ISSN: 3061-810X
Año 2, Núm. 1, Septiembre -Diciembre 2025 pp. **9**

Modelo neuronal para la identificación de elementos de protección personal aplicado a máquinas industriales de riesgo

Gilbert F Pérez-García[1]; Jorge A Orozco-Torres[1]; Jorge A Mijangos-López[1]; Oscar J Rincón-Zapata[1]; Héctor R Hernández-De León[1]; María C. Salgado Gutiérrez.

1 Tecnológico Nacional de México/I. T. de Tuxtla Gutiérrez, Tuxtla Gutiérrez, Chiapas, México.

Área: Informática e Inteligencia Artificial.

Resumen

En entornos industriales, alrededor del 70 % de los accidentes se relacionan con el uso inadecuado o con la omisión del equipo de protección personal (EPP) y más del 85 % de los incidentes graves están asociados específicamente a la falta de casco, chaleco o guantes. Para abordar esta situación, en el presente estudio se muestra el desarrollo y la evaluación de un modelo de visión artificial basado en la arquitectura YOLOv8, entrenado con un conjunto mixto de aproximadamente 3,000 imágenes provenientes de fuentes libres y de campo, anotadas manualmente en siete categorías de EPP (Helmet, Ear-Protection, Boots, Glass, Gloves, Vest, Mask). Los resultados obtenidos muestran una precisión general de hasta 75.3 %, un recall de 73.6 %, y un mAP@50 de 75.9 %, con valores superiores al 83 % de aciertos en al menos tres clases: Gloves (89 %), Vest (89 %) y Glass (83 %), según la matriz de confusión. La clase con mejor desempeño fue Gloves, mientras que casco alcanzó un acierto del 85 %, y Vest se mantuvo entre las clases más robustas. El sistema se ejecuta por encima de 25 fotogramas por segundo en GPU, lo que permite su implementación en tiempo real en entornos industriales. Con estos resultados se confirma la viabilidad de implementar sistemas de monitoreo no intrusivos para mejorar la seguridad laboral, al ofrecer alertas en tiempo real sobre el cumplimiento de normativas de EPP.

Palabras clave: detección de objetos; equipo de protección personal; YOLOv8; visión por computador; seguridad laboral.

Abstract

In industrial environments, approximately 70% of accidents are related to the improper use or omission of personal protective equipment (PPE), and over 85% of serious incidents are specifically associated with the absence of helmets, vests, or gloves. To address this issue, the present study presents the development and evaluation of a computer vision model based on the YOLOv8 architecture, trained on a mixed dataset of approximately 3,000 images from both public sources and field data, manually annotated into seven PPE categories (helmet, ear protection, glasses, mask, vest, gloves, and boots). The results show an overall precision of up to 75.3%, a recall of 73.6%, and a mAP@50 of 75.9%, with accuracy values exceeding 83% in at least three classes: gloves (89%), vest (89%), and glasses (83%), according to the confusion matrix.

The best-performing class was gloves, while helmets achieved an accuracy of 85%, and vests remained among the most robustly detected items. The system operates at over 25 frames per second on GPU, enabling real-time deployment in industrial settings. These results confirm the feasibility of implementing non-intrusive monitoring systems to enhance occupational safety by providing real-time alerts on PPE compliance.

Keywords: object detection; personal protective equipment; YOLOv8; computer vision; workplace safety.

Introducción

La seguridad industrial exige que cada trabajador utilice el equipo de protección personal adecuado en sitios de riesgo, como sectores de construcción o plantas industriales. Sin embargo, el control sobre el cumplimiento de esta normativa no siempre es riguroso, lo que se refleja en la alta incidencia de accidentes laborales (Al-Bayati et al. (2018)).

Diversos estudios (Kang. (2018); Khoshakhlagh et al. (2024); Al-Bayati & York. (2018)), indican que entre el 50 al 70% de los accidentes en la industria ocurrieron por la falta de uso de equipo de protección personal (casco, guantes, googles, etc.), y que más del 85% de los casos fatales documentados, se identificó específicamente por la falta del uso del casco, guantes y chaleco, elementos determinantes de protección.

De manera tradicional, la verificación del uso de equipo de protección personal se realiza mediante inspecciones manuales o, en algunos casos, con sensores específicos. Sin embargo, estas soluciones dependen de una intervención humana constante o de infraestructuras tecnológicas costosas y poco escalables. En contraste, un sistema de visión por computadora permite llevar a cabo un monitoreo continuo, no invasivo y en tiempo real durante toda la jornada laboral. Por ejemplo, los métodos basados en aprendizaje profundo tienen la capacidad de analizar imágenes y aprender a identificar automáticamente objetos de interés, para el caso de este estudio: guantes, chalecos, cascos, mascarillas, etc., sin la necesidad de que exista interacción física (Ahmed et al. (2023)).

Por este motivo, en el presente trabajo se propone el diseño de un modelo de visión artificial basado en el algoritmo de YOLOv8, empleando técnicas de aprendizaje profundo sobre un conjunto de aproximadamente 3,000 imágenes anotadas manualmente. El modelo fue entrenado para la identificación simultanea de siete clases de equipo de protección personal en una sola etapa. A diferencia de los métodos de detección en dos etapas, como los modelos R-CNN (Ren et al. (2015)), que dividen el proceso en propuestas de regiones y de clasificación. La arquitectura de la familia YOLO realiza la detección en una única fase, lo que permite alcanzar velocidades de procesamiento cercanas al tiempo real, haciéndolo adecuadas para su implementación en aplicaciones industriales (Redmon et al. (2016)).

Fundamentos teóricos

Los sistemas de detección de objetos modernos se basan en redes neuronales convolucionales (CNN) que aprenden a localizar y clasificar objetos en imágenes. En general existen dos enfoques: detectores de una sola etapa (one-stage), como YOLO (You Only Look Once), que realizan simultáneamente la localización y clasificación, y detectores de dos etapas (two-stage) como Faster R-CNN (Ren et al. (2015)), que primero proponen regiones candidatas y luego clasifican. Los detectores de una sola etapa suelen ser más rápidos y adecuados para aplicaciones en tiempo real, sacrificando en ocasiones algo de precisión. El algoritmo YOLO original de Redmon et al. (2016) introdujo esta idea de procesar toda la imagen en un solo escaneo para predecir múltiples cajas con etiquetas de clase en cada celda de la matriz. Sus versiones sucesivas han mejorado progresivamente la arquitectura (YOLOv3, YOLOv4, YOLOv5, etc.). La versión de YOLOv8, incorpora una arquitectura unificada capaz de abordar múltiples tareas de visión por computadora, como detección de objetos, segmentación por instancias, clasificación de imágenes, estimación de pose y detección con cajas orientadas (OBB). Esta versión introduce una cabeza de detección libre de anclas (anchor-free) y una estructura de red optimizada, eliminando la necesidad de cajas predefinidas y simplificando tanto el proceso de entrenamiento como el cálculo durante la inferencia (Ultralytics, 2023; Yaseen, 2024).

En el contexto del EPP, la literatura muestra múltiples trabajos de detección automática de casco, chaleco, gafas y guantes en entornos de trabajo. Por ejemplo, Ahmed et al. (2023) emplearon CNN para monitorear varios tipos de EPP en tiempo real, logrando detección de ocho clases (cascos de colores, personas, chalecos, gafas) en un dataset de 1.700 imágenes. De forma similar, Barlybayev et al. (2024) presentan un estudio comparativo usando YOLOv8 sobre datasets de EPP (cascos, chalecos, goggles, etc.), y concluyen que las variantes más grandes de YOLOv8 (modelos l y x) obtienen las mejores métricas de precisión y recall.

Materiales y métodos

En esta sección se describen los materiales utilizados y la metodología seguida para el desarrollo de esta investigación. El objetivo principal es entrenar un modelo de visión artificial capaz de identificar diferentes elementos de equipo de protección personal en ambientes industriales. Para ello, se propone la utilización de un sistema compuesto por una computadora monoplaca y una cámara de video convencional, sin requerimientos específicos de hardware, con esto se busca que físicamente sea flexible y robusta para operar en condiciones reales industriales, sin la necesidad de requerimientos especiales, lo que facilita su integración en diferentes procesos de la industria.

ISSN: 3061-810X
Año 2, Núm. 1, Septiembre -Diciembre 2025 pp. **12**

Modelo de visión artificial

Para la tarea de identificación de elementos de protección personal (EPP), se seleccionó el algoritmo YOLOv8 (Ultralytics, 2023), en su variante small (YOLOv8s), por su balance entre precisión y eficiencia computacional, lo cual permite su ejecución en una computadora monoplaca. El modelo YOLOv8s fue ajustado mediante fine-tuning a partir de un checkpoint preentrenado en el conjunto de datos COCO (Lin et al. (2014)), utilizando un total de 3,000 imágenes manualmente anotadas en formato YOLO. El conjunto de datos incluyó imágenes de licencia libre y capturas propias de campo, representando escenas reales de trabajadores en contextos industriales diversos. Se definieron siete clases de EPP: Helmet, Ear-Protection, Glass, Mask, Vest, Gloves y Boots. El conjunto fue dividido mediante separación aleatoria estratificada en 80 % para entrenamiento, 10 % para validación y 10 % para prueba.

Para incrementar la variabilidad y robustez del entrenamiento, especialmente en el subconjunto de imágenes propias, se aplicaron técnicas de aumento de datos como mosaic, mixup, rotaciones, espejados horizontales y otras transformaciones geométricas.

El entrenamiento se llevó a cabo durante 150 épocas, utilizando el optimizador SGD con un learning rate inicial de 0.0025, esquema de decaimiento cosenoidal (cosine LR scheduler) y un warm-up de 5 épocas. Los hiperparámetros clave incluyeron la pérdida BCE (Binary Cross-Entropy) para clasificación, y una combinación de Distributional Focal Loss (DFL) y CIoU (Complete Intersection over Union) para la regresión de los cuadros delimitadores.

Todo el proceso de entrenamiento fue implementado en Python mediante la librería oficial Ultralytics YOLO (Ultralytics, 2023), utilizando una unidad de procesamiento gráfico (GPU) para acelerar la optimización.

Flujo de trabajo

El flujo de trabajo seguido en el desarrollo del modelo se estructuró en cuatro etapas principales. En la Figura 1 se muestra un diagrama de la metodología general de la propuesta realizada.

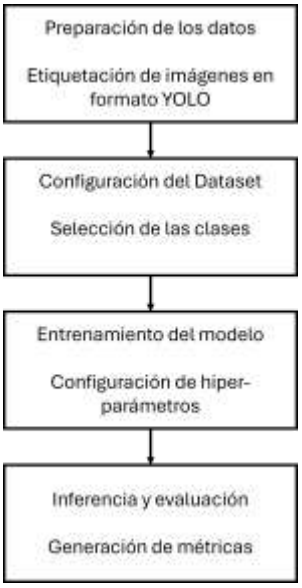

Figura 1. Metodología general de trabajo.

En la primera etapa, se realizó la preparación de los datos, que consistió en la recolección y anotación de imágenes mediante cuadros delimitadores (bounding boxes), con las etiquetas correspondientes en formato YOLO.

La estructura del conjunto de datos se organizó en carpetas separadas para imágenes (images/) y

anotaciones (labels/), facilitando su interpretación por el framework de entrenamiento.

En esta etapa se separaron las 3000 imágenes en tres grupos: el 80% de las imágenes se destinó para el entrenamiento, 10% para la validación y 10% restante para las pruebas. En la Tabla 1 se muestra la distribución del número de imágenes utilizados para cada una de las clases, considerando el 80% de imágenes utilizadas en el entrenamiento.

Tabla 1. Distribución de las imágenes de las clases utilizadas para el entrenamiento.

Clase de EPP	Patrones (imágenes)
Helmet	350
Ear-Protection	150
Glass	400
Mask	300
Vest	450
Gloves	500
Boots	250
Total (80%)	**2,400**

En la segunda etapa, se llevó a cabo la configuración del conjunto de datos, mediante la creación del archivo dataset.yaml, donde se especificaron las rutas a los subconjuntos de entrenamiento, validación y prueba, así como la lista de nombres de las clases correspondientes a los diferentes elementos de EPP.

- Helmet
- Ear-Protection
- Glass
- Mask
- Vest
- Gloves
- Boots

En la tercera etapa, se realizó el proceso de entrenamiento, para esto, se utilizó la función model.train(data="dataset.yaml", ...) de la librería Ultralytics, y se configuraron hiperparámetros como el número de épocas, el tamaño de imagen de entrada, el tamaño de lote (batch size), la tasa de aprendizaje y los métodos de aumento de datos.

Durante el proceso de entrenamiento se realizó una validación automática, y se seleccionó el modelo con mejor desempeño en el conjunto de validación, el cual fue almacenado en el archivo best.pt.

La última etapa fue la fase de inferencia y evaluación, en esta etapa se ejecutaron pruebas sobre el conjunto de prueba, generando métricas estándar de detección como precisión, recall y mAP (mean Average Precision), que permitieron cuantificar el rendimiento final del modelo y evaluar su viabilidad para su implementación en escenarios reales.

Resultados y discusión

En esta sección se presentan los resultados obtenidos durante el proceso de entrenamiento del modelo, junto con un análisis sobre las posibles causas de los patrones observados. El entrenamiento se llevó a cabo durante 150 épocas mediante una estrategia de fine-tuning aplicada sobre la arquitectura YOLOv8. El modelo fue entrenado para detectar siete clases de equipo de protección personal (EPP).

ISSN: 3061-810X
Año 2, Núm. 1, Septiembre -Diciembre 2025 pp. **14**

En la Figura 2, se muestra los resultados obtenidos de las curvas de pérdidas en la etapa de entrenamiento y de validación. Las curvas muestran que todos los componentes de pérdida disminuyen con el entrenamiento.

La pérdida de caja (box_loss) de entrenamiento desciende desde el valor inicial 1.69 (época 1) hasta el valor de 0.85 (época 150); mientras que en la etapa de validación desciende desde 1.44 hasta 1.04. En el caso de las pérdidas de clase (cls_loss), esta se reduce de 5.87 hasta aproximadamente 1.05 en el proceso de entrenamiento y de aproximadamente 3.67 hasta aproximadamente 1.66 en la etapa de validación.

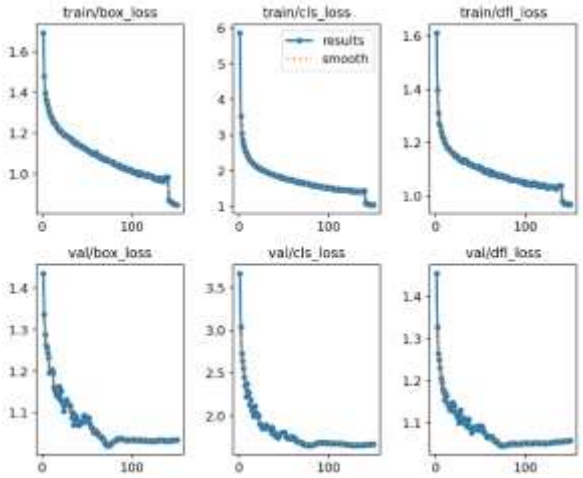

Figura 2. Curvas de pérdidas obtenidas después del entrenamiento.

La DFL loss (localización de bordes) desciende de 1.61 hasta 0.97 (durante el entrenamiento) y de aproximadamente 1.45 hasta 1.06 para la etapa de validación.

Durante la fase de warm-up (esta etapa corresponde a las primeras cinco épocas) se observa un descenso muy rápido en las pérdidas, luego las pérdidas continúan descendiendo hasta estabilizarse gradualmente alrededor de las épocas 80–100. A partir de entonces, las curvas de pérdida se aplanan, indicando convergencia; en particular, a final de entrenamiento la diferencia entreno/validación en cls_loss es notable (1.046 vs 1.668), sugiriendo que el modelo aprendió mejor la clasificación de entrenamiento que la generalización.

En el caso de las curvas de las métricas de precisión, recall y mAP (ver Figura 3), De manera general se puede mencionar que las métricas de detección mejoran continuamente.

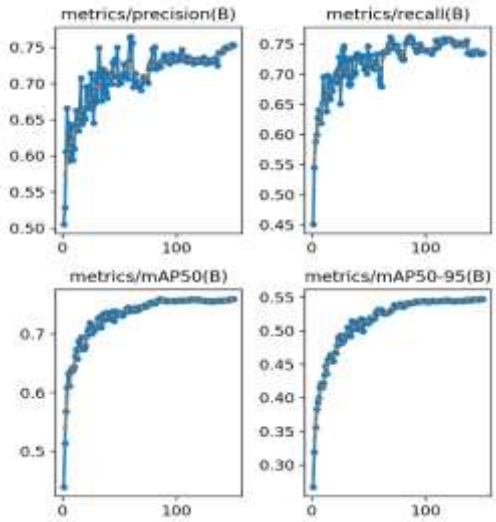

Figura 3. Curvas de las métricas obtenidas después del entrenamiento.

Durante el proceso de entrenamiento, se observó una mejora progresiva en todas las métricas de evaluación del modelo. La precisión aumentó desde un valor inicial de aproximadamente 0.50 hasta alcanzar un valor final de 0.75, reflejando una mejora significativa en la capacidad del modelo para reducir falsos positivos. De manera similar, la métrica de recall mostró un crecimiento

ISSN: 3061-810X
Año 2, Núm. 1, Septiembre -Diciembre 2025 pp. **15**

desde 0.45 hasta 0.74, indicando una mejora en la detección efectiva de los objetos relevantes.

En cuanto al mAP\@50 (métrica que evalúa la precisión media considerando un umbral de intersección sobre unión (IoU) de 0.50), el valor se incrementó de 0.44 a 0.76, mostrando un mejor desempeño en la localización general de los objetos. Finalmente, el mAP\@50-95 (métrica que evalúa el promedio de precisión media en un rango más estricto de umbrales IoU (de 0.50 a 0.95)), aumentó de 0.27 a 0.55, lo cual sugiere una mejora significativa en la precisión espacial de las predicciones a lo largo del entrenamiento.

De las gráficas se observa que las métricas de Precision, Recall, mAP@50, mAP@50-95, mejoran continuamente y alcanzan saturación hacia el final del entrenamiento.

Al inicio, en la época 1, el modelo obtiene valores bajos:

- Precision=0.505
- Recall=0.451
- mAP@50=0.439
- mAP@50-95=0.267

Hacia la época 50, se observaron valores altos, por ejemplo:

- Precision=0.706
- Recall=0.728
- mAP@50=0.734

Al final del entrenamiento (época 150) se alcanza

- Precision=0.753
- Recall=0.756
- mAP@50=0.759
- mAP@50-95=0.548

Esto significa que el modelo es capaz de detectar correctamente entre 75–76% de los objetos bajo IoU≥0.50, pero su desempeño cae al 55% promedio cuando se exige un IoU más estricto (≥0.95).

La matriz de confusión se observa en la Figura 4, en esta se observa el desempeño por clase. Se observa que la mayoría de las clases tiene alta exactitud de predicción, con valores diagonales elevados, pero existen confusiones consistentes con la clase background (falsos negativos) y, en menor medida, entre clases similares.

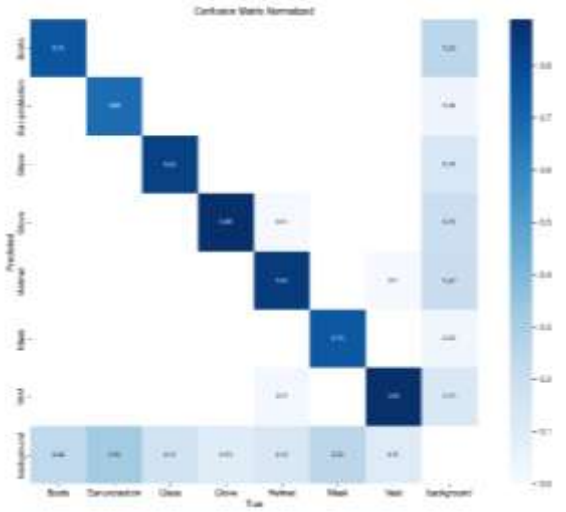

Figura 4. Matriz de confusión de las clases utilizadas durante el entrenamiento.

A continuación, se describe cada una de las clases utilizadas durante el entrenamiento de acuerdo a los datos obtenidos por la matriz de confusión normalizada.

Casco (Helmet): Aproximadamente 85% de los casos reales de casco se predicen correctamente como casco, mientras que un 13% son clasificados como fondo. Hay una pequeña confusión mutua con guantes (1% de guantes

reales detectados como casco y viceversa) y con mascarillas (aproximadamente 1%).

Protección auditiva (Ear-protection): solamente el 68% aciertos; en el caso del 32% de los casos los protectores de oído reales no se detectan (se confunden como fondo). El modelo predice falsamente protectores auditivos donde no los hay (producto de recursos), lo que indica baja precisión para esta clase.

Gafas (Glass): El 83% de las gafas reales se detectan correctamente. El resto (aproximadamente el 17%) son no detectadas (se confunden como fondo). Casi no hay confusión con otras clases.

Mascarilla (Mask): Alrededor del 75% de aciertos. Se ven errores hacia fondo (20%) y un pequeño 3% de guantes reales confundidos como mascarilla. Esto sugiere que las mascarillas (objetos planos) son detectadas con menos fiabilidad que otros EPP tridimensionales.

Chaleco (Vest): muy alto desempeño (el 89% son aciertos), con sólo un 11% de falsos negativos confundidos como fondo. Prácticamente no se confunde con otras clases.

Guantes (Glove): el 89% de aciertos. Se observan confusiones leves: el 1% de guantes reales fueron detectados como casco (quizá por similitud en color/textura) y el 10% fueron omitidos (predichos como fondo).

Botas (Boots): aproximadamente el 76% de aciertos. Un 24% de botas reales no se detectaron (fueron falsos negativos como fondo). No hay confusiones con otras clases, pero sí un nivel moderado de detecciones fantasma de botas en imágenes de fondo (aunque aproximadamente el 25% de las predicciones de botas son erróneas).

Conclusiones

Los resultados obtenidos por el entrenamiento demostraron datos de interés, tanto en términos de desempeño general como de identificación de áreas susceptibles de mejora para obtener un modelo más robusto. La matriz de confusión demostró puntos débiles en ciertas clases de objetos, particularmente si los objetos son pequeños o confusos, como es el caso de los elementos de protección auditiva o de las mascarillas. Una posible mejora consiste en aumentar la cantidad de datos para estas clases, ya sea recolectando más imágenes o mediante técnicas de aumentación focalizados como recortes o zoom. También podrían ajustarse hiperparámetros, incluyendo entrenamiento multiescala y redefinición de anchors y strides, para mejorar la detección de objetos pequeños. Otra alternativa para mejorar la detección de objetos pequeños, es aplicar estrategias como tilín, esto aumentaría la resolución local de las regiones de interés.

Finalmente, para facilitar su implementación en hardware limitado, es recomendable exportar el modelo a formatos optimizados (ONNX, TensorRT) y aplicar cuantización. Aunque YOLOv8s ofrece un buen equilibrio entre precisión y eficiencia, evaluar versiones más ligeras como YOLOv8n sería adecuado cuando la latencia es crítica.

Referencias bibliográficas

Chen, S., & Demachi, K. (2020). *A Vision-Based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station*. **Applied Sciences, 10**(15), 5129. https://doi.org/10.3390/app10155129

Elesawy, A., Abdelkader, E. M., & Osman, H. (2024). A Detailed Comparative Analysis of You Only Look Once-Based Architectures for the Detection of Personal Protective Equipment on Construction Sites. **Engineering, 5**(1), 19. https://doi.org/10.3390/eng5010019

Li, C., Luo, B., Yin, T., et al. (2025). SD-YOLOv5: a rapid detection method for personal protective equipment on construction sites. *Frontiers in Built Environment*, 11, Article 1563483. https://doi.org/10.3389/fbuil.2025.1563483

Liu, Y., & Wang, J. (2024). Personal Protective Equipment Detection for Construction Workers: A Novel Dataset and Enhanced YOLOv5 Approach. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3382817

Lo, J.-H., Lin, L.-K., & Hung, C.-C. (2023). Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm. **Sustainability, 15**(1), 391. https://doi.org/10.3390/su15010391

Malaikrisanachalee, S., Wongwai, N., & Kowcharoen, E. (2025). ESPCN-YOLO: A High-Accuracy Framework for Personal Protective Equipment Detection Under Low-Light and Small Object Conditions. **Buildings, 15**(10), 1609. https://doi.org/10.3390/buildings15101609

Massiris, M., Delrieux, C., & Fernández Muñoz, J. A. (2018). Detección de equipos de protección personal mediante red neuronal convolucional YOLO. *Memorias de las XXXIX Jornadas de Automática*, España. https://doi.org/10.17979/spudc.9788497497565.1022

Singh, S., Jain, A., Sharma, V., & Girdhar, P. (2024). PPE Detection for Construction Site Safety Leveraging YOLOv8. *Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024)*. SSRN: https://ssrn.com/abstract=4923318