Underground Space Technology*, vol. 150, Aug. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.tust.2024.105857
[24] U. Ulusoy, O. Eren, y A. Demirhan, "Development of an obstacle avoiding autonomous vehicle by using stereo depth estimation and artificial intelligence based semantic segmentation," *Engineering Applications of Artificial Intelligence*, vol. 126, Part A, Art. no. 106808, Nov. 2023. [En línea]. Disponible en: https://doi.org/10.1016/j.engappai.2023.106808
[25] Y. Li, H. Wei, Z. Han, J. Huang, y W. Wang, "Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks," Advances in Civil Engineering, vol. 2020, Art. no. 9703560, Sep. 2020. [En línea]. Disponible en: https://doi.org/10.1155/2020/9703560
[26] "Artificial intelligent systems for vehicle classification: A survey," Engineering Applications of Artificial Intelligence, vol. 129, Art. no. 107497, [En línea]. Disponible en: https://doi.org/10.1016/j.engappai.2023.107497
[27] J. Zapata-Paulini y L. Lescano, "Technological advances in high mountains: Development of an augmented reality and IoT Application for snow cover monitoring on Huascaran, Peru," Environmental Advances, [En línea]. Disponible en: https://doi.org/10.1016/j.envadv.2024.100586
[28] Nipun D. Nath, Amir H. Behzadan, Stephanie G. Paal, " Deep learning for site safety: Real-time detection of personal protective equipment" *Automation in Construction*, vol. 112, 2020. [En línea]. Disponible en: https://doi.org/10.1016/j.autcon.2020.103085
[29] H. Lyu, F. Qiu, L. An, D. Stow, R. Lewison, and E. Bohnett, "Deer survey from drone thermal imagery using enhanced faster R-CNN based on ResNets and FPN," *Ecological Informatics*, vol. 79, p. 102383, Mar. 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.ecoinf.2023.102383
[30] Jixiu Wu, Nian Cai, Wenjie Chen, Huiheng Wang, Guotian Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset" Automation in Construction, vol. 106, 2019. [En línea]. Disponible en: https://doi.org/10.1016/j.autcon.2019.102894
[31] Reis D., Kupe J., Hong J., & Daoudi A., "Real-Time Flying Object Detection with YOLOv8". ArXiv, 2023, [En línea]. Disponible en: https://doi.org/10.48550/arXiv.2305.09972
[32] Terven J., Córdova-Esparza D. M., "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas" Machine Learning and Knowledge Extraction, 2023, [En línea]. Disponible en: https://doi.org/10.48550/arXiv.2305.09972
[33] Shi P., Xu X., Ni J., Xin Y., Huang W., Han S., "Underwater Biological Detection Algorithm Based on Improved Faster-RCNN", Water, 13(17), 2420, 2021, [En línea]. Disponible en: https://doi.org/10.3390/w13172420
[34] Wang Q., Wang Z. Li J., Yang Y.,Liu P., "Improvement of Faster-RCNN Detection Algorithms for Small Size Line Accessory Equipment", Journal of Physics, 2020, [En línea]. Disponible en: https://doi.org/10.1088/1742-6596/1453/1/012007
[35] Wang Y., Niu P., Guo X., Yang G., Chen J., "Single Shot Multibox Detector With Deconvolutional Region Magnification Procedure" IEEE Access, 9, 47767-47776, 2021, [En línea]. Disponible en: https://doi.org/10.1109/ACCESS.2021.3068486
[36] Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C., Berg A., "SSD: Single Shot MultiBox Detector", 21-37,2015, [En línea]. Disponible en: https://doi.org/10.1007/978-3-319-46448-0_2
[37] Hafiz M., Afshin R., "SH17: A dataset for human safety and personal protective equipment detection in manufacturing industry", Journal of Safety Science and Resilience, 2024, [En línea]. Disponible en: https://doi.org/10.1016/j.jnlssr.2024.09.002