

Article

Autonomous Mobile Station for Artificial Intelligence Monitoring of Mining Equipment and Risks

Gabriel País Cerna ¹, Germán Herrera-Vidal ^{2,*} and Jairo R. Coronado-Hernández ^{3,*}

¹ Faculty of Engineering, Universidad Andrés Bello, Santiago 8320000, Chile; gabriel.pais@unab.cl

² Industrial Engineering Program, Ciptec Research Group, Fundación Universitaria Tecnológico Comfenalco, Cartagena 130007, Colombia

³ Department of Productivity and Innovation, Universidad de la Costa, Barranquilla 80007, Colombia

* Correspondence: herreravg@tecnocomfenalco.edu.co (G.H.-V.); jcoronad18@cuc.edu.co (J.R.C.-H.)

Abstract: Artificial intelligence in the mining industry is key to improving safety, optimizing resources, and ensuring sustainable operations in complex environments. The main objective of this research is to develop an autonomous mobile station equipped with artificial vision and artificial intelligence to identify and track equipment, people, and animals in critical areas of mining operations, issuing real-time alerts to reduce occupational risks and improve operational control. The research is applied with an experimental approach, designed to validate the effectiveness of the proposed system in real open-pit mining environments. The proposed methodology consisted of five stages: (i) Selection of data collection equipment, (ii) Definition of the positioning scheme, (iii) Incorporation of the communication system, (iv) Data processing and transformation, and (v) Equipment identification and tracking. The results showed an average accuracy of 98% in the validation and 95% in the test, achieving perfect performance (100%) in key categories such as excavators and drills, highlighting the potential of this technology to transform mining towards safer and more efficient standards.

Keywords: artificial intelligence; artificial vision; mining industry; autonomous mobile station



Academic Editor: Yosoon Choi

Received: 11 March 2025

Revised: 3 April 2025

Accepted: 8 April 2025

Published: 10 April 2025

Citation: Cerna, G.P.; Herrera-Vidal, G.; Coronado-Hernández, J.R. Autonomous Mobile Station for Artificial Intelligence Monitoring of Mining Equipment and Risks. *Appl. Sci.* **2025**, *15*, 4197. <https://doi.org/10.3390/app15084197>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The mining industry faces several complex challenges that threaten its sustainability, efficiency, and capacity for innovation in an ever-changing global environment. These challenges include declining ore grades in shallow deposits, the need to mine resources at extreme depths, and social and environmental pressures arising from the impacts of mining activity [1,2]. The integration of advanced technologies, such as artificial intelligence, Internet of Things (IoT), and automation, has been identified as a key strategy to address these constraints, improving operational efficiency and mitigating risks [3,4]. However, technology adoption is not without barriers, such as resistance to organizational change, lack of local technical capabilities, and high initial implementation costs [5,6]. In this context, solutions such as smart mining, combining advanced analytics and automation, and the circular economy, focused on water recovery and tailings management, have demonstrated a positive impact on productivity and sustainability [7,8]. Similarly, strengthening the links between local suppliers and global companies, together with the implementation of inclusive regulatory policies, is essential to overcome innovation limitations and promote sustainable development in strategic regions such as Latin America [9]. This comprehensive approach, which combines technological innovation, efficient resource management, and

coherent regulatory strategies, positions the mining industry as a key player to face the challenges of the 21st century, ensuring its competitiveness and social responsibility.

The main challenges in the adoption of artificial intelligence (AI) in the mining industry focus on the lack of technical skills, cultural resistance to change, and regulatory barriers to ensure ethical and safe use. Adapting advanced algorithms to dynamic and noisy environments, such as those that characterize mining processes, also represents a significant technical challenge [10,11]. In addition, inadequate technology infrastructure and limited interoperability between intelligent systems hinder the effective integration of AI into mining operations, while the development of legal frameworks to regulate autonomous technologies remains nascent in many regions [12–14]. On the other hand, specific challenges such as reducing greenhouse gas emissions, improving mining safety, and training personnel to work with advanced technologies remain priorities for the sustainability and efficiency of the sector [15–17].

AI applications in mining span a variety of critical processes, from drilling optimization, mineral extraction, and recovery to predictive maintenance and risk management in coal mining operations [10,18]. Intelligent systems have proven effective in advanced flotation monitoring, aerial inspection using drones, and failure prediction in critical machinery [11,19,20]. Likewise, AI technologies have been applied to reduce environmental impact through energy optimization of heavy machinery and structural damage monitoring of industrial equipment [15,21]. More recently, tools such as text mining have been used to analyze technology trends and foster strategic planning in specific regions, including Indonesia and South Africa [22,23].

Predominant methodologies in the use of AI in mining include supervised and unsupervised learning algorithms, hybrid models, and advanced neural network techniques. Expert systems and ConvLSTM models have been especially effective in predicting critical conditions and monitoring processes in real time [11,19]. In addition, data mining algorithms, combined with thematic analysis and natural language processing (NLP) techniques, have facilitated the analysis of large volumes of data, such as technology trends and industry behavioral patterns [24,25]. Other notable approaches include integrated cyber-physical systems for mining automation and advanced simulators for personnel training [14,17]. These technologies and methods have addressed technical, environmental, and operational challenges, laying the foundation for more efficient and sustainable mining.

From a theoretical standpoint, the integration of AI in industrial systems relies on a convergence of computational models capable of abstraction, reasoning, and generalization. This includes not only algorithmic sophistication but also systemic adaptability to contextual and environmental variability, which is especially relevant in the mining sector. Recent developments in neuro-symbolic AI, knowledge graph integration, and meta-learning techniques have extended the functional capacity of mining-focused AI systems by enabling better interpretability, modular transfer learning, and multi-task optimization without losing operational robustness [26–29]. These paradigms allow for more context-aware and resilient decision-making processes, while also enhancing the transparency of AI outputs in high-risk industrial contexts [28,30,31]. Their application within real-time systems, combined with edge processing infrastructure [32,33], supports scalable deployment in geographically distributed and connectivity-constrained operations such as remote mining facilities, where delay-sensitive decisions are critical to safety and efficiency.

The bibliometric analysis, based on a search in Scopus with the Boolean formula ('Neural Networks' OR 'Deep Learning' OR 'Artificial Intelligence') AND ('Mining' OR 'Mineral Processing'), identified 15 articles published between 2022 and 2025. Using VOSviewer, a co-occurrence network with 172 keywords was generated, setting a minimum threshold of 1 occurrence per term. Six main clusters were identified: the red cluster

(sixty terms) focuses on agricultural soils and atomic absorption spectrometry, reflecting environmental concerns. The green cluster (39 terms) groups terms such as ‘machine learning’ and ‘joint angle’, highlighting technical development. The blue cluster (30 terms) relates ‘mining industry’ to ‘environmental impact’, highlighting ecological risks. The yellow cluster (24 terms) emphasizes regulatory compliance, while the violet and light blue clusters (4 terms each) have less impact on the network. This analysis allows us to visualize the main trends at the intersection between artificial intelligence and mining.

In synthesis, artificial intelligence (AI) emerges as a key tool for transforming the mining industry towards a more sustainable and efficient operation. Its applications focus on environmental monitoring, air quality control, and agricultural land management, addressing critical challenges such as pollutant emissions reduction and land use sustainability. Advanced methods such as machine learning and neural networks lead their implementation, enabling predictive analysis and optimization of processes such as ventilation and maintenance in mines. This approach consolidates AI as a strategic pillar to minimize environmental impacts and maximize operational efficiency in mining (See Figure 1).

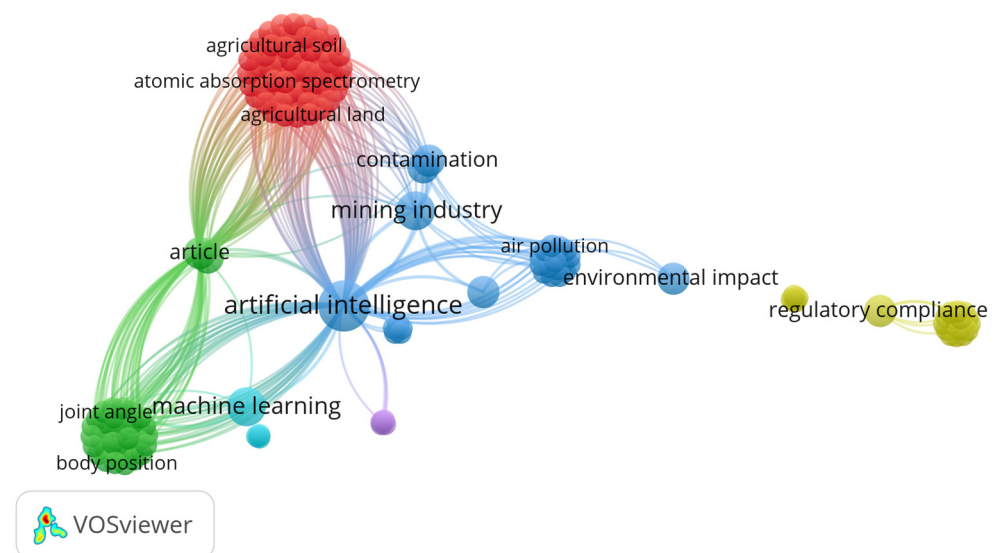


Figure 1. Cooccurrence of keywords with VOSviewer in recent research database.

From another perspective and based on an updated time horizon, the evolution of artificial intelligence (AI) in mining is highlighted, focusing on the transition to more advanced and sustainable technologies. An increase in the adoption of predictive algorithms and hybrid models is projected, with emphasis on air management, emissions reduction, and real-time environmental monitoring. These solutions integrate tools such as digital twins and deep learning systems, strengthening operational efficiency and control of environmental impacts. This scenario positions AI as a key catalyst to address contemporary sustainability challenges in the mining industry (See Figure 2).

The main contributions of this research are the following:

- Development of an autonomous mobile station with artificial vision and artificial intelligence capable of identifying equipment, people, and animals in critical operational areas, optimizing safety and surveillance.
- Implementation of deep learning algorithms to analyze movements and operation times, improving the allocation of mining equipment and correcting inefficiencies not considered by traditional systems.

- Integration of virtual delimitation of risk zones and issuance of automatic alerts in real time when unwanted presence is identified, significantly reducing occupational accidents.
- Application of data-augmented convolutional neural networks (CNN) to achieve 100% accuracy in the identification of key mining equipment during validation tests in real environments.

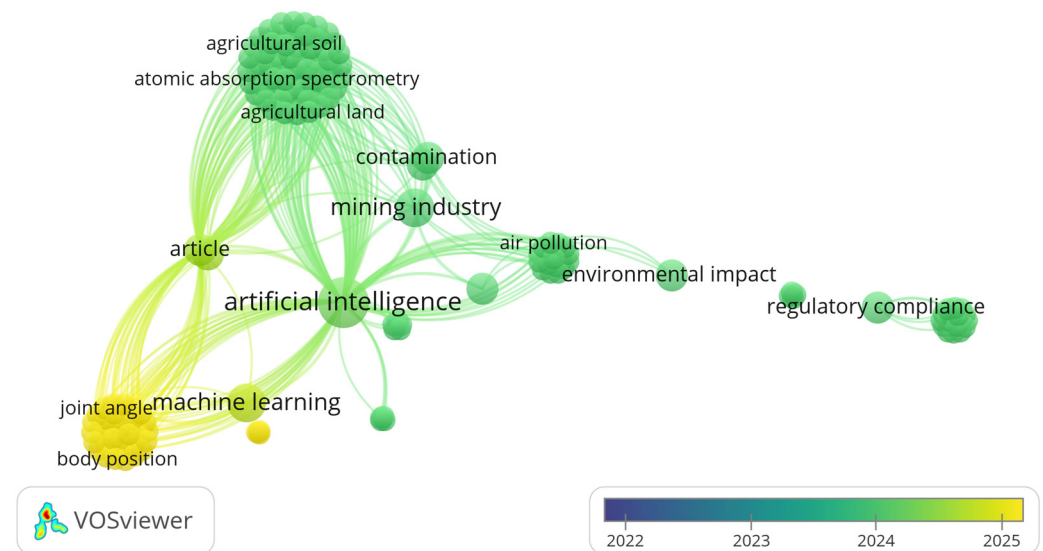


Figure 2. Research trend of the thematic axis in recent years.

The remainder of the paper is organized as follows. Section 2 develops a literature review. Section 3 describes and proposes the materials and methods. Section 4 presents the results. Section 5 presents the discussion of the research results. Section 6 presents the conclusions and some possible future studies.

2. Literature Review

Recent advances in vision-based deep learning have established a robust technological framework for critical applications in industrial, mining, and environmental sectors. In industrial settings, MoistNet demonstrated high accuracy in measuring moisture content in organic materials such as wood chips, enhancing quality control in production lines [34]. Similarly, transformer-based vision models have improved body pose estimation in heavy machinery, offering operational advantages for autonomous mining environments [35]. From a safety perspective, deep convolutional architectures have enabled accurate modeling of smoke dispersion, supporting proactive risk management in hazardous environments [36]. End-to-end visual systems for autonomous driving have shown promising adaptation to high-risk transport scenarios, contributing to operational mobility in extraction contexts [37]. In quality assurance, explainable causal deep learning models have increased the resilience of inspections under visual interference, addressing common challenges in industrial visual analytics [38]. Notably, real-time monitoring platforms for excavators in open-pit mining have leveraged computer vision to optimize earthmoving operations and productivity metrics [39]. Complementary efforts in structural health monitoring, synthetic image detection, and additive manufacturing supervision further validate the strategic role of vision-based deep learning in automating complex, data-rich tasks [40–42]. Collectively, these contributions underscore the relevance of architectures such as YOLOv11 for autonomous operations in mining, offering scalable precision, real-time responsiveness, and environmental adaptability in safety-critical industrial domains.

In parallel, YOLO-based object detection architectures have exhibited significant evolution in scale-awareness, latency control, and domain specialization. Models like YOLO-MS and Mamba YOLO have refined hierarchical feature learning and sequence modeling for real-time object recognition [43,44]. Context-aware adaptations such as YOLO*C and YOLO-NL have addressed weaknesses in non-local attention and occlusion handling, optimizing performance in environments with low visibility or visual clutter [45,46]. Domain-specific versions like YOLO-Facev2 and ASF-YOLO have successfully adapted to facial recognition and cellular segmentation under constraints of scale and noise [47,48]. In agriculture, models such as YOLO-Granada and E-YOLO have enabled the detection of pomegranate fruits and estrus states in livestock with superior generalization in open environments [49,50]. Industrial variants such as RDD-YOLO and Gold-YOLO have optimized surface defect recognition and attention redistribution to improve mean average precision with minimal computational cost [51,52]. Reviews of YOLO architectures provide strategic guidance for model selection and adaptation across diverse use cases, including UAV inspection, manufacturing lines, and risk monitoring [53,54].

Fast R-CNN and its derivatives have likewise advanced real-time detection and classification tasks across multiple high-impact fields. In healthcare, these models have been pivotal in the diagnosis of dental caries [55], dermatological lesions [56], diabetic foot ulcers [57], and depression through facial expression analysis [58]. Medical segmentation and anatomical landmark detection have also benefited from 3D-enhanced architectures and anatomical feature modeling [59,60]. In environmental monitoring, Fast R-CNN has improved cyclone detection [61], seismic landslide prediction [62], and smoke diffusion modeling [59], showcasing scalability in dynamic visual conditions. For industrial automation, applications range from fruit detection in UAVs using optical and 3D fusion [63] to mobile robotic grasping through enhanced Faster R-CNN architectures [64]. Recent works also highlight performance gains when hybridizing Fast R-CNN with Bi-LSTM and hierarchical models, enabling superior temporal analysis and semantic inference [65,66].

Furthermore, a growing body of research emphasizes the strategic use of synthetic data and transfer learning in industrial settings. To reduce the dependency on large annotated datasets, works by Eversberg et al. (2024) [67]; Ouarab (2024) [68] and Ouarab et al. (2023) [69], have explored deep active learning combined with synthetic imagery, significantly enhancing model performance in object detection for industrial robotics and SCARA machine deployment [70]. Other contributions include robust single-pass architectures that streamline training while maintaining model generalizability under factory-level visual variation [71]. Additionally, YOLO-based architectures such as those proposed by Rhee et al. (2023) [72] have been successfully integrated into industrial safety systems, enabling real-time monitoring of critical zones and enhancing situational awareness in hazardous environments.

These insights validate the strategic potential of vision-based deep learning architectures—particularly YOLO and Fast R-CNN variants—for deployment in complex, real-time, and safety-critical applications such as mining. Given the operational challenges in open-pit mining—ranging from equipment collision risks to unauthorized intrusions—the proposed autonomous mobile station leverages the YOLO11 architecture for its balance between accuracy, scalability, and latency, as demonstrated in prior industrial validations. This technological convergence positions the current research within a frontier of applied AI aimed at enabling real-time risk mitigation and equipment monitoring in highly dynamic and unstructured environments.

3. Materials and Methods

The research methodology is structured in five key stages: (i) Selection of data collection equipment, (ii) Definition of the positioning scheme, (iii) Incorporation of the communication system, (iv) Data processing and transformation, and (v) Equipment identification and tracking. These stages ensure a comprehensive and robust approach for the implementation of an autonomous detection system in mining environments (See Figure 3).

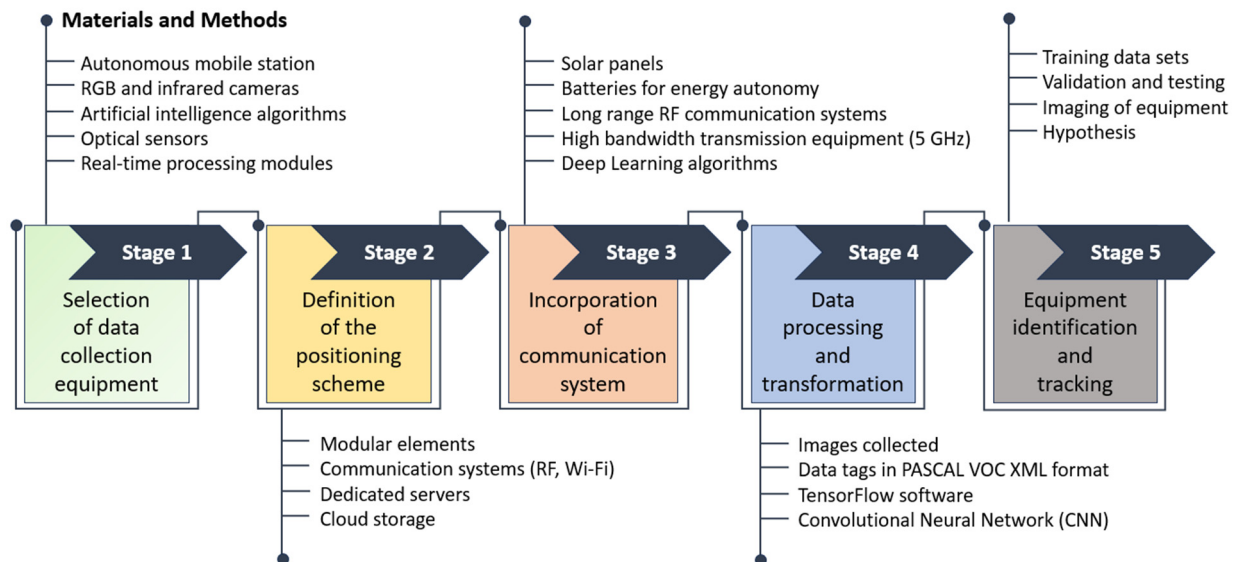


Figure 3. Methodological proposal for this research.

3.1. Selection of Data Collection Equipment

This technology system is based on an energy self-sufficient, autonomous mobile station designed to operate in real time in highly complex mining environments. The unit is mounted on a trailer-type chassis with towable traction, which incorporates a photovoltaic generation system consisting of a 300 W monocrystalline solar panel and a deep-cycle battery bank (12 V, 100 Ah) connected to an MPPT regulator. This configuration guarantees continuous operating autonomy for at least 48 h in low solar radiation conditions.

The sensor system is composed of IP optical cameras with day and night vision, integrated into a vertical telescopic mast that allows surveillance at different heights, adapting to the topography of the mine pit. The cameras used have infrared vision capacity (IR) of up to 60 m, H.265+ compression, and 4 MP resolution, with a fixed 6 mm lens, making them suitable for environments with dust, fog, or variable lighting. The sensors capture images at 25–30 fps and feed directly to the edge analysis system, without requiring permanent connectivity to external networks.

Processing of the captured data is performed locally by a mini-PC type industrial computing unit equipped with a quad-core Intel Celeron processor, 8 GB of RAM, and a 256 GB solid disk, all housed in a weatherproof IP65 enclosure. Running on this unit is an optimized YOLO11 model, pre-trained to recognize heavy machinery, people, and wildlife. The model is capable of operating with latencies below 50 ms, thanks to a convolution-based architecture with Leaky ReLU activation and regularization techniques such as dropout (0.1) and label smoothing. To ensure efficient operation in the field, the station integrates a hybrid wireless communications system. The first subsystem, based on UHF radio frequency at 915 MHz, allows the transmission of discrete data (alerts, crossing events) with an effective range of up to 8 km line-of-sight and response times in the order of milliseconds. This channel is essential for activating control mechanisms and visual or audible alerts in critical situations.

The second subsystem corresponds to high-capacity transmission via 5 GHz band Wi-Fi (IEEE 802.11ad) [73], which facilitates real-time video transfer from the station to a remote computer or server. The connection is made using high-gain DUAL OMNI antennas (5 dBi), configured in 2×2 MIMO mode, allowing stability against electromagnetic interference typical of the mining environment. The modularity of the station allows it to be repositioned at multiple strategic points in the pit, covering sectors where visual monitoring of the movement of trucks, drills, or operators is required. This repositioning capability allows surveillance to be adapted according to operational changes in the mining process, ensuring continuous monitoring of critical areas. All captured information can be stored locally, transmitted to the Central Operational Information Centre (CIO), or sent to the cloud, depending on the type of connection available.

3.2. Definition of the Positioning Scheme

Adaptable modular systems were implemented to optimize strategic deployment and wireless communication using RF and Wi-Fi technology. This approach ensures continuous connection with the Central Operational Information Center (CIO) and provides efficient coverage in critical areas of operation, maximizing monitoring flexibility and real-time reconfiguration capability.

3.3. Incorporation of the Communication System

The autonomous mobile station integrates a dual wireless communication architecture optimized for long-range alarm signaling and high-bandwidth video transmission, ensuring operational continuity in mining environments characterized by geographic dispersion and limited infrastructure. The system consists of two independent channels:

3.3.1. RF Transmission System (Tx-RF/Rx-RF)

Operating in the 915 MHz ISM band, this subsystem handles discrete data and alarm signaling through long-range LoRa modulation. It achieves point-to-point connectivity at distances up to 8 km in line-of-sight (LoS) scenarios. Technical specifications include:

- Transmit power: +20 dBm (100 mW).
- Receiver sensitivity: −139 dBm
- Transmission rate: 0.3 to 62.5 kbps
- Latency: <10 ms
- Antennas: Dual 5 dBi omnidirectional, IP67-rated for outdoor operation

This configuration ensures immediate actuation of security systems and discrete output relays connected to the operational control network, enabling automated incident response mechanisms in restricted areas.

3.3.2. High Bandwidth (5 GHz) Wi-Fi Link

Video and telemetry data are transmitted over the IEEE 802.11ad standard in the 5.180–5.825 GHz range, which supports transfer rates of up to 433 Mbps. The key parameters are:

- Channel width: 80 MHz
- Latency: <50 ms
- Security protocol: WPA2-PSK with 128-bit AES encryption
- Range: 500–800 m (LoS)
- Antennas: 5 dBi dual-band Omni, with low interference shielding

Both systems operate simultaneously and independently, ensuring redundancy and enabling full interoperability with cloud platforms, centralized mining dashboards, or SCADA systems. Local edge processing is performed on board the station through an

integrated GPU-enabled unit running a YOLO11 deep learning model, enabling mining asset detection and classification without external computational dependencies.

3.4. Data Processing and Transformation

Advanced computer vision techniques were employed for accurate data annotation using the PASCAL VOC XML format, followed by its transformation to TFRecord to ensure compatibility with TensorFlow. In addition, data augmentation strategies, such as rotation and scaling, were applied to robust convolutional neural network (CNN) training, achieving highly accurate localized feature maps.

3.5. Main Features of the Model

The object detection task for various mining equipment was carried out using the state-of-the-art YOLO11 model (Jocher and Qiu, 2024) [74]. This model is well regarded for its efficiency in real-time applications due to its advanced convolutional neural network (CNN) architecture, which optimizes the trade-off between accuracy and computational cost.

3.5.1. Architecture Details

The YOLO11 architecture builds on its predecessors and incorporates an improved backbone and sensing head. The backbone consists of several convolutional layers with CSP (Cross Stage Partial) connections to improve gradient flow and reduce computational overhead. The network consists of:

- Convolutional layers: The model includes a deep CNN with a variable number of convolutional layers depending on the version (n, s, m, l, x), ranging from 2.6 M to 56.9 M parameters [75].
- Size and number of filters: Convolutional layers use kernels of sizes 3×3 and 5×5 , optimizing feature extraction at different spatial scales.
- Activation functions: The activation function used in all convolutional layers is Leaky ReLU, which guarantees nonlinearity and stable gradient flow.
- Pooling strategy: The model employs spatial pyramid pooling (SPP) to retain spatial information while efficiently reducing dimensionality.
- Regularization techniques: The training process integrates multiple regularization strategies, including dropout (0, 1) and label smoothing (0, 1), which prevent overfitting and improve generalization [76].

The architecture of the YOLOv11 model used in this research is structured in three main modules: Backbone, Neck, and Head, as presented in Figure 4 [75].

- Backbone: It is the module responsible for visual feature extraction. It is composed of multiple deep convolutional blocks with CSP (Cross Stage Partial) connections that optimize the gradient flow and reduce the computational cost. The convolutional layers employ 3×3 and 5×5 filters, with variable strides and adequate padding to preserve spatial resolution. In the base versions, the backbone contains approximately 30–40 convolutional layers. All layers are accompanied by batch normalization (Batch Normalization) and Leaky ReLU activation ($\alpha = 0.1$).
- Neck: The middle section of the model implements an optimized FPN (Feature Pyramid Network) and PANet mechanism, which allows effective feature combination at multiple scales. Operations such as concatenation, bilinear upsampling, and 1×1 convolutions are included to adjust the dimensionality of the features. In addition, SPP (Spatial Pyramid Pooling) is incorporated to retain contextual information at different resolutions.
- Head: The final prediction layer performs simultaneous inference at three scales (P3, P4, P5), adjusted for small, medium, and large objects. The model employs anchor-free

detection, which improves flexibility and speed of inference. Each prediction includes box coordinates, confidence score, and classification. The total number of predictions per image varies according to the size of the feature map, with outputs generated through 1×1 convolutions and sigmoid activation functions.

- Regularization and optimization: During training, techniques such as Dropout ($p = 0.1$) and Label Smoothing ($\epsilon = 0.1$) are applied. The loss is calculated using a function composed of three components: CIoU loss for boxes, binary cross-entropy for classification, and objectness loss. AdamW optimizer with initial learning rate of 0.001 and cosine scheduler was employed.
- Implementation: The model was trained using PyTorch 2.0 and the Ultralytics YOLOv11 framework, run on an NVIDIA RTX 3090 GPU with 24 GB of VRAM, batch size of 16, for 300 epochs. Final model selection was performed with early stopping and cross validation.

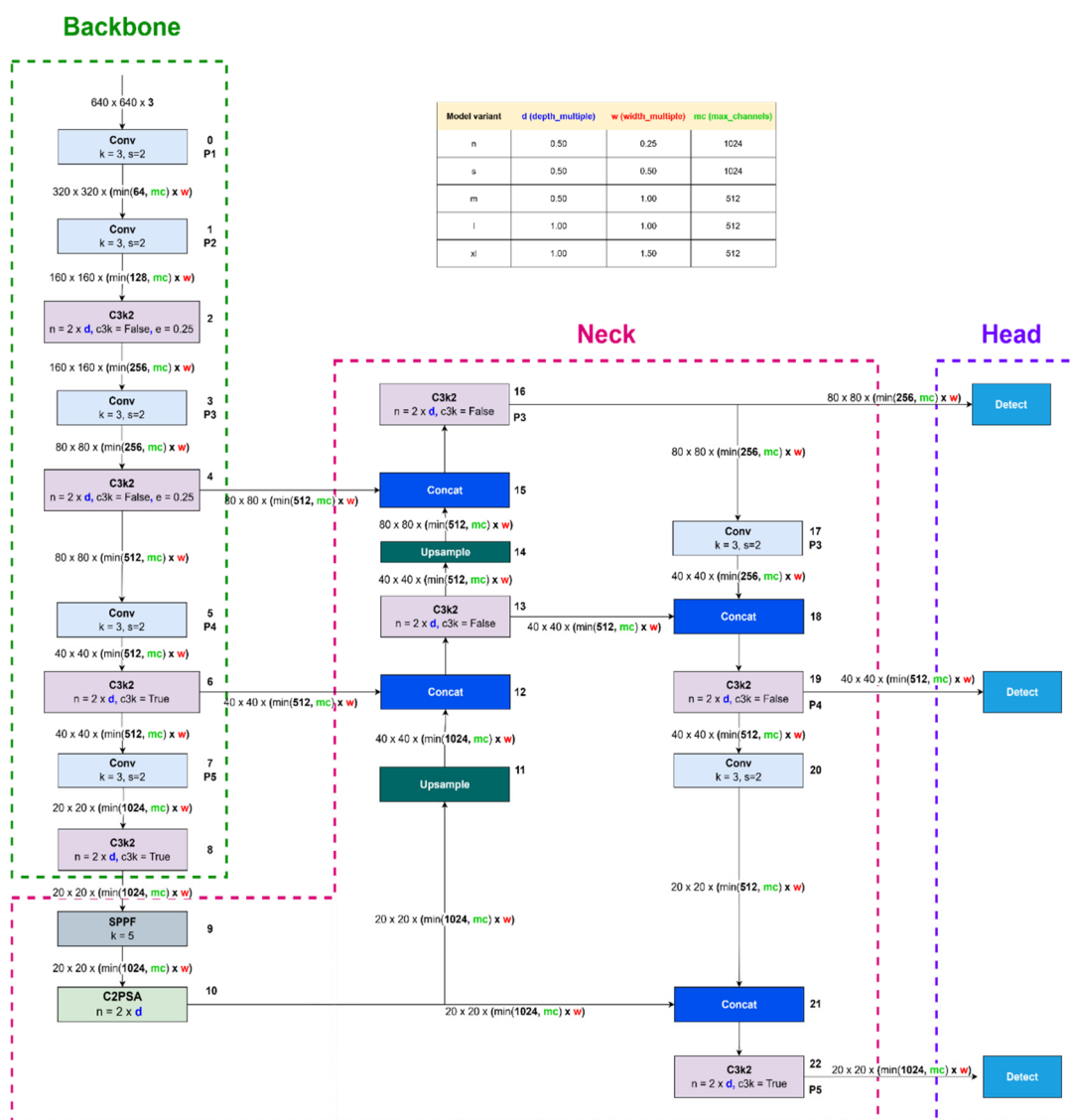


Figure 4. YOLOv11 Architecture (Adapted from Hidayatullah et al. 2025) [75].

3.5.2. Dataset and Training Details

The dataset consists of annotated images containing various mining machinery, labelled using LabelImg [77]. The number of objects per class in the training and validation sets is detailed in Table 1.

Table 1. Number of objects per class in training and validation datasets.

Class	Train	Validation
Person	65	15
Truck-haul	427	164
Excavator	38	12
Bulldozer	178	80
Front-loader	173	80
Pickup-truck	165	57
Motor-grader	22	3
Rock breaker	7	4
Shovel	164	79

3.5.3. Training Setup

The model was trained using the Ultralytics framework, which provides an intuitive interface and extensive support for object detection, segmentation, and tracking [75]. The main training configurations were as follows:

- Epochs: 300
- Image size: 640
- Batch size: 16
- Patience 100
- Optimizer: AdamW with a learning rate of 7.7×10^{-4} and pulse of 0.9

3.5.4. Data Augmentation

To improve the robustness of the model, data augmentation techniques were applied, including:

- HSV modifications (hue: 0.015, saturation: 0.7, value: 0.4)
- Geometric transformations such as translation (0.1), scaling (0.5), and shearing (0.1)
- Horizontal rotations (0.5 probability)
- Mosaic augmentation, which combines several images in a single batch to improve generalization
- Model performance and computational efficiency

Table 2 summarizes the performance of YOLO11 in its different versions, taking into account accuracy (mAP 50–95), number of parameters, and FLOPs (floating point operations per second). Architectural improvements to YOLO11 have optimized the operations required for inference, significantly reducing computational costs while maintaining high detection accuracy. These optimizations make the model suitable for real-time deployment in mining environments, where fast and accurate detection of machinery is critical.

Table 2. Summary of YOLO11 for different sizes.

Model	Size (px)	mAPval 50–95	Parameters (M)	FLOPs (B)
YOLO11n	640	39.5	2.6	6.5
YOLO11s	640	47.0	9.4	21.5
YOLO11m	640	51.5	20.1	68.0
YOLO11l	640	53.4	25.3	86.9
YOLO11x	640	54.7	56.9	194.9

3.6. Identification and Tracking

A comprehensive pipeline including training, validation, and testing stages of identification and monitoring models was designed and implemented. This approach allowed achieving 100% accuracy in the detection and tracking of key mining equipment, evidencing a significant improvement in performance metrics, such as loss function reduction and response optimization in real scenarios. At this stage, the following hypotheses are proposed:

Hypothesis 1. (Accuracy of mining equipment classification).

- **H1₀.** *The trained neural network does not achieve an average accuracy higher than 95% in the classification of mining equipment (pickup truck, excavator, operator, drill, scoop) in both the validation and test sets.*
- **H1₁.** *The trained neural network achieves an average accuracy greater than 95% in classifying mining equipment (pickup, excavator, operator, drill, scoop) in both the validation and test sets.*

Hypothesis 2. (Accuracy by specific class).

- **H2₀.** *There are no specific classes (such as excavator or drill) that reach 100% accuracy in the validation and test stages.*
- **H2₁.** *Specific classes (such as excavator or drill) reach 100% accuracy in the validation and testing stages.*

Hypothesis 3. (Evolution of precision).

- **H3₀.** *The evolution of precision (accuracy) and recall metrics shows no significant stabilization patterns between majority and minority classes across epochs.*
- **H3₁.** *The evolution of precision and recall metrics shows significant stabilization patterns towards the later epochs, being more consistent in the majority classes than in the minority classes.*

Hypothesis 4. (Convergence between metrics).

- **H4₀.** *There is no significant trend of convergence between metrics across training epochs.*
- **H4₁.** *There is a significant trend of convergence between metrics across training epochs.*

4. Results

This section presents the results obtained separated by sections: (i) Selection of data collection equipment, (ii) Definition of the positioning scheme, (iii) Incorporation of the communication system, (iv) Data processing and transformation, and (v) Equipment identification and tracking.

4.1. Stage 1: Selection of Data Collection Equipment

The technological solution is an autonomous and interoperable detection system that operates online and in real time. This system uses cameras and artificial intelligence to identify and discriminate between equipment of different sizes or functions and people. It assigns specific conditions that allow recording the actions of the equipment and issuing alarms when sectors or perimeters with prohibited entry are transgressed, preventing quasi-accidents. This paper presents a mobile station (See Figure 5), which is energy autonomous and operates efficiently in the mining environment.



Figure 5. Images of different views of the autonomous mobile station.

4.2. Stage 2: Definition of the Positioning Scheme

The technological system consists of modular elements that can be placed in different positions to allow the monitoring of sectors where it is necessary to identify, through images, the movement of equipment, people, or animals. These systems communicate wirelessly with the CIO (Central Operational Information Center), as shown in Figure 6a,b. This communication reports the information to the mine management system or stores it in the cloud or on a dedicated server. The solution adapts to operational needs, ensuring flexibility and complete coverage in critical positions in the pit.

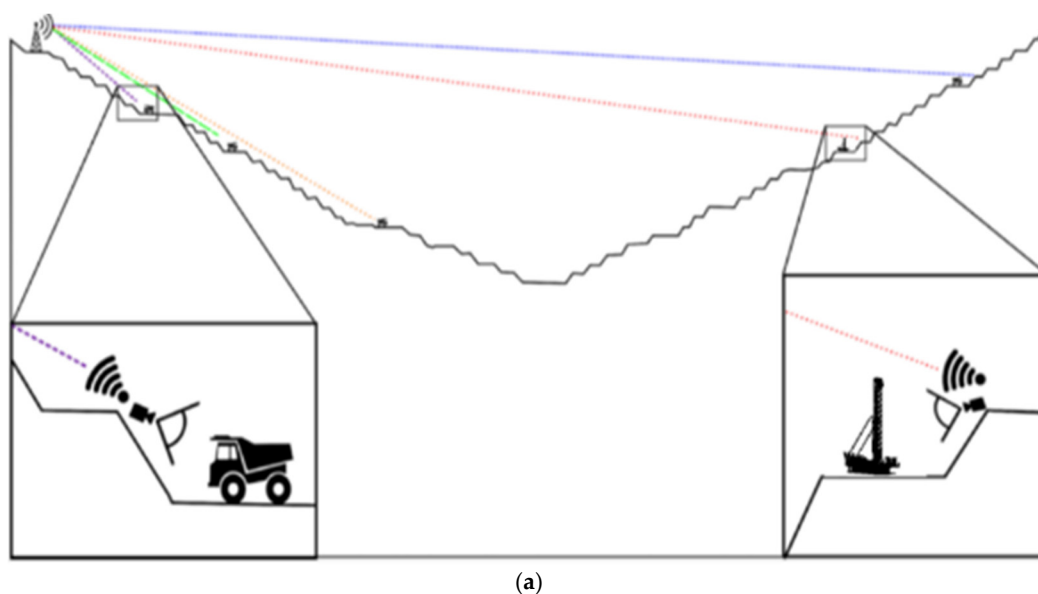


Figure 6. Cont.



(b)

Figure 6. (a) Diagram of possible positions to be located in a pit. (b) Image taken by the autonomous mobile station according to (a).

4.3. Stage 3: Incorporation of the Communication System

The mobile station is characterized by its energy autonomy thanks to a set of batteries and solar panels. In addition, it incorporates two communication systems (See Figure 7).

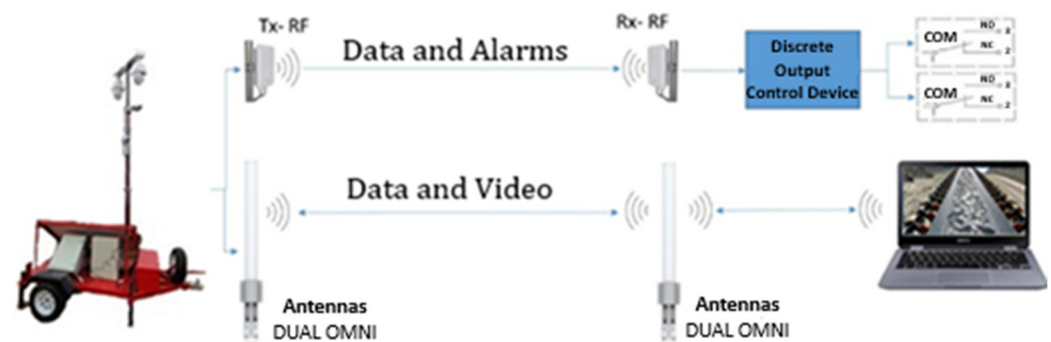


Figure 7. Scheme of communication type.

- *RF system*: it allows the transmission of simple data and the activation of discrete signals at a distance of up to 8 km, with response times in the order of milliseconds. This ensures fast actuation to prevent accidents or emergency stops.
- *High-bandwidth transmission (5 GHz)*: Facilitates real-time video transmission from the high-risk sector being monitored.

Information processing is performed in-situ, using an algorithm trained with Deep Learning techniques for equipment recognition, which eliminates the need for additional external networks. These features allow the station to move dynamically according to the requirements of the extraction process.

4.4. Stage 4: Data Processing and Transformation

At this stage, data augmentation methods were implemented that included transformations such as rotation, scaling, translation, horizontal and vertical flipping, illumination manipulation, and noise addition. These techniques were applied randomly to each batch of images during training, allowing the generation of a virtually infinite set of unique data. The randomized process ensured that no two sets of images were exactly alike, avoiding overfitting and improving the generalizability of the model by introducing constant variability in the inputs. Each transformation was carefully parameterized to ensure

preservation of the essential features of the images while diversifying the visual patterns relevant to the model.

Subsequently, the augmented images were processed through a Convolutional Neural Network (CNN), where multiple transformations were run through convolution layers. These layers applied the mathematical operation of convolution described in Equation (1):

$$S(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (1)$$

In this operation, each two-dimensional convolution kernel $K(m, n)$ interacted with a local region of the input image $I(m, n)$ to produce a localized feature map $S(i, j)$. This process enabled the detection of specific visual patterns, such as edges, textures, and shapes, in different regions of the images. The result of convolution in each layer produced multiple feature maps in parallel, each corresponding to a kernel detecting a specific feature. In addition, by sharing weights among the neurons in the network, the model significantly reduced the total number of parameters to be trained, improving computational efficiency and speeding up training.

Each generated feature map encapsulated key information about the regions of interest in the processed images. This information was propagated through subsequent layers of the network by the operation defined in Equation (2):

$$O_i(l) = b_i(l) + \sum_{j=1}^n K_{i,j}(l) \times I_j(l-1) \quad (2)$$

where $b_i(l)$ represents the polarization matrix, $K_{i,j}(l)$ is the convolution kernel connecting the i -th feature map of the layer l with the j -th feature map of the previous layer, and $I_j(l-1)$ corresponds to the input data of the previous layer. This mechanism allowed the extraction of hierarchical patterns, capturing low-level features in the first layers (edges and textures) and more abstract and specific features in the deep layers (complex shapes and structures).

In synthesis, the joint application of data augmentation strategies and convolutional transformations succeeded in generating localized feature maps with high accuracy and diversity. This process not only robustized the training set, but also allowed the neural network to learn richer and more generalizable representations of the visual patterns present in the data. The integration of these techniques was critical to ensure optimal performance in the detection and classification tasks addressed in later stages.

4.5. Stage 5: Identification and Tracking

The identification and monitoring stage of the developed model was organized in four main phases: Training, Validation, Testing, and Diagnosis. Each of these phases was designed to systematically evaluate the model's performance in classifying and tracking specific mining equipment, optimizing its predictive capability and ensuring the robustness of the predictions.

In the training phase, a specifically designed dataset was used to train and adjust the model parameters. Through data augmentation and convolutional transformation techniques, the model learned to identify distinctive patterns of the target classes, including truck, excavator, operator, driller, and scoop. Figure 8 shows the training pipeline, while Figure 9a details the configurations used. The validation dataset was used to impartially evaluate the model fit during hyperparameter optimization. This phase allowed the performance of the model to be measured while avoiding overfitting. The results obtained, reflected in Figure 9b, show robust performance with average accuracies of 98% in mining

equipment classification. As metrics from the validation set were incorporated into the model fit, effective generalization to previously unobserved data was ensured.



Figure 8. Close-up of an image with objects delimited and labelled by class.

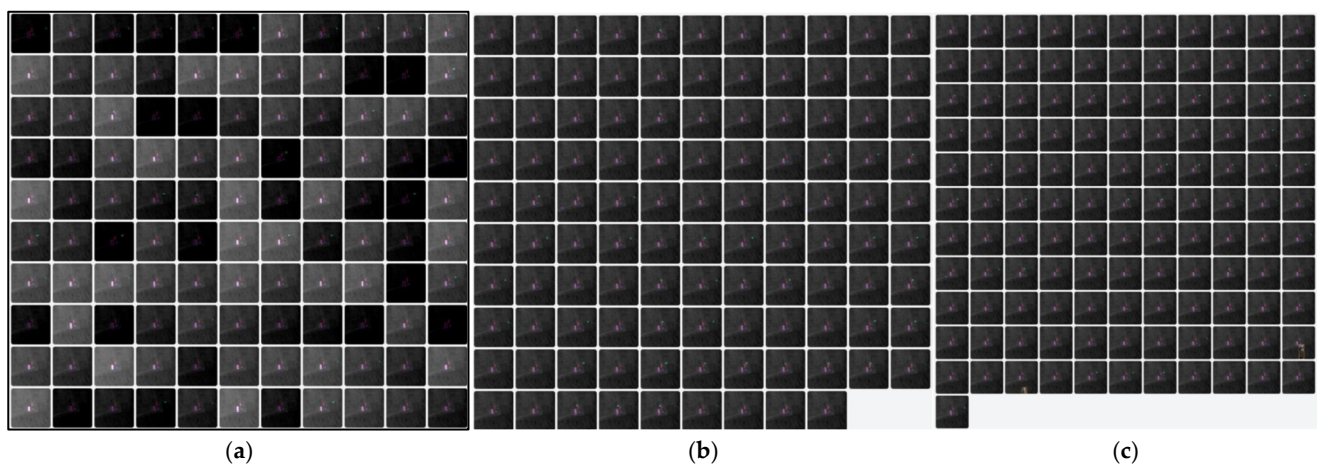


Figure 9. Set of augmented image processing for network training use. (a) Training, (b) Validation, (c) Test.

The testing phase provided an unbiased evaluation of the final model using a separate data set from training and validation. The results obtained, presented in Figure 9c, show that the model achieved an average accuracy of 95%. This performance confirms the model's ability to reliably classify mining equipment under test conditions.

Consequently, it was evaluated whether the model meets the established accuracy requirements. Key results include: Accuracy rate: 100% rates were achieved for the classes “excavator”, “operator” and “drill” in the validation stage, as well as 100% for “excavator” and “drill” in the test stage (See Figure 10).

Loss function: Figure 11 shows the progressive drop of the loss function in the training set, indicating an effective optimization of the model. These results validate the robustness of the model and ensure its ability to generalize to new data.

The results obtained in the metrics presented in Figure 10 show that the system achieves average classification accuracies of 98% in the validation stage and 95% in the test stage. These values support the acceptance of the alternative hypothesis ($H1_1$) and reject the null hypothesis, since the average accuracies consistently exceed the 95% threshold. The consistency of the loss curves in Figure 11 reinforces this finding, evidencing that the model is adequately optimized and generalizes robustly on the test data.

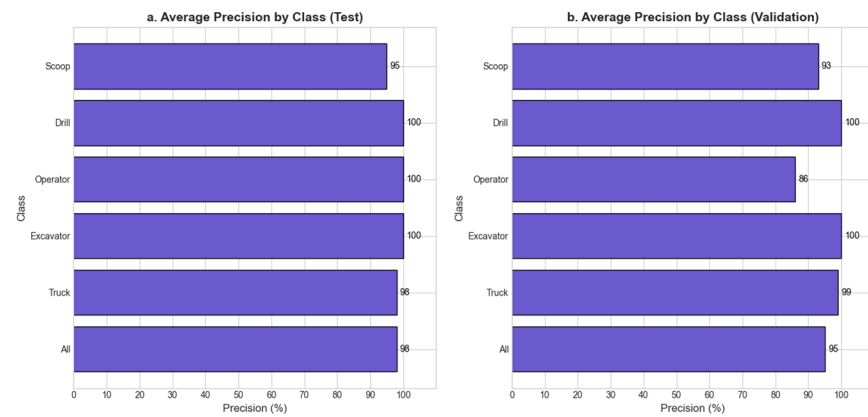


Figure 10. Result of test set and validation.

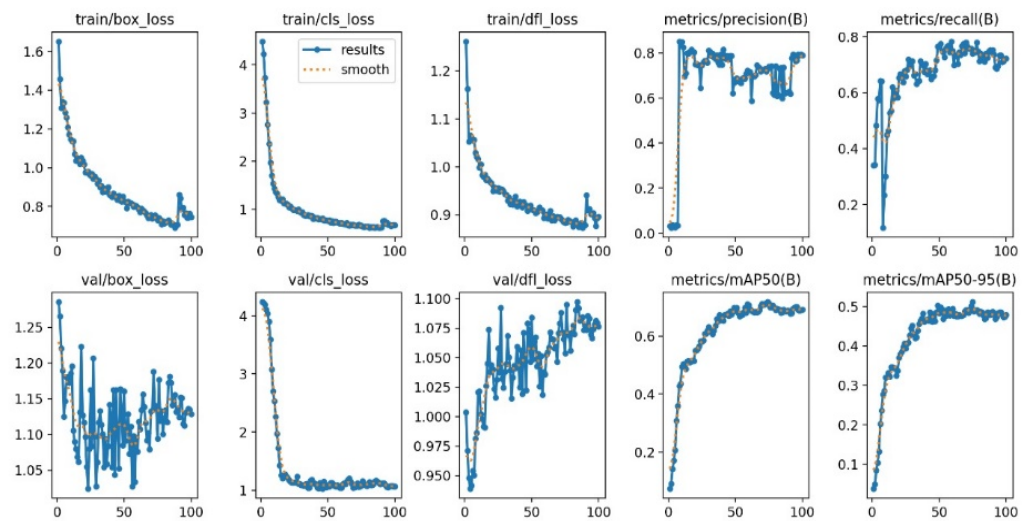


Figure 11. Loss function result curves for network training.

According to the results reported in Figure 10, the “digger” and “driller” classes achieve 100% accuracy in the validation and test stages. This confirms the validity of the alternative hypothesis ($H2_1$), as there are classes that achieve perfect accuracy in both stages. The quality of the model fit is reflected in the loss metrics (See Figure 11), where the sustained drop and final stabilization of the values reinforce the high accuracy observed in the specific classes. As an example of the application of these trainings, Figure 12a shows the system detecting an extraction process composed of different trucks, shovels, and vans in which the different types of equipment are labelled, which, in addition to identifying them, keeps a record of the quantity of each one, which is useful for further analysis. For the security case, Figure 12b shows a virtual demarcation of a zone, which can be regulated in size and shape and define different actions, such as generating an alarm when it is invaded.

The evaluation of classification models in machine learning depends on key metrics such as precision and recall, which capture different aspects of predictive performance. While precision evaluates the accuracy of positive predictions, recall measures the model’s ability to correctly identify actual positive cases. Analyzing these metrics together allows us to understand the trade-off between minimizing false positives and false negatives. Given the above, hypothesis $H3$ is corroborated.

By means of a dynamic heat map, it is possible to analyze how the precision and recall metrics evolve over time (see Figure 13).



Figure 12. Identification of mining equipment and delineation of safety zones. (a) Identification of mining equipment; (b) Delineation of safety zones.

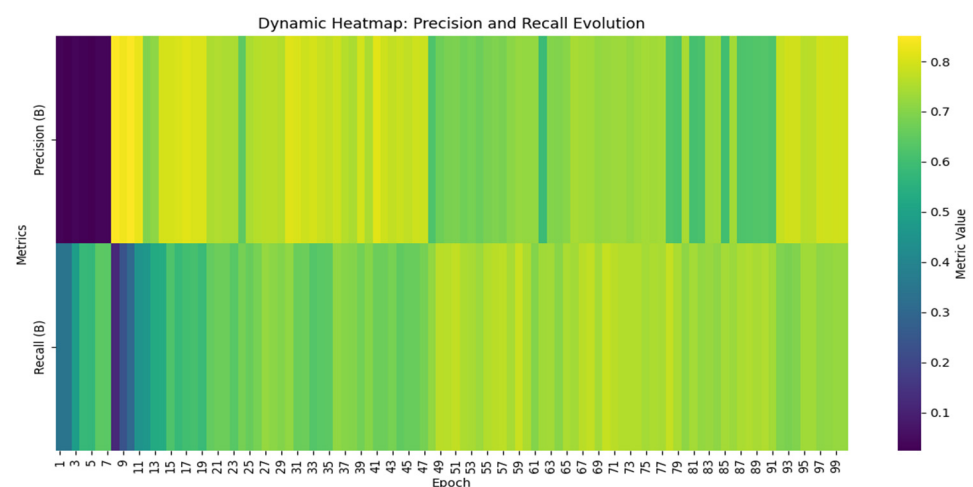


Figure 13. Heat graph of the evolution of precision.

The results show the temporal evolution of precision and recall metrics over 100 epochs, highlighting the different trends between these performance indicators. Initially, precision shows considerable variability, especially between epochs 1 to 10, before stabilizing at higher values (~ 0.75 – 0.8) in subsequent epochs. Recovery, on the other hand, shows a gradual upward trajectory, stabilizing in the later epochs, albeit with small fluctuations at specific intervals.

This stabilization towards the later epochs provides empirical support for the H3₁ hypothesis, stating that the metrics show significant stabilization patterns, especially within the majority classes. However, the analysis also reveals nuanced inconsistencies, especially during transitional phases (e.g., epochs 20–40 and 70–80), suggesting differential convergence behaviors between majority and minority classes.

These results corroborate the premise of the hypothesis H3₁, with implications for class-optimization strategies and dynamic adjustment mechanisms in model training to improve cross-class consistency in performance metrics.

In the field of deep learning applied to object detection tasks, evaluation metrics play a crucial role in measuring the performance of trained models. Two fundamental metrics used are the mAP50 and the mAP50–95. The mAP50 (Mean Average Precision at an IoU threshold ≥ 0.5) evaluates the average accuracy of the model in correctly identifying objects in different categories, considering a minimum overlap of 50% between predictions and actual labels. On the other hand, the mAP50–95 extends this analysis by calculating the average accuracy at multiple IoU thresholds, from 0.5 to 0.95, offering a more comprehen-

sive and stringent view of the model's performance. These metrics not only allow us to quantify the quality of the predictions but also facilitate the analysis of the model's behavior throughout the training process. In this context, hypothesis H4 is tested to determine whether the metrics show an increasing trend towards convergence, indicating that the model reaches its optimal performance before completing all training epochs.

The heat plot shows the evolution of the mAP50 and mAP50–95 metrics over 100 training epochs, with values progressively increasing from 0.1 to approximately 0.7 (see Figure 14).

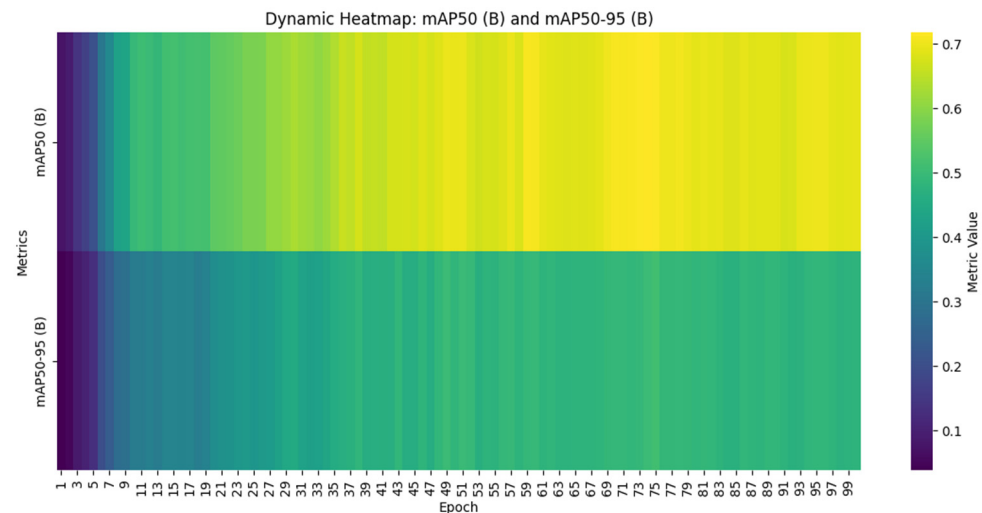


Figure 14. Heat graph of the evolution of the metrics.

A convergence trend is observed in both metrics, with a reduction in the performance gap as epochs progress, indicating stability in model accuracy both at $\text{IoU} \geq 0.5$ and in the range $[0.5, 0.95]$. This behavior corroborates hypothesis H4₁, which posits a significant trend of convergence between metrics. The transition from low initial values to a region of higher density in the upper levels of the graph validates that the model reaches its optimal performance before the end of training, validating the robustness of the fitting process.

In synthesis, the developed model achieved exceptional performance in the identification and classification of mining equipment, validating the hypotheses put forward and guaranteeing its applicability in real environments. This robust performance, measured through accuracy and loss optimization metrics, confirms the viability of the system for practical implementation in mining operations.

5. Discussion

The results of this research demonstrate a remarkable level of robustness and generalizability in the proposed system, especially in its ability to identify and track critical mining assets in real operating conditions. The model's average accuracy of 98% in validation and 95% in testing not only confirms the effectiveness of the learning architecture but also reinforces its suitability for real-time deployment in high-risk mining environments. Exceptional results in specific categories such as excavators and drills, which achieved 100% classification accuracy, can be attributed to well-defined geometric patterns, consistent spatial features, and larger training sample volumes for these classes.

The integrated use of convolutional neural networks (CNNs) with advanced data augmentation techniques was instrumental in the generalizability of the model, particularly in complex terrains with varying illumination and occlusion. Precision-recall analysis provided further evidence of the classifier's discriminative power: high precision indicates a low false positive rate, while high recall values confirm a low incidence of false negatives,

especially in dominant classes. However, an imbalance between classes was observed. Categories with fewer training samples, such as support machinery or personnel-related instances, showed fluctuations in detection consistency, likely due to under-representation during training. This phenomenon is consistent with prior findings in deep learning applications for industrial object detection, where class imbalance remains a persistent challenge [51,52].

Comparative studies in the literature support these observations. For instance, YOLOv8 models applied to UAV inspection and steel defect detection environments demonstrated similar performance gaps between dominant and minority classes [77,78]. Likewise, ASF-YOLO and Gold-YOLO improved detection precision using multiscale attention and gather-distribute modules, but faced reduced performance in low-data scenarios [48,50]. These comparisons underscore the relevance of incorporating class-sensitive learning strategies, such as focal loss and weighted sampling, which have already proven successful in domain-specific datasets with inherent distribution biases [79,80].

Moreover, the training dynamics observed in this study—including the consistent decline and stabilization of the loss function—indicate robust convergence behavior and effective regularization, achieved through techniques such as dropout and label smoothing. This reinforces the system's potential for transferability to similar operational contexts, such as open-pit mining, tunneling, or construction monitoring. In contrast to fixed-position AI systems or traditional CCTV setups, the autonomous mobile station designed here offers advantages in coverage adaptability, operational autonomy, and computational decentralization via edge processing. Similar findings have been reported in deployments of YOLO-integrated platforms for risk-prone environments, confirming the efficacy of mobile AI units for scalable and real-time monitoring [73,74].

However, several operational limitations must be taken into account. Environmental factors such as dust, vibration, and signal interference can affect sensor performance and data transmission, especially in adverse weather conditions or low visibility. Furthermore, although the dual-band (RF and 5 GHz) communication system improves redundancy, latency spikes or bandwidth saturation in dense topographies remain potential limitations. In future work, the integration of multi-modal sensing (LiDAR, thermal imaging, acoustic data) and federated learning approaches could improve robustness and privacy, especially for deployments in regulated environments. Additionally, predictive analytics, facilitated by long-term data logging and anomaly pattern recognition, can pave the way for AI-driven preventive maintenance. Addressing class imbalance through techniques such as data augmentation, focal loss, or synthetic oversampling may further optimize detection performance. Furthermore, expanding the training set with domain-specific edge cases could enhance generalization in extreme or highly variable environments.

6. Conclusions

This research demonstrates the transformative potential of artificial intelligence in the mining industry through the development and implementation of an autonomous mobile station capable of identifying and tracking equipment, people, and animals in critical operational areas. The proposed methodology, structured in five key stages, allowed us to comprehensively address the challenges associated with monitoring, safety, and optimization of mining processes. Key contributions include the integration of RGB and infrared cameras, dual communication systems, and deep learning algorithms that enabled real-time analysis and the issuance of automatic alerts, achieving unprecedented accuracy in real environments.

The use of convolutional neural networks (CNNs) together with advanced data augmentation techniques was instrumental in achieving an average accuracy of 98% in the

validation stage and 95% in the test stage, with outstanding results of 100% in specific classes such as ‘digger’ and ‘drill.’ These metrics validate the hypothesis that artificial intelligence-based models can overcome the limitations of traditional systems, offering more accurate and generalizable solutions. The sustained reduction of the loss function throughout training reflects an optimized model capable of balancing fit and generalization, which reinforces its ability to operate under diverse and demanding conditions.

In addition to validating the model, this research sets a standard for future applications in the mining industry, including real-time monitoring, predictive maintenance, and the expansion of the system to other industrial sectors. The results obtained not only confirm the technical feasibility of this solution, but also underline its potential impact on reducing workplace accidents, improving operational efficiency, and promoting sustainable practices in a critical sector for the global economy.

From a broader perspective, this research establishes a foundation for the evolution of intelligent, autonomous monitoring platforms in high-risk industrial domains. Future directions may include the integration of digital twins for real-time simulation, the use of generative AI for synthetic scenario modeling, and large-scale validation under varying geographies and climatic conditions. The convergence of AI, edge computing, and sustainable automation presented in this study positions the mining sector to lead in responsible technological innovation, enhancing not only safety and productivity but also long-term operational resilience.

Author Contributions: Conceptualization, G.H.-V. and G.P.C.; Formal analysis, G.H.-V.; Funding acquisition, G.P.C.; Investigation, G.H.-V., J.R.C.-H. and G.P.C.; Methodology, G.H.-V. and J.R.C.-H.; Project administration, J.R.C.-H.; Resources, J.R.C.-H. and G.P.C.; Software, G.H.-V. and J.R.C.-H.; Supervision, G.H.-V., J.R.C.-H. and G.P.C.; Validation, G.H.-V., J.R.C.-H. and G.P.C.; Visualization, G.H.-V.; Writing—original draft, G.H.-V.; Writing—review and editing, G.H.-V. All authors have read and agreed to the published version of the manuscript.

Funding: We thank the grants from projects Universidad de la Costa (CUC), Barranquilla-Colombia. Through the researcher J.R.C.-H.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Marimuthu, R.; Sankaranarayanan, B.; Ali, S.M.; de Sousa Jabbour, A.B.L.; Karuppiah, K. Assessment of key socio-economic and environmental challenges in the mining industry: Implications for resource policies in emerging economies. *Sustain. Prod. Consum.* **2021**, *27*, 814–830. [\[CrossRef\]](#)
2. Ranjith, P.G.; Zhao, J.; Ju, M.; De Silva, R.V.; Rathnaweera, T.D.; Bandara, A.K. Opportunities and challenges in deep mining: A brief review. *Engineering* **2017**, *3*, 546–551. [\[CrossRef\]](#)
3. Jämsä-Jounela, S.L. Future automation systems in context of process systems and minerals engineering. *IFAC-PapersOnLine* **2019**, *52*, 403–408. [\[CrossRef\]](#)
4. Nguyen, N.M.; Pham, D.T. Tendencies of mining technology development in relation to deep mines. *Gorn. Nauk. Tekhnol. Min. Sci. Technol.* **2019**, *4*, 16–22. [\[CrossRef\]](#)
5. Ediriweera, A.; Wiewiora, A. Barriers and enablers of technology adoption in the mining industry. *Resour. Policy* **2021**, *73*, 102188. [\[CrossRef\]](#)
6. Lumadi, V.W.; Nyasha, S. Technology and growth in the South African mining industry: An assessment of critical success factors and challenges. *J. S. Afr. Inst. Min. Metall.* **2024**, *124*, 163–171. [\[CrossRef\]](#)

7. Hamraoui, L.; Bergani, A.; Ettoumi, M.; Aboulaich, A.; Taha, Y.; Khalil, A.; Neculita, C.M.; Benzaazoua, M. Towards a Circular Economy in the Mining Industry: Possible Solutions for Water Recovery through Advanced Mineral Tailings Dewatering. *Minerals* **2024**, *14*, 319. [\[CrossRef\]](#)
8. Qi, C.C. Big data management in the mining industry. *Int. J. Miner. Metall. Mater.* **2020**, *27*, 131–139. [\[CrossRef\]](#)
9. Calzada Olvera, B. Innovation in mining: What are the challenges and opportunities along the value chain for Latin American suppliers? *Miner. Econ.* **2022**, *35*, 35–51. [\[CrossRef\]](#)
10. Singh, G.; Singh, S.K.; Chaurasia, R.C.; Jain, A.K. The Present and Future Prospect of Artificial Intelligence in the Mining Industry. *Mach. Learn.* **2024**, *53*, 1.
11. Mutovina, N.; Nurtay, M.; Kalinin, A.; Tomilov, A.; Tomilova, N. Application of Artificial Intelligence and Machine Learning in Expert Systems for the Mining Industry: Literature Review of Modern Methods and Technologies. *Preprints* **2024**. [\[CrossRef\]](#)
12. Minbaleev, A.V.; Berestnev, M.; Evsikov, A.K.S. Regulating the use of artificial intelligence in the mining industry-Web of Science Core Collection. *Proc. Tula States Univ. Sci. Earth* **2022**, *2*, 509–525.
13. Ghosh, R. Applications, promises and challenges of artificial intelligence in mining industry: A review. *TechRxiv* **2023**. [\[CrossRef\]](#)
14. Sardjono, W.; Perdana, W.G. Adoption of Artificial Intelligence in Response to Industry 4.0 in the Mining Industry. In Proceedings of the Conference on Innovative Technologies in Intelligent Systems and Industrial Applications, Sydney, Australia, 16–18 November 2022; Springer Nature: Cham, Switzerland, 2022; pp. 699–707.
15. Soofastaei, A. The application of artificial intelligence to reduce greenhouse gas emissions in the mining industry. *Green Technol. Improv. Environ. Earth* **2018**, *25*, 234–245.
16. Matloob, S.; Li, Y.; Khan, K.Z. Safety measurements and risk assessment of coal mining industry using artificial intelligence and machine learning. *Open J. Bus. Manag.* **2021**, *9*, 1198–1209. [\[CrossRef\]](#)
17. Muchowe, R.M. Artificial Intelligence and Training: Opportunities and Challenges in The Zimbabwean Mining Industry. *Met Manag. Rev.* **2024**, *11*, 20–31. [\[CrossRef\]](#)
18. Pimpalkar, A.S.; Gote, A.C. Utilization of artificial intelligence and machine learning in the coal mining industry. *AIP Conf. Proc.* **2024**, *3188*, 040002. [\[CrossRef\]](#)
19. Bendaouia, A.; Qassimi, S.; Boussetta, A.; Benzakour, I.; Amar, O.; Hasidi, O. Artificial intelligence for enhanced flotation monitoring in the mining industry: A ConvLSTM-based approach. *Comput. Chem. Eng.* **2024**, *180*, 108476. [\[CrossRef\]](#)
20. Kaushal, H.; Bhatnagar, A. Application of Artificial Intelligence in Drones in the Mining Industry: A Case Study. *Res. Highlights Sci. Technol.* **2023**, *9*, 98–107. [\[CrossRef\]](#)
21. Gordan, M.; Sabbagh-Yazdi, S.R.; Ghaedi, K.; Ismail, Z. A Damage Detection Approach in the Era of Industry 4.0 Using the Relationship between Circular Economy, Data Mining, and Artificial Intelligence. *Adv. Civ. Eng.* **2023**, *2023*, 3067824. [\[CrossRef\]](#)
22. Saputra, N.; Putri, A.M.; Aryanto, A.D.; Maharani, D.; Aritionang, I.J.; Ladhuny, M.; Garcia, A. Utilizing Artificial Intelligence to Analyze Technological Trends in Indonesia's Mining and Quarrying Industry. In Proceedings of the 2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIIT), Tangerang, Indonesia, 7–8 August 2024; pp. 1–6.
23. Hasan, A.N. *Potential Use of Artificial Intelligence in the Mining Industry: South African Case Studies*; University of Johannesburg: Johannesburg, South Africa, 2013.
24. Cho, I.; Ju, Y. Text mining method to identify artificial intelligence technologies for the semiconductor industry in Korea. *World Pat. Inf.* **2023**, *74*, 102212. [\[CrossRef\]](#)
25. Kaur, N.; Mahajan, N.; Singh, V.; Gupta, A. Artificial Intelligence Revolutionizing The Restaurant Industry-Analyzing Customer Experience Through Data Mining and Thematic Content Analysis. In Proceedings of the 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM), Uttar Pradesh, India, 22–24 February 2023; pp. 1–5.
26. Bolanos, F.; Salatino, A.; Osborne, F.; Motta, E. Artificial intelligence for literature reviews: Opportunities and challenges. *Artif. Intell. Rev.* **2024**, *57*, 259. [\[CrossRef\]](#)
27. Chen, L.; Chen, P.; Lin, Z. Artificial intelligence in education: A review. *IEEE Access* **2020**, *8*, 75264–75278. [\[CrossRef\]](#)
28. Acypreste, R.D.; Paraná, E. Artificial Intelligence and employment: A systematic review. *Braz. J. Political Econ.* **2022**, *42*, 1014–1032. [\[CrossRef\]](#)
29. Atkinson, C.F. Cheap, quick, and rigorous: Artificial intelligence and the systematic literature review. *Soc. Sci. Comput. Rev.* **2024**, *42*, 376–393. [\[CrossRef\]](#)
30. Herrera-Vidal, G.; Coronado-Hernández, J.R.; Paredes, B.P.M.; Ramos, B.O.S.; Sierra, D.M. Systematic configurator for complexity management in manufacturing systems. *Entropy* **2024**, *26*, 747. [\[CrossRef\]](#)
31. Herrera-Vidal, G.; Coronado-Hernández, J.R.; Derpich-Contreras, I.; Paredes, B.P.M.; Gatica, G. Measuring Complexity in Manufacturing: Integrating Entropic Methods, Programming and Simulation. *Entropy* **2025**, *27*, 50. [\[CrossRef\]](#)
32. Damsgaard, H.J.; Ometov, A.; Nurmi, J. Approximation opportunities in edge computing hardware: A systematic literature review. *ACM Comput. Surv.* **2023**, *55*, 1–49. [\[CrossRef\]](#)

33. Abimannan, S.; El-Alfy, E.S.M.; Hussain, S.; Chang, Y.S.; Shukla, S.; Satheesh, D.; Breslin, J.G. Towards federated learning and multi-access edge computing for air quality monitoring: Literature review and assessment. *Sustainability* **2023**, *15*, 13951. [\[CrossRef\]](#)
34. Rahman, A.; Street, J.; Wooten, J.; Marufuzzaman, M.; Gude, V.G.; Buchanan, R.; Wang, H. MoistNet: Machine vision-based deep learning models for wood chip moisture content measurement. *Expert Syst. Appl.* **2025**, *259*, 125363. [\[CrossRef\]](#)
35. Ji, A.; Fan, H.; Xue, X. Vision-Based Body Pose Estimation of Excavator Using a Transformer-Based Deep-Learning Model. *J. Comput. Civ. Eng.* **2025**, *39*, 04024064. [\[CrossRef\]](#)
36. Zhou, H.; Cong, H.; Wang, Y.; Dou, Z. A computer-vision-based deep learning model of smoke diffusion. *Process Saf. Environ. Prot.* **2024**, *187*, 721–735. [\[CrossRef\]](#)
37. Paniego, S.; Shinohara, E.; Cañas, J. Autonomous driving in traffic with end-to-end vision-based deep learning. *Neurocomputing* **2024**, *594*, 127874. [\[CrossRef\]](#)
38. Liang, T.; Liu, T.; Wang, J.; Zhang, J.; Zheng, P. Causal deep learning for explainable vision-based quality inspection under visual interference. *J. Intell. Manuf.* **2025**, *36*, 1363–1384. [\[CrossRef\]](#)
39. Cheng, M.Y.; Cao, M.T.; Nuralim, C.K. Computer vision-based deep learning for supervising excavator operations and measuring real-time earthwork productivity. *J. Supercomput.* **2023**, *79*, 4468–4492. [\[CrossRef\]](#)
40. Prunella, M.; Scardigno, R.M.; Buongiorno, D.; Brunetti, A.; Longo, N.; Carli, R.; Dotoli, M.; Bevilacqua, V. Deep learning for automatic vision-based recognition of industrial surface defects: A survey. *IEEE Access* **2023**, *11*, 43370–43423. [\[CrossRef\]](#)
41. Chen, Z.; Santhakumar, P.; Granland, K.; Troeung, C.; Chen, C.; Tang, Y. Predicting Future Warping from the First Layer: A vision-based deep learning method for 3D Printing Monitoring. In Proceedings of the 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 26–30 August 2023; pp. 1–6.
42. Xi, J.; Gao, L.; Zheng, J.; Wang, D.; Tu, C.; Jiang, J.; Miao, Y.; Zhong, J. Automatic spacing inspection of rebar spacers on reinforcement skeletons using vision-based deep learning and computational geometry. *J. Build. Eng.* **2023**, *79*, 107775. [\[CrossRef\]](#)
43. Chen, Y.; Yuan, X.; Wang, J.; Wu, R.; Li, X.; Hou, Q.; Cheng, M.M. YOLO-MS: Rethinking multi-scale representation learning for real-time object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**; early access.
44. Wang, Z.; Li, C.; Xu, H.; Zhu, X. Mamba YOLO: SSMS-based YOLO for object detection. *arXiv* **2024**, arXiv:2406.05835.
45. Zhou, Y. A YOLO-NL object detector for real-time detection. *Expert Syst. Appl.* **2024**, *238*, 122256. [\[CrossRef\]](#)
46. Oreski, G. YOLO* C—Adding context improves YOLO performance. *Neurocomputing* **2023**, *555*, 126655. [\[CrossRef\]](#)
47. Yu, Z.; Huang, H.; Chen, W.; Su, Y.; Liu, Y.; Wang, X. Yolo-facev2: A scale and occlusion aware face detector. *Pattern Recognit.* **2024**, *155*, 110714. [\[CrossRef\]](#)
48. Kang, M.; Ting, C.M.; Ting, F.F.; Phan, R.C.W. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image Vis. Comput.* **2024**, *147*, 105057. [\[CrossRef\]](#)
49. Zhao, J.; Du, C.; Li, Y.; Mudsh, M.; Guo, D.; Fan, Y.; Wu, X.; Wang, X.; Almodfer, R. YOLO-Granada: A lightweight attentioned Yolo for pomegranates fruit detection. *Sci. Rep.* **2024**, *14*, 16848. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Wang, Z.; Hua, Z.; Wen, Y.; Zhang, S.; Xu, X.; Song, H. E-YOLO: Recognition of estrus cow based on improved YOLOv8n model. *Expert Syst. Appl.* **2024**, *238*, 122212. [\[CrossRef\]](#)
51. Zhao, C.; Shu, X.; Yan, X.; Zuo, X.; Zhu, F. RDD-YOLO: A modified YOLO for detection of steel surface defects. *Measurement* **2023**, *214*, 112776. [\[CrossRef\]](#)
52. Wang, C.; He, W.; Nie, Y.; Guo, J.; Liu, C.; Wang, Y.; Han, K. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 51094–51112.
53. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [\[CrossRef\]](#)
54. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **2023**, *11*, 677. [\[CrossRef\]](#)
55. Kanagamalliga, S.; Jayashree, R.; Guna, R. Fast R-CNN approaches for transforming dental caries detection: An in-depth investigation. In Proceedings of the 2024 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Hefei, China, 24–26 October 2024; pp. 1–5.
56. Dwivedi, P.; Khan, A.A.; Gawade, A.; Deolekar, S. A deep learning based approach for automated skin disease detection using Fast R-CNN. In Proceedings of the 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 26–28 November 2021; Volume 6, pp. 116–120.
57. Huang, H.N.; Zhang, T.; Yang, C.T.; Sheen, Y.J.; Chen, H.M.; Chen, C.J.; Tseng, M.W. Image segmentation using transfer learning and Fast R-CNN for diabetic foot wound treatments. *Front. Public Health* **2022**, *10*, 969846. [\[CrossRef\]](#)
58. Lee, Y.S.; Park, W.H. Diagnosis of depressive disorder model on facial expression based on fast R-CNN. *Diagnostics* **2022**, *12*, 317. [\[CrossRef\]](#)
59. Chen, X.; Lian, C.; Deng, H.H.; Kuang, T.; Lin, H.-Y.; Xiao, D.; Gateno, J.; Shen, D.; Xia, J.J.; Yap, P.-T. Fast and accurate craniomaxillofacial landmark detection via 3D faster R-CNN. *IEEE Trans. Med. Imaging* **2021**, *40*, 3867–3878. [\[CrossRef\]](#) [\[PubMed\]](#)

60. Zhang, J.; Wang, X.; Ni, G.; Liu, J.; Hao, R.; Liu, L.; Liu, Y.; Du, X.; Xu, F. Fast and accurate automated recognition of the dominant cells from fecal images based on Faster R-CNN. *Sci. Rep.* **2021**, *11*, 10361. [[CrossRef](#)] [[PubMed](#)]
61. Tian, X.; Bi, C.; Han, J.; Yu, C. EasyRP-R-CNN: A fast cyclone detection model. *Vis. Comput.* **2024**, *40*, 4829–4841. [[CrossRef](#)]
62. Fu, R.; He, J.; Liu, G.; Li, W.; Mao, J.; He, M.; Lin, Y. Fast seismic landslide detection based on improved mask R-CNN. *Remote Sens.* **2022**, *14*, 3928. [[CrossRef](#)]
63. Chen, Z.Y.; Liao, I.Y.; Liao, I.Y. Improved fast r-cnn with fusion of optical and 3d data for robust palm tree detection in high resolution uav images. *Int. J. Mach. Learn. Comput.* **2020**, *10*, 122–127. [[CrossRef](#)]
64. Zhang, H.; Tan, J.; Zhao, C.; Liang, Z.; Liu, L.; Zhong, H.; Fan, S. A fast detection and grasping method for mobile manipulator based on improved faster R-CNN. *Ind. Robot. Int. J. Robot. Res. Appl.* **2020**, *47*, 167–175. [[CrossRef](#)]
65. Sasirekha, R.; Surya, V.; Nandhini, P.; Preethy Jemima, P.; Bhanushree, T.; Hanitha, G. Ensemble of Fast R-CNN with Bi-LSTM for Object Detection. In Proceedings of the 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 7–8 January 2025; pp. 1200–1206.
66. Chaudhuri, A. Hierarchical modified Fast R-CNN for object detection. *Informatica* **2021**, *45*, 67–82. [[CrossRef](#)]
67. Eversberg, L.; Lambrecht, J. Combining synthetic images and deep active learning: Data-efficient training of an industrial object detection model. *J. Imaging* **2024**, *10*, 16. [[CrossRef](#)]
68. Ouarab, S.; Boutteau, R.; Romeo, K.; Lecomte, C.; Laignel, A.; Ragot, N.; Duval, F. Industrial Object Detection: Leveraging Synthetic Data for Training Deep Learning Models. In Proceedings of the International Conference on Industrial Engineering and Applications, Nice, France, 10–12 January 2024; Springer Nature: Cham, Switzerland, 2024; pp. 200–212.
69. Ouarab, S. Enhancing Industrial Object Detection with Synthetic Data for Deep Learning Model Training. Ph.D. Thesis, Higher School of Computer Science, Sidi Bel Abbès, Algeria, 2023.
70. Kapusi, T.P.; Erdei, T.I.; Husi, G.; Hajdu, A. Application of deep learning in the deployment of an industrial scara machine for real-time object detection. *Robotics* **2022**, *11*, 69. [[CrossRef](#)]
71. Puttemans, S.; Callemeyn, T.; Goedemé, T. Building robust industrial applicable object detection models using transfer learning and single pass deep learning architectures. *arXiv* **2020**, arXiv:2007.04666.
72. Rhee, J.; Park, J.; Lee, J.; Ahn, H.; Pham, L.H.; Jeon, J. A Safety System for Industrial Fields using YOLO Object Detection with Deep Learning. In Proceedings of the 2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC), Grand Hyatt Jeju, Republic of Korea, 25–28 June 2023; pp. 1–6.
73. IEEE 802.11ad-2012; IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band. IEEE Standards Association: Piscataway, NJ, USA, 2012.
74. Jocher, G.; Qiu, J.; Chaurasia, A. *Ultralytics YOLO11*; Version 11.0.0; Ultralytics: Frederick, MD, USA, 2024.
75. Hidayatullah, P.; Syakrani, N.; Sholahuddin, M.R.; Gelar, T.; Tubagus, R. YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review. *arXiv* **2025**, arXiv:2501.13400.
76. Jocher, G.; Chaurasia, A.; Qiu, J. *Ultralytics YOLO (Version 8.0.0) [Software]*; Ultralytics: Frederick, MD, USA, 2023.
77. Tzutalin, D. *Tzutalin/Labelimg*; Github: Online, 2015.
78. Cao, X.; Su, Y.; Geng, X.; Wang, Y. YOLO-SF: YOLO for fire segmentation detection. *IEEE Access* **2023**, *11*, 111079–111092. [[CrossRef](#)]
79. Zhou, J.; Zhang, B.; Yuan, X.; Lian, C.; Ji, L.; Zhang, Q.; Yue, J. YOLO-CIR: The network based on YOLO and ConvNeXt for infrared object detection. *Infrared Phys. Technol.* **2023**, *131*, 104703. [[CrossRef](#)]
80. Su, P.; Han, H.; Liu, M.; Yang, T.; Liu, S. MOD-YOLO: Rethinking the YOLO architecture at the level of feature information and applying it to crack detection. *Expert Syst. Appl.* **2024**, *237*, 121346. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.