www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 18(1), 2027, 32-49. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2027.v18i1.1483

Intelligent Industrial Risk Assessment using Computer Vision and Hierarchical Fuzzy Logic

Oman Fernando García-Franco[1], Kevin Martin Trujeque Sánchez[1], Manuel Iván Manríquez Calderón[2], Maury del mar Zarco Calderón[2].

[1].Universidad del Valle de Puebla, Puebla, México

[2].Tecnológico Nacional de México, Pachuca, Hgo., México

II46511@uvp.edu.mx; manuel.mc@pachuca.tecnm.mx; maury.zc@pachuca.tecnm.mx

Corresponding Author: ptc.investigacion4@uvp.mx

Abstract. Computer vision-based safety monitoring often focuses on PPE compliance, but operational risk is gradual, spatial, temporal, and context-dependent. This work proposes an interpretable hierarchical risk-scoring system that integrates YOLOv11-based PPE detection, regions of interest, homography-based spatial correction, and zero-order Sugeno-Takagi fuzzy inference. Seven observable components are considered: weighted PPE compliance, critical PPE absence, homography-corrected ROI proximity, dwell time, machinery severity, local density, and movement intensity. These variables are aggregated into protection deficit, hazard exposure, and operational context to estimate instantaneous and temporally smoothed risk scores in [0,100]. The pilot validation was conducted in a university manufacturing laboratory using 1,611 PPE images, ROI activation tests, and six controlled scenarios. YOLOv11 achieved mAP@0.5=0.886; homography reduced activation errors by up to 93.1%. The proposed approach should be understood as an online-compatible decision-support tool for safety supervision and training, rather than a replacement for institutional safety protocols or comprehensive occupational risk assessment.

Keywords: Fuzzy logic, ROI, YOLO, industrial safety, Sugeno inference, risk assessment.

Article Info
Received: Jun 16, 2026
Accepted: Aug 01, 2026

1. Introduction

Industrial safety requires monitoring mechanisms capable of identifying risk conditions before they escalate into incidents. Industrial spaces bring together machinery, bounded operating areas, the movement of people, changes in lighting, and the mandatory use of personal protective equipment (PPE). Although human supervision remains indispensable, its continuous application can become limited when multiple work areas, visual occlusions, shifting activities, and users with varying levels of experience in operating machinery coexist.

Computer vision has been used as a supporting tool in safety monitoring tasks, particularly through the automatic detection of people and PPE. In addition, detection models such as YOLO have proven useful in surveillance and protocol compliance verification applications, owing to their ability to operate in near real time. However, this perspective is insufficient to represent operational risk, since a person wearing partial PPE does not face the same level of risk when standing far from a machine as when standing near a hazardous region or inside a high severity operating zone.

Beyond PPE compliance, the location of the person within the environment is a decisive factor. Regions of interest (ROIs) make it possible to computationally represent risk zones associated with machinery or operating areas; nevertheless, evaluating them directly on the image can produce errors when the camera perspective creates visual overlap between a person and a hazardous zone. Under these conditions, a person may appear to be inside a ROI without physically occupying the risk area. For this reason, geometric correction through homography becomes relevant, as it projects each person's support point onto an approximate top-down plane and improves the spatial consistency of the assessment.

Even with visual detection and geometric correction, risk cannot be adequately estimated through rigid thresholds. Safety conditions are gradual and uncertain: PPE compliance may be partial, proximity to the hazardous zone may vary, dwell time may accumulate over time, and the operational context may change due to the density of people or rapid movement. In this sense, fuzzy logic offers a suitable strategy to integrate heterogeneous variables and produce a continuous, interpretable score. In particular, zero order Sugeno-Takagi systems allow efficient numerical outputs to be obtained while preserving traceability over the activated rules.

Building on the above, this work proposes a hierarchical fuzzy risk scoring system for industrial safety monitoring in a university manufacturing laboratory, used as a controlled pilot environment. The proposal integrates person and PPE detection through YOLOv11, the association of protective elements per person, the delimitation of regions of interest, spatial correction through homography, and a hierarchical fuzzy inference system. The model computes three intermediate dimensions: protection deficit, hazard exposure, and operational context. These dimensions are then aggregated through Sugeno-Takagi rules to produce an instantaneous risk index and a temporally smoothed version.

The main contribution of this work is a hierarchical, interpretable, and traceable system for continuous risk scoring based on a fuzzy architecture that operationalizes the multifactorial nature of industrial risk from visual, geometric, temporal, and contextual variables, integrating seven analysis components: 1) weighted PPE compliance; 2) absence of critical elements; 3) proximity to regions of interest projected through homography; 4) dwell time in risk zones; 5) severity associated with the machinery; 6) local density of people; and 7) speed or movement intensity. This integration yields a more informative score and facilitates the interpretation of the conditions that increase risk.

The study is carried out in a university manufacturing laboratory because this environment allows conditions relevant to operational safety to be evaluated in a controlled manner, such as machinery, delimited zones, user movement, variation in PPE compliance, proximity to risk areas, and dwell time within supervised regions.

The experimental evaluation was conducted using a proprietary PPE dataset, ROI activation tests with and without homography, and six controlled scenarios designed to represent low, moderate, high, and critical risk conditions. The results make it possible to analyze the performance of the visual detector, the usefulness of geometric correction, and the consistency of the fuzzy response to changes in protection, exposure, and operational context.

2. State of the Art

The review of the state of the art is organized around three complementary axes: 1) PPE detection and verification systems based on computer vision; 2) contextual, spatial, and operational analysis for risk assessment; and 3) fuzzy scoring systems and interpretable risk indices.

First, computer vision assisted occupational safety has become a relevant topic across different settings, particularly in construction and industry, where continuous human supervision may prove limited or insufficient. Recent research has noted that digital technologies applied to safety improve detection capability, event traceability, and early response to hazardous conditions (Dodoo et al., 2024). Along these lines, computer vision stands out for its ability to operate on video streams from surveillance cameras and to extract useful information about people, objects, trajectories, and environmental conditions (Li et al., 2026).

One of the most actively developed fields has been the monitoring of PPE compliance. Vukicevic et al. (2024) present a systematic review of computer vision systems applied to PPE compliance in industrial practice, identifying significant advances in the automatic detection of helmets, vests, gloves, glasses, and other safety items.

More specifically, work on PPE compliance has shown important progress through deep learning architectures, mainly models derived from YOLO, Faster R-CNN, and real time detection approaches (Rasouli et al., 2024; López et al., 2025). Iannizzotto et al. (2021) explore the combination of deep neural networks and fuzzy logic for PPE detection on embedded devices, showing that the relationship among deep learning, PPE, and fuzzy logic has already been considered in the literature, primarily to strengthen the visual decision.

Although these studies differ in architecture, experimental domain, or type of PPE evaluated, they share a common idea: transforming manual visual inspection into an automated system for compliance verification. This shows that automatic PPE detection is a well-established line of work; however, its main objective usually remains verifying compliance or identifying the absence of specific protective elements.

The second axis corresponds to contextual, spatial, and operational analysis for risk assessment. In safety monitoring systems, the actual level of exposure depends on where events occur, on dwell times, and on the dynamics between people and machinery in the environment. For this reason, regions of interest, exclusion zones, distance to machinery, dwell time in hazardous areas, zone severity, and the density and movement of people constitute relevant variables that complement PPE verification.

From this perspective, approaches have been proposed for real time hazard recognition, monitoring of entry into hazardous areas, surveillance of restricted zones, and safety distance warning (Ayoubi & Arashpour, 2026; Tran et al., 2025). These studies show that risk depends not only on the detected objects but also on the spatial relationships among workers, machinery, and operating zones.

Spatial analysis has also been addressed through systems that combine visual detection with surveillance of restricted or hazardous zones. Toth et al. (2026) propose a modular computer vision approach for industrial safety, integrating PPE compliance monitoring with hazardous zone surveillance. Their system represents one of the closest precedents, although its output is oriented mainly toward detecting non-compliance and spatial violations. In turn, Li et al. (2026) develop a fast safety distance warning framework to detect proximity between people and hazardous devices during electrical construction operations, using oriented object detection, tracking, and distance estimation with a monocular camera.

In a related manner, there is also evidence of the use of transformations to a planar map to visualize the proximity between workers and equipment, employing homography to represent safety related spatial relationships in a top-down view (Shin & Kim, 2022). Likewise, multi-camera detection and tracking approaches highlight the importance of preserving information about the location, trajectory, and dwell time of people in complex scenes (López-Cifuentes et al., 2022).

Taken together, this prior work shows that spatial and operational analysis enriches the interpretation of risk, although many approaches are oriented mainly toward detecting intrusions, distance violations, or specific exposure events.

The third axis corresponds to the use of fuzzy scoring systems and interpretable risk indices. In occupational safety, risk can rarely be adequately represented through strictly binary criteria, because heterogeneous, gradual, and uncertain factors are involved. Fuzzy logic makes it possible to model linguistic variables, incorporate expert judgment, and transform incomplete or ambiguous information into scores useful for supporting decision-making.

Gürcanli and Müngen (2009) propose an occupational risk analysis method for construction sites based on fuzzy sets, aimed at handling the uncertainty associated with estimating probability, severity and safety conditions. Similarly, Ilbahar et al. (2018) develop an integrated approach for occupational health and safety risk assessment through a fuzzy analytic hierarchy process and a fuzzy inference system, combining criteria derived from traditional schemes, such as Fine-Kinney, with techniques capable of representing ambiguity and subjectivity.

Ahmed et al. (2023) present an integrated approach for occupational health and safety risk assessment in small enterprises in Bangladesh, while Aziz et al. (2025) use fuzzy logic to estimate risk and precaution levels for explosive, flammable, and toxic substances. These works reinforce the suitability of fuzzy systems for representing gradual risk conditions in the presence of multiple variables and different levels of severity.

The integration of visual perception and fuzzy inference has also been explored in safety applications related to collisions, moving objects, workspace management, and ergonomic assessment (Kim et al., 2016; Martínez et al., 2026). Likewise, predictive risk models in construction confirm the trend toward computational systems capable of combining multiple factors to anticipate or classify hazardous conditions (Sadeghi et al., 2020).

Based on the previous review, the literature has advanced in three complementary directions: the automatic detection of PPE through computer vision, the contextual analysis of hazardous zones, and the construction of risk indices through fuzzy logic. Nevertheless, these components are usually addressed only partially. Some studies focus on verifying PPE compliance or improving the visual detection of protective elements; others prioritize identifying unauthorized entries, distance violations, or interaction with hazardous zones; and others build fuzzy indices from expert judgment, historical records, specific sensors, ergonomic conditions, or particular collision risks.

In this regard, the present research is positioned at the intersection of these three axes. Its contribution is oriented toward articulating visual, spatial, temporal, and operational evidence within a hierarchical Sugeno-Takagi structure, in which risk is decomposed into intermediate dimensions of protection deficit, hazard exposure, and operational context. This organization makes it possible to integrate weighted PPE compliance, absence of critical elements, homography corrected proximity to regions of interest, dwell time, machinery severity, local density, and movement intensity to produce a continuous, interpretable, and traceable score of operational risk per person.

3. Methodology

The proposed system is structured as a hierarchical visual, spatial, and fuzzy analysis pipeline aimed at estimating operational risk per person inside a manufacturing laboratory. Accordingly, the system integrates computational perception, geometric correction, and fuzzy reasoning modules to convert visual observations into a continuous risk score.

The methodology is organized into three main stages. The first corresponds to the extraction of visual and spatial evidence, where video frames are acquired, people are detected through YOLOv11, a support point is computed for each individual, that point is projected onto the top-down plane through homography, and its relationship with the regions of interest associated with machinery or hazard zones is evaluated. In this same stage, PPE elements are associated with each person in order to identify compliance, absence of critical elements, and visually relevant safety conditions. This stage produces the observable variables that feed the fuzzy system.

The second stage corresponds to the definition and fuzzification of the input variables, in which the evidence obtained from the video is transformed into membership degrees through triangular and trapezoidal functions.

Finally, the third stage implements a zero-order hierarchical Sugeno-Takagi inference architecture, where three intermediate dimensions are computed: protection deficit, hazard exposure, and operational context. These dimensions are subsequently integrated into a final rule matrix to obtain the instantaneous risk index and its temporally smoothed version.

With this organization, the system separates visual perception from risk inference, maintaining traceability among detections, generated variables, and the final score.

3.1 Visual and spatial evidence extraction

This stage transforms each video frame into a set of observable variables that subsequently feed the fuzzy subsystems of protection deficit, hazard exposure, and operational context.

Evidence extraction is organized into four functional blocks: 1) video capture and preprocessing; 2) person detection and support point estimation; 3) geometric correction through homography and evaluation of regions of interest; and 4) PPE association per person and estimation of contextual variables. The system was designed to operate both on near real time video and on previously recorded sequences.

In particular, visual processing yields bounding boxes, support points, homography corrected positions, membership in or proximity to ROIs, PPE compliance, critical absence, dwell time, local density, and movement intensity. These variables do not constitute the final risk score; they represent the input evidence that will subsequently be integrated by the Sugeno-Takagi architecture.

The system's process diagram, together with its input and output elements and the variables to be generated, which will be detailed in the following sections, can be seen in Figure 1.

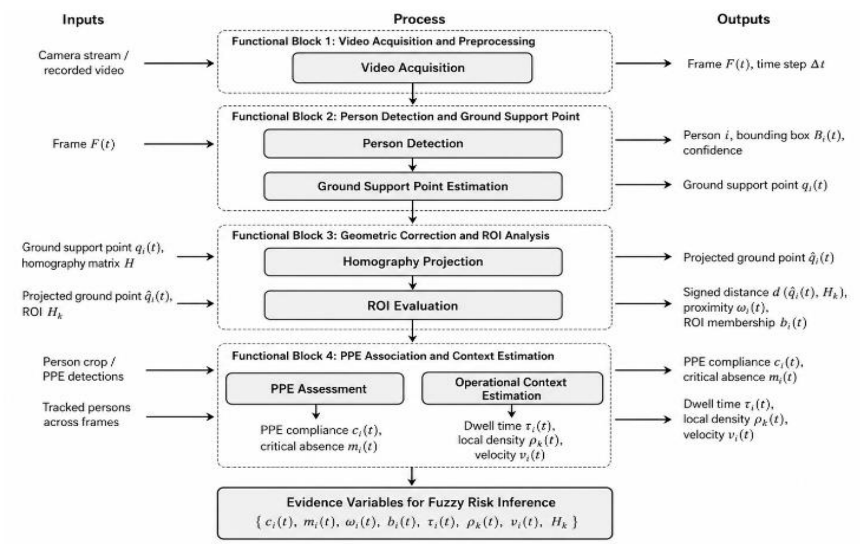

Fig. 1. Process diagram, inputs, outputs, and variables of the visual and spatial evidence extraction system.

3.1.1 Video capture and preprocessing.

The test environment corresponds to the manufacturing laboratory of the Universidad del Valle de Puebla (UVP), where operating zones associated with machinery such as a lathe, milling machine, electrode welder, and cutter are identified. Each zone is represented by a region of interest H_k on the top-down plane of the laboratory and can be defined within the interface of the ROI delimitation program. The specifications of the video capture device, as well as of the video clips obtained for the tests, are reported (Table 1).

Table 1. Table of specifications of video equipment used for pilot tests

Item	Description / value to be reported
Test Type	Video and recorded footage
Camera used	USB 2.0 HD UVX WebCam
Resolution	1280×720 px for recorded footage and 1920×1080 for video
Average FPS	17-45 fps
Total Video Length	Samples from 60 to 90 secs
Number of sequences	70
Number of participants	Approximately 5 per video
ROI Zones	Lathe, milling cutter, welding machine and cutter
Observed conditions	Lighting, occlusion, aisle transit, ROI permanence

From the camera stream or recorded video, the system obtains frames F(t) and computes the temporal interval Δt between consecutive frames. These two outputs are necessary to synchronize the subsequent stages of detection, tracking, dwell-time estimation, and speed computation. In this sense, the capture stage provides not only the input image but also the temporal reference on which the system's dynamic variables are computed.

3.1.2. Person detection and support point

In the stage of detecting people in the frame, the YOLOv11 image recognition model was used. This model was chosen for its ability to perform object detection in near real time, a necessary feature for continuous monitoring applications.

For each identified person, the detector produces a bounding box, a class label, and a confidence value. For each frame F(t), the detector identifies the people present in the scene and marks them with a bounding box $B_i(t) = (x_1, y_1, x_2, y_2)$; from it, a support point is computed, associated with the person's approximate contact with the ground plane. This point is defined as the lower center of the box and is used to evaluate the spatial relationship with the ROIs, as shown in Figure. 2.

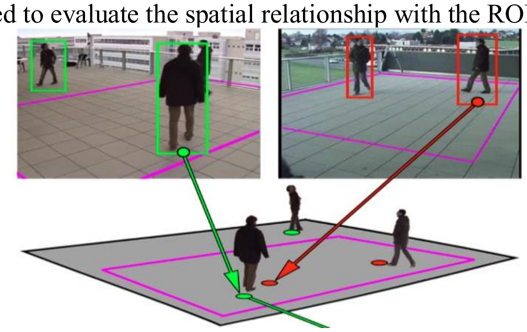

Fig. 2. Recognition of points x_1, y_1, x_2, y_2 at the lower half of the bounding box (red, green) and their projection onto the ROI (purple). Adapted from (López-Cifuentes et al., 2022).

Choosing the lower-center point avoids using the geometric center of the box, which may fall on the torso and fail to correctly represent the person's position on the floor. This decision is relevant for cameras with a pronounced perspective, where evaluating directly on the raw image can produce false ROI activations.

3.1.3 Geometric correction through spatial homography.

To reduce the errors arising from the perspective change between the real 3D environment and its dimensional reduction to a 2D image, the system incorporates a homographic transformation of the ground plane.

The homography is obtained from reference points selected in the laboratory, which belong to the dominant working plane.

These points allow a transformation matrix to be computed, through which each person's support point is projected from the original image onto an approximate top-down view. In this way, the spatial evaluation is not performed directly on the raw image, but on a plane where the relationships between people and risk zones are more consistent with the physical layout of the laboratory.

This transformation is particularly important in situations where a person visually crosses in front of a machine or marked zone without actually occupying the hazardous area.

On the other hand, the k-th regions of interest H_k are defined as polygons on the top-down plane. The system computes their spatial relationship with each ROI through a point-in-polygon test. From this test, the signed distance $d(q_{i(t)}, H_k)$ and the proximity $\omega_i(t)$ to the ROI are obtained. These variables make it possible to represent the spatial relationship between each detected person and the risk regions defined on the top-down plane, reducing errors arising from the camera perspective.

3.1.4 PPE association per person.

PPE compliance is computed from the elements detected and associated with each person. This association is carried out from the person crops produced by YOLO and the direct detection of PPE classes performed in Roboflow, using a proprietary dataset of 1,611 images adjusted to the laboratory environment (Table 2).

Table 2. Composition by number of dataset images

PPE	Number of images	Train	Valid	Test
Safety boots	637	441	127	69
Gloves	352	287	40	25
Safety Googles	400	274	81	45
Lab Coat	115	80	24	11
Face Shield	107	97	5	5
Total	1611	1179	277	155

The expected result is a binary variable $d_{ij}(t)$, indicating whether element j was detected for person i at time t. This PPE association distinguishes between required elements and critical elements. This separation is necessary because the absence of a critical element, such as a helmet or gloves in certain zones, must affect the protection deficit more strongly than the absence of a conditional element.

In addition to PPE association, the system estimates contextual variables from the people active in consecutive frames. The dwell time $\tau_i(t)$is computed by accumulating the interval Δt during which person (i) remains inside or near a ROI. The local density $\rho_k(t)$ represents the concentration of people associated with a region H_k, normalized with respect to an expected maximum value. In turn, the speed or movement intensity $v_i(t)$ is obtained from the displacement of the projected point.

3.2. Definition of variables and fuzzification functions.

The second stage of the system takes the evidence gathered by the computational perception module to assign a risk score.

Let F(t) be the video frame acquired at time t. For each person i, the system builds a dynamic state containing smoothed PPE compliance, dwell time in hazardous zones, last observed position, smoothed risk, and activated rules. The final risk score $\mathrm{r}_{\mathrm{i(t)}}$ is modeled as a hierarchical function of three dimensions:

$$\mathrm{r}_{\mathrm{i(t)}} = \Phi\big(\mathrm{P}_{\mathrm{i(t)}}, \mathrm{E}_{\mathrm{i(t)}}, \mathrm{K}_{\mathrm{i(t)}}\big): \quad \mathrm{r}_{\mathrm{i(t)}} \in [0,100]. \tag{1}$$

Where $\mathrm{P}_{\mathrm{i(t)}}$ denotes the protection deficit, $\mathrm{E}_{\mathrm{i(t)}}$ the hazard exposure, and $\mathrm{K}_{\mathrm{i(t)}}$ the operational context. The system is implemented as a composition of three first level fuzzy subsystems and a final second level fuzzy aggregator. Each dimension in turn has different sub-dimensions, where:

$$P_{i(t)} = \psi_p\big(c_i(t), m_i(t)\big) \in [0,1]. \tag{2}$$

$$\mathrm{E}_{\mathrm{i(t)}} = \psi_E\big(\omega_i(t), s_k, \tau_i(t)\big) \in [0,1]. \tag{3}$$

$$\mathrm{K}_{\mathrm{i(t)}} = \psi_K\big(\rho_k(t), v_i(t)\big) \in [0,1]. \tag{4}$$

Where $c_i(t)$ denotes PPE compliance, $m_i(t)$ the absence of critical PPE, $\omega_i(t)$ proximity to the hazard zone, $\tau_i(t)$ dwell time in the hazard zone, s the hazard severity coefficient, $\rho(t)$ the congestion or crowding of the zone, and $v_i(t)$ the intensity of the movement recorded in the zone.

3.2.1 Protection deficit $\mathbf{P_{i(t)}}$

This dimension evaluates the instantaneous compliance of a set of protective elements; the definition of which elements are required and their criticality depend on the type of machinery located in the laboratory (Table 3). For the PPE compliance parameter $c_i(t)$, the set R is defined as the set of PPE elements required by institutional protocol, and $w_j > 0$ is the criticality assigned to element j; the values assigned to w_j (Table 4).

Table 3 & 4. PPE matrix and its safety condition for each machinery available in the laboratory (left) and criticality weight associated with each PPE requirement condition w_j (right) tables Source: Authors

Machinery	Safety Glasses	Safety Boots	Leather Gloves	Laboratory Coat	Face Shield
Lathe	Critical	Critical	Not Applicable	Conditional	Not Applicable
Milling Machine	Critical	Critical	Not Applicable	Conditional	Not Applicable
Stick Welding Machine	Not Applicable	Critical	Critical	Critical	Critical
Cutting Machine	Critical	Conditional	Critical	Conditional	Not Applicable

Safety Condition	Value w_j
Not applicable	0.2
Conditional	0.6
Critical	1

If d_{ij}(t) ∈ {0,1} indicating whether element j was detected for person i at time t, then the instantaneous compliance is computed as:

$$c_{i(t)} = \frac{\sum_{j\in R} w_j d_{ij}(\mathrm{t})}{\sum_{j\in R} w_j} \tag{5}$$

Thus, $c_i(t) = 1$ implies full compliance and zero in the total absence of required PPE.

On the other hand, for the parameter $m_i(t)$, is defined as the objects classified as critical under the usage contexts of each machine, taking the value 1 when at least one element classified as critical for the current work zone is missing, and 0 otherwise.

3.2.2 Hazard exposure $\mathbf{E_{i(t)}}$.

For this dimension, the distances associated with each person are evaluated with respect to the plane of the ROIs, which are placed to delimit the working zones of each machine located in the laboratory; these ROIs are in turn defined through the homography associated with the laboratory floor plane.

To carry out the normalized proximity analysis $\omega_i(t)$ for each hazardous ROI zone H_k, a maximum reference distance $d_{max}(H_k)$ is defined; likewise, a representative point $q_i(t)$is assigned to each person, located on the top-down plane of the homography of H_k; membership and the signed distance to the contour are then evaluated through a PointInPolygon (PIP) test. If the observed distance $d(q_i, H_k)$ is positive inside the polygon and negative outside it, the effective distance to the hazard $\delta_i(t)$is defined as:

$$\delta_i(t) = \max[0, -d(q_i(t), H_k)]. \tag{6}$$

Finally, the proximity is expressed as a bounded measure in [0,1], normalized from a maximum measured distance $d_{max}(H_k)$ in pixels on the top-down plane of the homography, defined by design and calibration of the system.

$$\omega_i(t) = min\left[\frac{\delta_i(t)}{d_{max}(H_k)}, 1\right] \tag{7}$$

To compute the dwell time $\tau_i(t)$in H_k a binary criterion $b(t)$ is defined indicating whether person i is inside H_k at time t; the accumulated dwell time $\tau_i(t)$ is updated incrementally and with decay memory when the worker leaves the zone:

$$\tau_i(t) = \begin{cases} \tau_i(t-1) + \Delta t : \ b(t) = 1 \\ \lambda(t)\, \tau_i(t-1) : b(t) = 0 \end{cases} \tag{8}$$

Where $\lambda(t)$ is the temporal memory factor that gradually reduces the accumulated persistence when the person leaves H_k, defined as:

$$\lambda(t) = e^{-\frac{\Delta t}{T}} \tag{9}$$

Where T is proposed as a memory window calibrated according to system conditions (T=5s, configurable), and the temporal interval Δt is computed dynamically from the system clock

$$\Delta t = t_{actual} - t_{previous} \approx \frac{1}{fps} \tag{10}$$

This allows dwell-time accumulation and speed computation to adapt to real variations in the processing rate, especially in scenarios where YOLO model inference and homography processing can modify the effective number of frames per second of the camera used.

The severity s_k is taken directly from the active ROI associated with the machine position and is constant depending on which working zone is active (Table 5), these active zones can be configures manually by the user in the interface.

Table 5. Level of severity s_k associated with each machine present in the laboratory

Machine	Value s_K	Machine Severity
Lathe	0.3	Low
Milling Machine		
Electrode welder	0.7	Medium
Cutter	0.9	High

3.2.3 Operational context $K_i(t)$

This dimension seeks to define the situations surrounding people located near or inside the ROIs H_k , namely the concentration of people and their movement intensity.

The local density of workers $\rho_k(t)$ around the zone is defined from the number $n_i(t)$ of people active in the vicinity of the ROI, normalized by the maximum number of people detected in the zone n_{max}, which in the pilot version was set to 5 people.

$$\rho_k(t) = min\left[\frac{n_k(t)}{n_{max}}, 1\right] \tag{11}$$

The movement intensity $v_i(t)$ is estimated from the individual's 2D reference position $x_i(t)$ and its previous positions, from the displacement between frames of the people near H_k:

$$v_i(t) = min\left[\frac{||x_i(t) - x_i(t-1)||}{\Delta t \cdot V_{max}}, 1\right] \tag{12}$$

Finally, V_{max} is a raw saturation speed measured in pixels/s, defined by experimental calibration of the environment; for the pilot tests it was set to a saturation speed of 250 px/s.

3.2.4 Proposed membership functions

The membership functions are implemented through parameterized triangular and trapezoidal shapes. To maintain consistency with low resource devices, the generic definitions of triangular and trapezoidal membership functions are used.

Through the membership functions μ, the design establishes three unified linguistic labels for PPE compliance $c_i(t)$: low, medium, and high; while for the proximity $\omega_i(t)$: near, medium, far. The boundary values of each membership function, as well as their overlaps and correspondence rules, are shown in Figure 3.

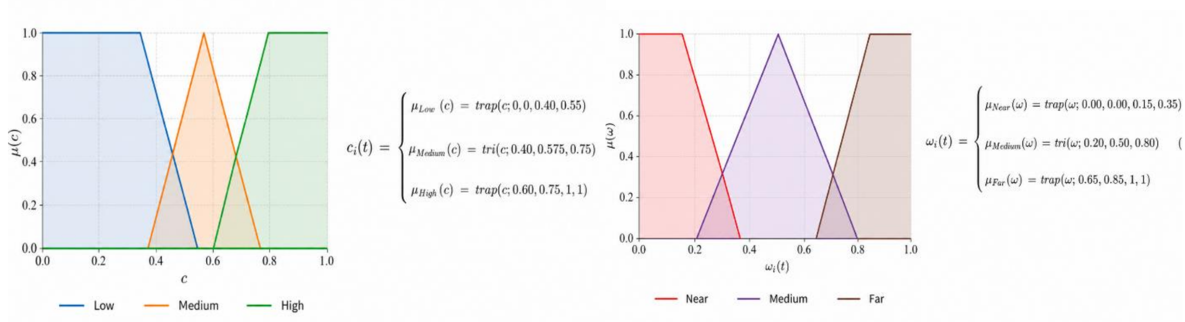

Fig. 3. Graphs of the membership functions for $c_i(t)$ (left) and $\omega_i(t)$ (bottom) and associated equations. Source: Authors.

For the variables $\tau_i(t)$ the labels short, medium, and long are used (measured in seconds); for $\rho_k(t)$: low, medium, and high; and finally for $v_i(t)$: slow, medium, and fast. The values of the membership functions and their equations are shown in Figure 4.

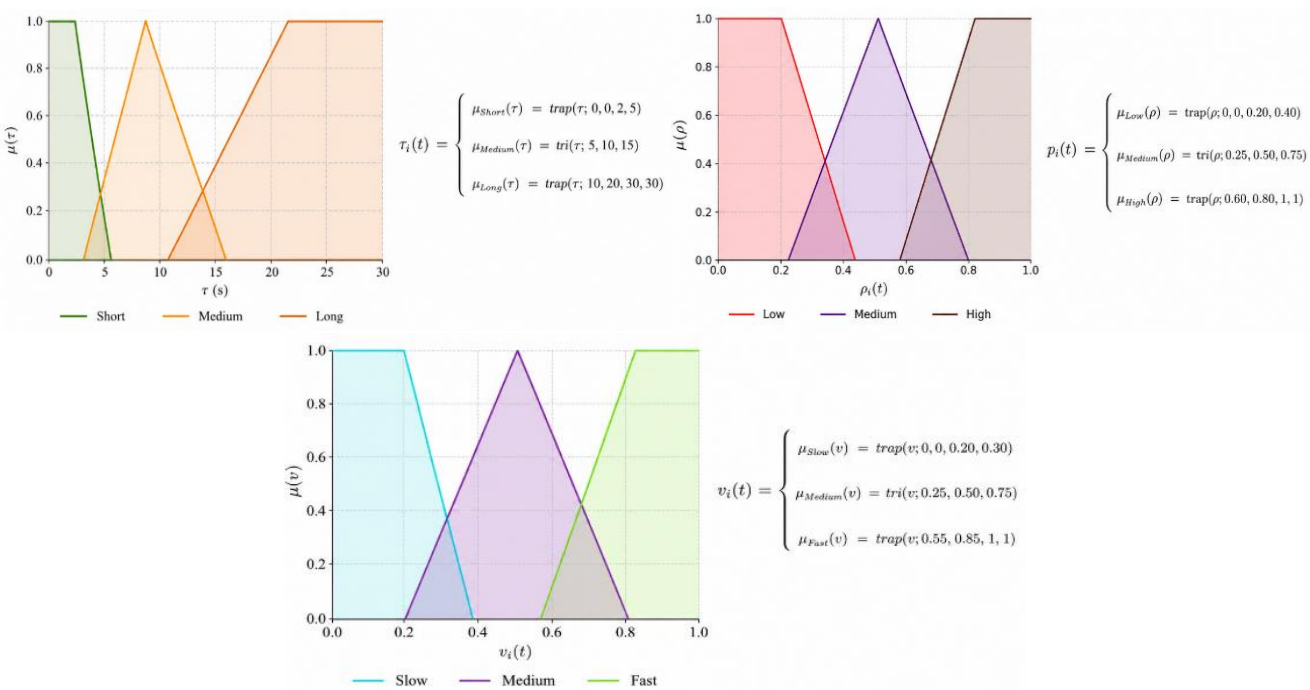

Fig. 4. Graphs of the membership functions proposed for $\tau_i(t)$ (top left), $\rho_k(t)$ (top right) and $v_i(t)$ (bottom).

3.3 Implementation of the hierarchical Sugeno-Takagi architecture.

The inference architecture was designed to transform the previously fuzzified variables and membership functions into a final risk score per person.

The proposed fuzzy inference model is structured as a zero-order hierarchical Sugeno-Takagi architecture. This choice responds to two main reasons. First, operational risk depends on heterogeneous variables, personal protection, spatial proximity, dwell time, severity, density, and movement, which cannot be adequately interpreted through binary rules. Second, a flat fuzzy architecture would generate a combinatorial explosion of rules by combining all input variables into a single system; by dividing the model into the three intermediate subsystems, complexity is considerably reduced. Considering three linguistic labels per input, the protection subsystem $P_i(t)$ requires 6 rules, the operational context $K_i(t)$ requires 9 rules, while the exposure subsystem $E_i(t)$ requires 27 and the final module $r_{i(t)}$ require 36. Thus, the system's hierarchical architecture requires 78 rules, compared with the 1,458 combinations it could require when considering six inputs, three labels per variable, and one binary variable.

Although this number of combinations can be computationally processable on modern equipment, its main limitation lies in the difficulty of designing, reviewing, calibrating, and interpreting a large flat knowledge base. Moreover, in a video system, inference must be executed repeatedly per frame and per detected person, so a more compact rule base favors a more stable, traceable, and scalable implementation.

The choice of a zero order Sugeno-Takagi architecture is not posed as a direct replacement for other fuzzy inference schemes, but as a methodological decision consistent with the nature of the output required by the system: a continuous numerical risk score per person, amenable to temporal smoothing. Sugeno-Takagi models allow linguistic rules to be represented in the antecedents and constant or functional consequents in the output, which facilitates obtaining scalar values through a weighted average of the activated rules (Takagi & Sugeno, 1985). This feature is suitable when the fuzzy system must be integrated with a computer vision computational pipeline, where the input variables come from normalized measurements, as is the case here.

From the membership functions for the previously defined input variables, risk estimation is performed through a zero-order hierarchical Sugeno-Takagi fuzzy architecture. In this type of system, the consequents of the rules are not fuzzy sets but constant values associated with output levels. If a rule q has antecedents A, B, and C, its firing strength (weight) $\beta_{q(t)}$ is:

$$\beta_{q(t)} = \min[\, \mu_A(x_1), \mu_B(x_2), \mu_C(x_3)] \tag{13}$$

In turn, the output of any subsystem is obtained as a weighted average $y_{s(t)}$ of its activated consequents γ_q^R. Therefore, the final risk depends simultaneously on the activated linguistic labels and on the intensity with which they are activated.

$$y_{s(t)} = \frac{\sum_{q=1}^{Qs} \beta_q^S(t)\gamma_q^S}{\sum_{q=1}^{Qs} \beta_q^S\,(t)} \tag{14}$$

Each of the rule subsystems $P_{i(t)}, E_{i(t)}, K_{i(t)}$ has an individually constructed matrix of rules and consequents. Although these intermediate outputs are continuous values, they are re-evaluated through the linguistic labels low, medium, and high to activate the rules of the final aggregator.

To clarify the internal organization of the model, Figure 5 presents the hierarchical Takagi-Sugeno inference architecture used to estimate operational risk. It shows how the variables extracted from visual processing are grouped and then integrated into a final aggregator to obtain the instantaneous risk.

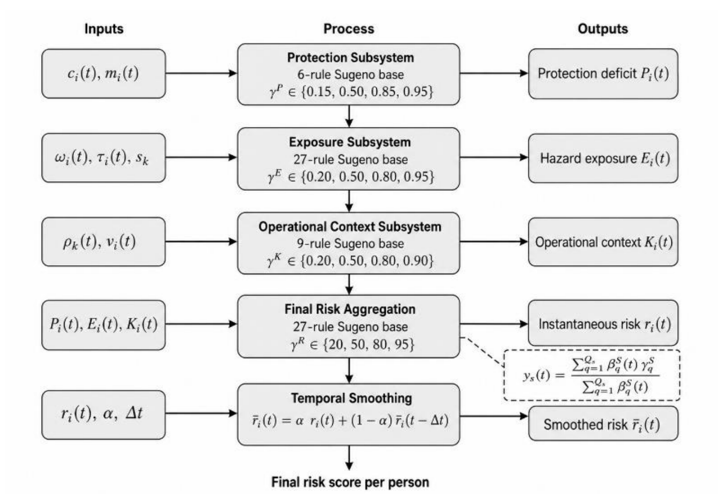

Fig. 5. Hierarchical Takag-Sugeno architecture for operational risk inference.

3.3.1 Sugeno rule bases and aggregation matrices

The rule bases of the fuzzy subsystems were defined through expert judgment and preliminary calibration of the laboratory environment, prioritizing three principles: monotonicity, interpretability, and a conservative response to higher risk conditions. Accordingly, the consequents increase as PPE compliance decreases, as a critical absence of protective elements appears, as proximity to a ROI increases, as dwell time lengthens, as machinery severity increases, as local density grows, or as the movement of people intensifies.

Table 6 presents the rule matrix of the protection deficit subsystem $P_i(t)$, built from the weighted PPE compliance $c_i(t)$ and the critical absence $m_i(t)$. This matrix assigns higher values when compliance decreases and, especially, when a critical element is missing.

Table 6. Sugeno rule matrix for the protection deficit $P_i(t)$. The consequents {0.15, 0.50, 0.85, 0.95} correspond to low, medium, high, and extreme deficit, respectively

PPE Compliance $c_i(t)$	Critical PPE Absence $m_i(t)$	
	No ($m_i(t) = 0$)	Yes ($m_i(t) = 1$)
High	0.15	0.50
Medium	0.50	0.85
Low	0.85	0.95

Table 7 groups the matrices of the hazard exposure subsystem $E_i(t)$, defined as a function of the proximity to the ROI $\omega_i(t)$, the dwell time $\tau_i(t)$ and the machinery severity s_k (Table 5). They show that exposure increases when the person is closer to the risk region, remains in it for longer, or the associated machinery has greater severity.

Table 7. Rule matrices for hazard exposure $E_i(t)$. The consequents {0.20, 0.50, 0.80} correspond to low, medium and high exposure, respectively

Low Severity s_k					Medium Severity s_k					High Severity s_k			
	Dwell Time $\tau_i(t)$					Dwell Time $\tau_i(t)$					Dwell Time $\tau_i(t)$		
Proximity $\omega_i(t)$	Short	Medium	Long		Proximity $\omega_i(t)$	Short	Medium	Long		Proximity $\omega_i(t)$	Short	Medium	Long
Far	0.20	0.20	0.20		Far	0.20	0.20	0.50		Far	0.20	0.50	0.50
Medium	0.20	0.50	0.50		Medium	0.20	0.50	0.50		Medium	0.50	0.80	0.80
Near	0.50	0.50	0.80		Near	0.50	0.80	0.80		Near	0.80	0.80	0.80

Table 8 shows the matrix of the operational context subsystem $K_i(t)$, built from the local density $\rho_k(t)$ and the movement intensity $v_i(t)$. This matrix represents scene conditions that can increase the probability of unsafe interaction, especially when high density of people and rapid movements coincide.

Table 8. Rule matrix for the operational context $K_i(t)$. The consequents used are ({0.20, 0.50, 0.80}), corresponding to low, medium, and high context, respectively

Local density $\rho_k(t)$	Movement Intensity $v_i(t)$		
	Slow	Medium	Fast
Low	0.20	0.20	0.50
Medium	0.50	0.50	0.80
High	0.50	0.80	0.80

Finally, Table 9 presents the final aggregation matrix of the instantaneous risk index $r_i(t)$, built from the intermediate outputs. For this purpose, the continuous outputs of the subsystems are reinterpreted through the linguistic labels low, medium, and high in order to activate the final risk rules. This organization allows the final risk to increase gradually as the protection deficit, the hazard exposure, and the adverse operational conditions increase.

Table 9. Rule matrix for the final aggregation of the instantaneous risk $r_i(t)$. The final consequents are ({20, 50, 80, 95}), corresponding to low, moderate, high, and critical risk

Low operational context $K_i(t)$.					Medium operational context $K_i(t)$.					High operational context $K_i(t)$.			
	Hazard exposure $E_i(t)$					Hazard exposure $E_i(t)$					Hazard exposure $E_i(t)$		
Protection deficit $P_i(t)$	Low	Medium	High		Protection deficit $P_i(t)$	Low	Medium	High		Protection deficit $P_i(t)$	Low	Medium	High
Low	0.20	0.20	0.20		Low	0.20	0.20	0.50		Low	0.20	0.50	0.50
Medium	0.20	0.50	0.50		Medium	0.20	0.50	0.80		Medium	0.50	0.80	0.80
High	0.50	0.50	0.80		High	0.50	0.80	0.80		High	0.80	0.80	0.95
Extreme	0.50	0.80	0.95		Extreme	0.80	0.95	0.95		Extreme	0.95	0.95	0.95

To avoid abrupt oscillations caused by visual noise, lighting changes, momentary occlusions, or false negatives, an exponential smoothing is applied to $r_i(t)$, where α is the exponential smoothing coefficient.

$$\bar{r}_i(t) = \alpha r_i(t) + (1-\alpha)\bar{r}_i(t-\Delta t) : 0 < \alpha < 1. \tag{15}$$

In the pilot implementation, a smoothing parameter of α=0.25 was used, so that the final index retains 75% of the previous smoothed value and incorporates 25% of the instantaneous risk computed in the current frame. This value was selected empirically to reduce oscillations due to visual noise without excessively delaying the system's response to real increases in risk. The value $\bar{r}_i(t)$ being the final score displayed on screen.

4. Results

The results are presented sequentially in order to evaluate the contribution of each component of the proposed hierarchical system. First, the ability of the visual perception module to identify the PPE classes relevant in the manufacturing laboratory is analyzed. Next, geometric correction through homography is examined as a mechanism to improve the spatial consistency between the detected people and the regions of interest associated with machinery. Finally, the response of the fuzzy system is evaluated in controlled scenarios designed to represent gradual conditions of low, moderate, high, and critical risk. This structure makes it possible to distinguish among the performance of the visual detector, the reliability of the spatial activation of the ROIs, and the final response of the Sugeno-Takagi model to combinations of protection, exposure, and operational context.

4.1 Evaluation of the visual perception module

To evaluate the visual perception module, a proprietary dataset oriented to the detection of personal protective equipment in the manufacturing laboratory was used. The classes considered were lab coat, boots, face shield, gloves, and glasses, as these are the most relevant elements for safe operation within the experimental setting. The behavior of the YOLOv11 detector was analyzed through the precision-recall curve (Figure 6), the normalized confusion matrix (Figure 7), and the training curves (Figure 8) in order to identify both the overall performance of the model and the classes most susceptible to error.

The precision-recall curve shows a performance with an overall value of (mAP@0.5=0.886) on the validation set. This metric considers as correct those detections whose intersection over union reaches at least 0.5, so the result indicates that the model maintains an adequate balance between precision and recall for most of the evaluated classes. In particular, the best performing classes are glasses (mAP=0.940), face shield (mAP=0.938), and boots (mAP=0.934), followed by gloves (mAP=0.898). The lab coat class shows the lowest value (mAP=0.797), suggesting greater sensitivity to posture variations, partial occlusion, visual deformation of the garment, or similarity to clothing not labeled as PPE.

Figure 11 shows the normalized confusion matrix, which allows the detector's error patterns to be identified in greater detail. The highest diagonal values correspond to face shield (1.00), glasses (0.95), boots (0.92), and gloves (0.90), while lab coat shows the lowest diagonal value (0.78). The most relevant errors are not concentrated in direct confusions among PPE classes, but in the relationship between some classes and the background. For example, part of the real lab coat instances are classified as background, and some background regions generate false positives mainly associated with boots, glasses, and lab coat.

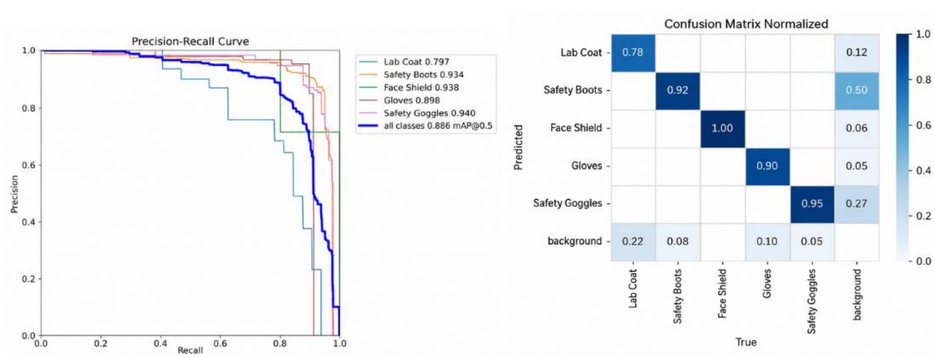

Fig. 6 & 7. Precision-recall curve of the model trained with the dataset in YOLOv11 for PPE detection (left) and YOLOv11 Model Normalized Confusion Matrix for PPE Detection (right).

The training curves, shows a progressive reduction of the losses associated with localization, classification, and edge distribution, both in training and in validation. This behavior indicates a functional convergence of the model and does not evidence severe overfitting, since the validation losses also decrease during training. Likewise, the precision, recall, and F1-score metrics per class show functional values for their integration into the risk system. Lab coat, face shield, and glasses maintain high recall values, while boots and gloves show greater susceptibility to false positives or false negatives. In particular, the gloves class must be interpreted with care, owing to its smaller visual size, possible occlusions by tools or body posture, and its relevance as critical PPE in certain ROIs.

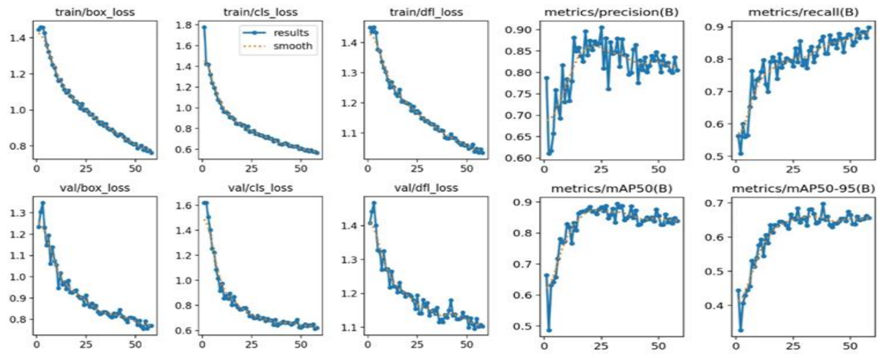

Fig. 8. Evolution of losses and metrics during the training of the YOLOv11 model for PPE detection.

4.2 Geometric evaluation of ROIs through homography

On the other hand, the geometric evaluation was aimed at verifying whether incorporating homography improves the spatial consistency of the system when activating regions of interest. For this purpose, the system's behavior was compared under two conditions: ROI activation on the original image, without geometric correction, and ROI activation on the top-down plane obtained through homography.

The evaluation was carried out over four representative scenarios of the manufacturing laboratory: crossing an aisle in front of a machine, real entry into an operating zone, dwell within a ROI, and the presence of a partially occluded person. For each scenario, 90 video events were analyzed, considering sequences of approximately two minutes (Table 10). In each case, the number of activation errors produced without homography and with homography was recorded.

An activation error was considered to be any spatial response inconsistent with the real situation of the scenario, including false activations in aisle crossings, missed activations at real entries, instability during dwell, and errors arising from partial occlusion of the support point.

Table 10. ROI activation comparison with and without homography

Scenario	Events evaluated	Activation errors without homography	Activation errors with homography	Relative reduction
Crossing aisle in front of machine	90	65	8	87.7%
Real Entry to ROI		72	5	93.1%
Permanence within ROI		48	6	87.5%
Partially occluded person		51	12	76.5%

The results show that evaluation on the top-down plane substantially reduces ROI activation errors across all analyzed scenarios. The most representative case corresponds to crossing an aisle in front of a machine, where the method without homography produced 65 errors, while the homography based evaluation reduced the value to 8 errors. In the real entry and dwell-within-ROI scenarios, a considerable reduction of errors was also observed.

These results indicate that homography not only reduces false activations in transit zones but also improves spatial stability when the person is effectively inside an operating zone.

The scenario with a partially occluded person showed the smallest relative reduction, although still relevant, going from 51 to 12 errors. This behavior was expected, since partial occlusion can affect the estimation of the person's support point and, therefore, the projection onto the top-down plane. Even so, the observed reduction suggests that the geometric approach retains advantages over direct evaluation on the image.

Taken together, the geometric evaluation shows that the use of homography provides a functional improvement to the risk analysis system, by reducing incorrect spatial activations and making the relationship among person, ROI, and the laboratory's physical plane more consistent. This improvement is especially important for the proposed fuzzy system, since the proximity ($\omega_i(t)$) and dwell ($\tau_i(t)$) variables depend directly on the spatial activation of the ROIs. Therefore, a reduction of geometric errors decreases the possibility that the exposure subsystem ($E_i(t)$) artificially increases the risk index due to perspective effects.

4.3 Design of controlled scenarios

To complement the evaluation of the fuzzy system, six controlled scenarios were defined through the systematic modification of the variables that feed the protection, exposure, and operational context subsystems. Each scenario was recorded through five independent video sequences of 60 to 90 seconds, obtaining on average 380 seconds of observation per scenario and 35 total minutes of analyzed material. The experimental conditions considered are summarized in Table 11. The scenarios were structured to elicit differentiated responses in the risk index.

Table 11. Proposed scenarios and their respective experimental conditions linked to the modified variables of the fuzzy system

Scenario	PPE Condition	Distance to ROI δ_1 (in px)	Permanence $\tau_i(t)$ (in s)	$\rho_k(t)$ (in No. of persons)	Movement $v_i(t)$ (in px/s)
S1	Full PPE	50≤ δ_1 <100	$\tau_i(t) < 5$	1	$v_i(t)$<150
S2	Partial PPE	20≤ $\delta_1 < 50$	5≤ $\tau_i(t) < 10$	1	150≤ $v_i(t)$ <200
S3	Lack of critical PPE	0≤ δ_1 <20	5<10≤ $\tau_i(t)$	1	150 ≤ $v_i(t)$<200
S4	Lack of critical PPE	0≤ $\delta_1 < 20$	$\tau_i(t)$>10	1	150≤ $v_i(t)$ <200
S5	Variable or partial PPE	δ_1 < 20 px compared to δ_2	$\tau_i(t) < 5$	2 - 4	200 ≤ $v_i(t)$<250
S6	Partial PPE	50≤ $\delta_1 < 20$	5.≤ $\tau_i(t) < 10$	5	$v_i(t)$>250

S1 considered a low risk condition, characterized by full PPE, moderate distance to the ROI, low severity, short dwell, low density, and slow movement. In S2, risk was increased by introducing partial PPE, greater proximity to the ROI, medium severity, intermediate dwell, and moderate speed. In S3 and S4, the effect of the absence of critical PPE inside the ROI was evaluated, differentiating both cases by zone severity and dwell time; thus, S3 represents a high condition, while S4 corresponds to a critical condition due to the combination of lack of protection, direct occupation of the hazardous zone, high severity, and prolonged dwell. S5 was designed to analyze cases of crossing or close spatial overlap between regions, where homography must avoid undue activations or overestimations of risk. Finally, S6 made it possible to evaluate the effect of the operational context by combining partial PPE, proximity to the ROI, greater density of people, and high speed. Together, this configuration made it possible to verify whether the fuzzy system responds gradually and consistently to increases in exposure, severity, and dynamic complexity of the environment.

4.4 Fuzzy system response per scenario

For each scenario, the average values per person and per frame were computed (Table 12), then aggregated per sequence and finally averaged over the five corresponding recordings. The results show behavior consistent with the logic of the proposed system.

Table 12. Average response of the fuzzy system in all videos of all controlled scenarios and their standard deviation

Scenario	$P_i(t)$	$E_i(t)$	$K_i(t)$	$r_i(t)$	$\bar{r}_i(t)$
S1	0.25±0.04	0.3±0.05	0.3±0.03	28±4	25±3
S2	0.60 ±0.06	0.58±0.02	0.20±0.03	55±3	50±5
S3	0.88±0.02	0.78±0.03	0.38±0.01	79±2	73±2
S4	0.92±0.01	0.93±0.02	0.55±0.04	92±1	90±2
S5	0.5±0.07	0.28±0.06	0.35±0.04	40±3	36±3
S6	0.58±0.05	0.6±0.05	0.86±0.03	78±2	80±2

In addition to the quantitative results obtained by the system, representative screenshots were selected in order to visually document the experimental conditions considered in each scenario. These images are not used as the main performance metric, but as qualitative evidence of the real test context.

In scenario S1, where the person has full PPE, is far from the risk regions, and there is low operational density, the three intermediate outputs remain at reduced values, generating a low risk level (Figure 9). This result confirms that the system does not unnecessarily increase the risk score when the visual, spatial, and contextual evidence does not indicate a hazardous condition.

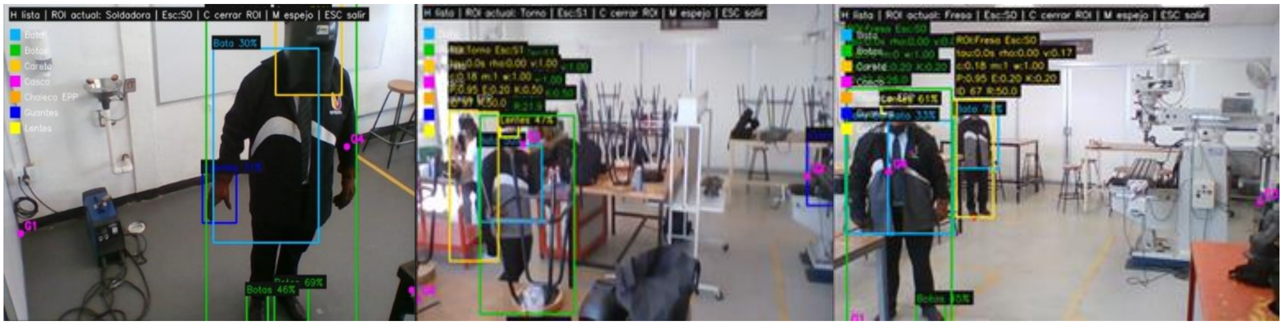

Fig. 9. Experimental evidence corresponding to Scenario S1 in the welding station (left), lathe station (center) and milling station (right). The images illustrate a baseline operating condition characterized by safe distances from hazardous regions and/or full PPE equipment, while standing still or walking slowly. Under these conditions, workers exhibit low risk levels.

In scenario S2, the protection deficit increases due to the partial presence of PPE, while exposure also increases due to proximity to the ROI and a medium dwell time (Figure 10). However, the operational context remains low, so the final risk is at a moderate level.

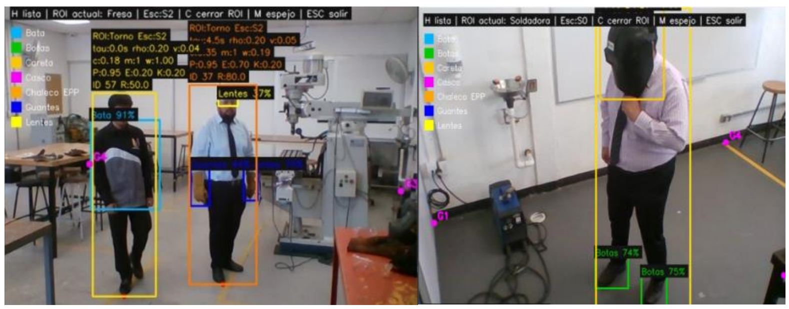

Fig. 10. Experimental evidence corresponding to Scenario S2 in the milling station (left) and welding station (right). The operators approach more to the ROI while the system automatically detects missing personal protective equipment (PPE) that are not critical, which combined with the reduced distance between workers and hazardous areas increases exposure levels, resulting in higher risk estimates.

Scenarios S3 and S4 reveal the system's sensitivity to the absence of critical PPE. In S3, the lack of a critical element inside a ROI produces a considerable increase in protection deficit and exposure, generating a high risk (Figure 11). In S4, the combination of critical absence, high zone severity, and prolonged dwell raises the index to a critical level.

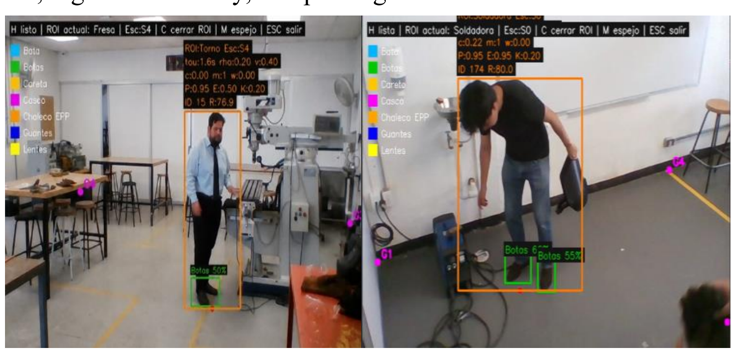

Fig. 11. Experimental evidence corresponding to Scenario S3 in the milling station (left),and Scenario S4 in welding station (right). The operators remain within the ROI for extended periods with some critical PPE equipment missing, resulting in increased exposure to hazardous areas and low PPE index.

Scenario S5 allows the effect of geometric correction to be observed. Although the person may visually overlap with a risk zone while crossing the aisle (Figure 12), the evaluation on the top-down plane reduces the computed exposure, keeping the risk between low and moderate. In addition, the inclusion of a greater density of people increases the risk, also due to the speed at which they cross the risk zone.

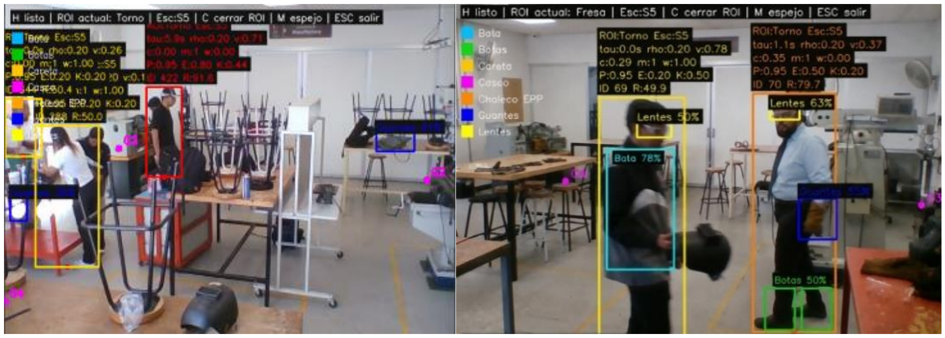

Fig. 12. Experimental evidence corresponding to Scenario S5 in the lathe station (left) and milling station (right). Multiple operators perform simultaneous activities within the monitored environment while having some or none PPE on. Differences in worker positions, proximity to hazardous areas, and velocity distinct individual risk levels can be seen.

Finally, scenario S6 shows that the operational context can increase risk even when the protection deficit is not extreme (Figure 13). The presence of high local density and rapid movement increases ($K_i(t)$), which raises the final index toward a high level.

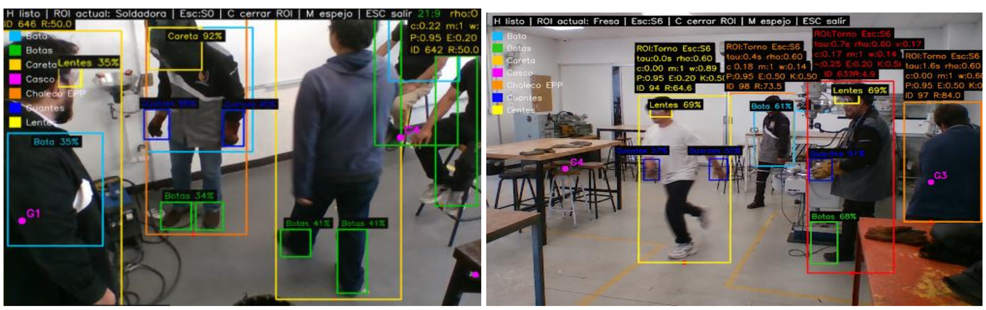

Fig. 13. Experimental evidence corresponding to Scenario S6 in the welding station (left)and milling station (right). The images illustrate a complex operational scenario involving multiple workers simultaneously performing activities or moving quickly within the monitored environment.

Although the experimental validation was carried out with recorded sequences and controlled scenarios, the proposed architecture was designed as a frame-by-frame processing pipeline.

The system estimates risk indicators from each frame through visual, spatial, and temporal evidence, and these indicators can be updated as new frames arrive. In this sense, the proposal should be understood as a safety supervision support tool compatible with near real time monitoring, capable of providing early risk indications without losing usefulness for auditing or retrospective analysis.

In 2024 alone, Mexico's Secretariat of Labor and Social Welfare (STPS) has recorded more than 4,500 accidents linked to the failure to use PPE (Secretaría del Trabajo y Previsión Social [STPS], 2026). For that reasons it is important to note that the proposed system does not intend to authorize the operation of machinery under unsafe conditions, nor to replace institutional safety protocols. In industrial and educational manufacturing settings, the use of the required PPE must be verified before the start of the activity (Secretaría del Trabajo y Previsión Social [STPS], 2024). For this reason, the scenarios with partial PPE or absence of critical elements were considered only as controlled experimental conditions to evaluate the model's sensitivity to different levels of non-compliance and exposure, not as acceptable operational practices.

Likewise, occupational risk is a multifactorial phenomenon that cannot be fully represented through visual evidence. Factors such as the maintenance condition of the machinery, operator experience, environmental conditions, ergonomic load, work organization, and human factors also influence safety.

In this way, the generated score should not be understood as a comprehensive occupational assessment that replaces audits, regulations, or human supervision (Parlamento Europeo y Consejo de la Unión Europea, 2024), but as a continuous, interpretable, and traceable indicator of operational risk to support supervision, training, event documentation, and safety decision-making.

This interpretation is consistent with recent reviews on AI applied to occupational safety, which stress that these systems can support monitoring, early detection, and user assistance, but require ethical integration, a human centered approach, risk assessment, transparency, and human supervision to avoid over-reliance or inappropriate uses of automation (Huber et al., 2025; Baraza, 2026).

5. Conclusions.

This work presented a hierarchical system for continuous scoring of operational risk in a university manufacturing laboratory, integrating computer vision, regions of interest, geometric correction through homography, and zero order Sugeno-Takagi fuzzy inference. The proposal is based on the premise that risk should not be interpreted solely as a binary condition of PPE compliance or non-compliance, but as a gradual, spatial, and context dependent condition.

The results show that the visual perception module provides a functional basis to feed the fuzzy system. The YOLOv11 detector achieved favorable overall performance in identifying PPE, although differences among classes were observed. Elements such as glasses, face shield, and boots showed better results, while lab coat and gloves showed greater sensitivity to occlusions, reduced apparent size, posture variations, and lighting conditions. This behavior confirms that the quality of the perception module directly influences the computation of the protection deficit and, therefore, the final risk index.

The geometric evaluation showed that homography improves spatial consistency in the activation of ROIs. By projecting each person's support point onto an approximate top-down plane, the system reduced errors associated with camera perspective, especially in situations where a person visually overlaps with a machine or hazardous region without physically occupying that zone. This correction is relevant because the proximity and dwell variables depend on a consistent spatial estimation.

The response of the fuzzy system across the six controlled scenarios showed gradual behavior consistent with the experimental conditions. In low risk scenarios, characterized by full PPE, low exposure, and reduced operational context, the final index remained at low levels. In intermediate conditions, such as partial PPE or greater proximity to a ROI, the system produced moderate values. When there was an absence of critical PPE inside a risk zone, high severity, or prolonged dwell, the index increased toward high or critical levels. Likewise, the scenarios with greater local density and rapid movement showed that the operational context can increase risk even when the protection deficit is not extreme.

Taken together, the results indicate that the proposed hierarchical architecture makes it possible to decompose risk estimation into interpretable dimensions: protection deficit, hazard exposure, and operational context. This organization facilitates identifying whether an alert relates mainly to a lack of PPE, proximity to a hazardous zone, prolonged dwell, high density of people, or intense movement. In addition, temporal smoothing helps reduce oscillations caused by visual noise, momentary false negatives, or partial occlusions.

Nevertheless, the results should be interpreted as a pilot validation in a controlled environment. The system was evaluated with a specific camera configuration, calibrated ROIs, and a dataset adjusted to the laboratory, so its performance may vary under changes in lighting, spatial layout, type of machinery, clothing, or safety protocols. Likewise, although the architecture processes information frame by frame and is compatible with near real time monitoring, prolonged tests under real operating conditions, latency measurements, and evaluation of false positives and false negatives are still required before considering it a fully deployed solution.

As future work, it is proposed to expand and balance the PPE dataset, incorporate more examples with occlusions and lighting variations, strengthen multi-object tracking, improve geometric calibration, and compare the system's behavior with Mamdani, ANFIS, or type-2 fuzzy variants. It is also envisaged to develop an automatic reporting module that records relevant events, risk values, and activated rules, in order to support auditing, training, and continuous improvement of safety protocols.

References

Ahmed, T., Hoque, A. S. M., Karmaker, C. L., & Ahmed, S. (2023). Integrated approach for occupational health and safety (OHS) risk assessment: An empirical (case) study in small enterprises. *Safety Science, 164*, Article 106143. https://doi.org/10.1016/j.ssci.2023.106143

Al-Azani, S., Luqman, H., Alfarraj, M., Sidig, A. A. I., Khan, A. H., & Al-Hamed, D. (2024). Real-time monitoring of personal protective equipment compliance in surveillance cameras. IEEE Access, 12, 121882–121895. https://doi.org/10.1109/ACCESS.2024.3451117

Ayoubi, M., & Arashpour, M. (2026). Enhancing workplace safety through assistive computer vision: Real-time hazard recognition using the Workplace Hazards Dataset (WHD). *Computer Vision and Image Understanding, 265*, Article 104681. https://doi.org/10.1016/j.cviu.2026.104681

Aziz, A., Suzon, M. M., & Hasan, R. (2025). A fuzzy logic-based risk evaluation and precaution level estimation of explosive, flammable, and toxic chemicals for preventing damages. Heliyon, 11(1), e41216. https://doi.org/10.1016/j.heliyon.2024.e41216

Baraza, X., & Torrent-Sellens, J. (2026). Artificial intelligence and emerging risks in occupational safety and health. *Encyclopedia, 6*(1), Article 25. https://doi.org/10.3390/encyclopedia6010025

Dodoo, J. E., Al-Samarraie, H., Alzahrani, A. I., Lonsdale, M., & Alalwan, N. (2024). Digital innovations for occupational safety: Empowering workers in hazardous environments. *Workplace Health & Safety, 72*(3), 84–95. https://doi.org/10.1177/21650799231215811

Gürcanli, G. E., & Müngen, U. (2009). An occupational safety risk analysis method at construction sites using fuzzy sets. *International Journal of Industrial Ergonomics, 39*(2), 371–387. https://doi.org/10.1016/j.ergon.2008.10.006

Huber, J., Anzengruber-Tanase, B., Schobesberger, M., Haslgrübler, M., Fischer-Schwarz, R., & Ferscha, A. (2025). Evaluating user safety aspects of AI-based systems in industrial occupational safety: A critical review of research literature. *International Journal of Environmental Research and Public Health, 22*(5), Article 705. https://doi.org/10.3390/ijerph22050705

Iannizzotto, G., Lo Bello, L., & Patti, G. (2021). Personal protection equipment detection system for embedded devices based on DNN and fuzzy logic. *Expert Systems with Applications, 184*, Article 115447. https://doi.org/10.1016/j.eswa.2021.115447

Ilbahar, E., Karaşan, A., Cebi, S., & Kahraman, C. (2018). A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP & fuzzy inference system. *Safety Science, 103*, 124–136. https://doi.org/10.1016/j.ssci.2017.10.025

Kim, H., Kim, K., & Kim, H. (2016). Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects. *Journal of Computing in Civil Engineering, 30*(4), Article 04015075. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562

Li, L., Huang, Z., Wang, J., Du, B., & Dai, L. (2026). Automated construction monitoring based on computer vision: A comprehensive review. *Developments in the Built Environment, 25*, Article 100832. https://doi.org/10.1016/j.dibe.2025.100832

López, L., Suárez-Ramírez, J., Alemán-Flores, M., & Monzón, N. (2025). Automated PPE compliance monitoring in industrial environments using deep learning-based detection and pose estimation. *Automation in Construction, 176*, Article 106231. https://doi.org/10.1016/j.autcon.2025.106231

López-Cifuentes, A., Escudero-Viñolo, M., Bescós, J., & Carballeira, P. (2022). Semantic-driven multi-camera pedestrian detection. *Knowledge and Information Systems, 64*(5), 1211–1237. https://doi.org/10.1007/s10115-022-01673-w

Martinez, C. F., Galvan, F. S., Santos, H. B., Estrada, O. V., De La Cruz, D. I. M., & Ponce del Angel, F. G. (2026). Real-time ergonomic assessment with fuzzy logic and RGB-D sensors. *Engineering, Technology & Applied Science Research, 16*(1), 31609–31617. https://doi.org/10.48084/etasr.14044

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). (2024). *Diario Oficial de la Unión Europea, L*, 2024/1689. EUR-Lex official text

Rasouli, S., Alipouri, Y., & Chamanzad, S. (2024). Smart personal protective equipment (PPE) for construction safety: A literature review. *Safety Science, 170*, Article 106368. https://doi.org/10.1016/j.ssci.2023.106368

Sadeghi, H., Mohandes, S. R., Hosseini, M. R., Banihashemi, S., Mahdiyar, A., & Abdullah, A. (2020). Developing an ensemble predictive safety risk assessment model: Case of Malaysian construction projects. *International Journal of Environmental Research and Public Health, 17*(22), Article 8395. https://doi.org/10.3390/ijerph17228395

Salhab, D., Sabek, M., Pourrahimian, E., Gonzalez, V. A., & Hamzeh, F. (2026). Integrated framework of computer vision and fuzzy systems for lean workspace management in construction. *Automation in Construction, 181*, Article 106618. https://doi.org/10.1016/j.autcon.2025.106618

Secretaría del Trabajo y Previsión Social. (2025, March 28). *Norma Oficial Mexicana NOM-017-STPS-2024, Equipo de protección personal-Selección, uso y manejo en los centros de trabajo*. *Diario Oficial de la Federación*. Official DOF record

Secretaría del Trabajo y Previsión Social. (n.d.). *Riesgos de trabajo registrados en el IMSS* [Interactive data portal]. Retrieved August 29, 2026, from STPS occupational-risk statistics portal

Shin, Y.-S., & Kim, J. (2022). A vision-based collision monitoring system for proximity of construction workers to trucks enhanced by posture-dependent perception and truck bodies' occupied space. *Sustainability, 14*(13), Article 7934. https://doi.org/10.3390/su14137934

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15*(1), 116–132. https://doi.org/10.1109/TSMC.1985.6313399

Toth, K., Singhal, M., Chauhan, D., Polra, J., Zhou, C., Page, G., & Fisher, C. (2026). A dual-model approach to industrial safety: Computer vision for PPE compliance and hazard-zone monitoring in steel production. *Integrating Materials and Manufacturing Innovation, 15*, 326–340. https://doi.org/10.1007/s40192-026-00461-6

Tran, S. V.-T., Pham, H. C., Le, Q. T., & Lee, U.-K. (2025). Digital technologies for monitoring hazardous area entry in construction sites. *Automation in Construction, 177*, Article 106357. https://doi.org/10.1016/j.autcon.2025.106357

Vukicevic, A. M., Petrovic, M., Milosevic, P., Peulic, A., Jovanovic, K., & Novakovic, A. (2024). A systematic review of computer vision-based personal protective equipment compliance in industry practice: Advancements, challenges and future directions. *Artificial Intelligence Review, 57*, Article 319. https://doi.org/10.1007/s10462-024-10978-x

Zhang, M., Cao, Z., Yang, Z., & Zhao, X. (2020). Utilizing computer vision and fuzzy inference to evaluate level of collision safety for workers and equipment in a dynamic environment. *Journal of Construction Engineering and Management, 146*(6), Article 04020051. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001802