Artificial Intelligence System for Detecting the Use of Personal Protective Equipment

Josue Airton Lopez Cabrejos[1]*, Avid Roman-Gonzalez[1,2]
Aerospace Sciences & Health Research Laboratory (INCAS-Lab), Universidad Nacional Tecnológica de Lima Sur, Lima, Perú[1, 2]
Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Lima, Perú[2]

***Abstract*—In recent years, occupational accidents have been increasing, and it has been suggested that this increase is related to poor or no supervision of personal protective equipment (PPE) use. This study proposes developing a system capable of identifying the use of PPEs using artificial intelligence through a neural network called YOLO. The results obtained from the development of the system suggest that automatic recognition of PPEs using artificial intelligence is possible with high precision. The recognition of gloves is the only critical object that can give false positives, but it can be addressed with a redundant system that performs two or more consecutive recognitions. This study also involved the preparation of a custom dataset for training the YOLO neural network. The dataset includes images of workers wearing different types of PPEs, such as helmets, gloves, and safety shoes. The system was trained using this dataset and achieved a precision of 98.13% and a recall of 86.78%. The high precision and recall values indicate that the system can accurately identify the use of PPEs in real-world scenarios, which can help prevent occupational accidents and ensure worker safety.**

Keywords—Personal protective equipment (PPE); artificial intelligence (AI); YOLO (You Only Look Once); object detection; neural network; custom PPE dataset

I. Introduction

Workplace accidents pose a serious threat to the safety and well-being of workers worldwide. The human cost of such accidents is immeasurable, with the victims and their families often enduring long-term physical, emotional, and financial consequences. Besides, accidents at work can also have significant economic impacts, such as lost productivity, increased healthcare costs, and legal liabilities for employers. Thus, preventing and mitigating the risks of occupational accidents should be a top priority for governments, organizations, and individuals alike.

Unfortunately, recent statistics indicate that the frequency of occupational accidents has been on the rise, leading to an increase in injuries and fatalities. For instance, the Ministry of Labor and Employment Promotion reported a staggering 23.7% increase in accidents in April 2022 [1] compared to the same period the previous year. The report showed that Lima had the highest number of accidents, with 2,132 cases, of which six were fatal. The most common types of accidents were due to object strikes and falls, with tools being the leading cause. Fingers and eyes were the most affected body parts.

In a recent study, it was determined that, on average, 87% of workers do not use PPE properly, linking the lack of PPE usage to both fatal and non-fatal accidents [2]. Furthermore, the use of PPE is a legal right for workers, who are obligated to protect their health and lives [3].

To address this alarming trend, researchers and practitioners have been exploring various solutions to prevent or mitigate the risks of occupational accidents. One promising approach is the use of artificial intelligence systems to detect the use of personal protective equipment in work environments. PPEs are essential safety devices that protect workers from hazards, such as falling objects, sharp tools, chemicals, or radiation. However, the effectiveness of PPEs relies on their proper use and maintenance, which is only sometimes guaranteed in practice.

AI-based systems can help monitor and enforce the proper use of PPEs in real-time, thus reducing the risk of accidents and injuries. However, an adequate and relevant dataset is a significant challenge in developing such systems. The performance of AI algorithms depends heavily on the quality and quantity of data used to train them. Collecting and labeling large and diverse datasets of workers wearing different types of PPEs in various work environments can be costly and time-consuming, so a pre-trained neural network such as YOLO can help reduce time in training and deploying a model [4].

Researchers had to create a step-by-step dataset using internet images to overcome this limitation, using various techniques, such as web scraping, data augmentation, and transfer learning, to generate a large and diverse dataset of workers wearing different types of PPEs in various work environments. Researchers must ensure a balanced and representative dataset of the real-world distribution of PPEs and work environments.

With a custom dataset of images obtained from github's user [5], an innovative AI-based system was developed that uses advanced image processing techniques, specifically convolutional neural networks and object detection, to detect the use of PPEs in work environments. The system can recognize various types of PPEs, such as hard hats, safety glasses, vests, shoes, and gloves, and their proper use and placement on the worker's body.

One of the advantages of this system is its non-invasiveness. The system can be used on cameras installed in the work environment to capture images of workers wearing PPEs, without requiring physical contact or interference with their daily activities [6]. This feature makes the system more acceptable and practical for workers and supervisors, as it does not disrupt their workflow or privacy.

In this paper, a review was conducted to determine which neural network would perform most efficiently. Section III details how a neural network can be trained in the cloud, thus avoiding the need for powerful hardware. Section IV presents the results regarding precision and recall to provide a comprehensive understanding of the system's performance. Finally, Section V shows the conclusions about the findings.

II. Literature Review

"Design of an artificial vision system to determine the quality of mandarins" is a thesis about of the development of an algorithm that classifies mandarins based on their size, shape, and colour using digital image processing and region-based segmentation through a webcam using MATLAB and Arduino, achieving an accuracy of 93.3% in classification[7].

Another thesis mentions that techniques such as You Only Look Once, Region proposals + CNN, Single Shot Detector, among others, can be used to detect objects based on computer vision, and these techniques are associated with deep learning. The advantages and disadvantages of each technique mentioned above are also discussed [8].

It is possible to classify lemons, using artificial vision, based on their shape and dimensions. The development of the algorithm follows the phases of an artificial vision solution acquisition, pre-processing, segmentation, description and recognition and interpretation, and the efficiency achieved by the system is 83.9% in "Development of an artificial vision system to perform a uniform classification of lemons" [9].

An article named "Real-time Personal Protective Equipment Detection Using YOLOv4 and TensorFlow" consists of developing a mask and face shield detector as preventive measures for COVID-19 in real-time using YOLOv4, obtaining a system effectiveness of 79% [10].

A thesis called "Electronic system for quality control of chicken eggs using image processing" from the Universidad Técnica de Ambato (Ecuador), consists in a quality control system for chicken eggs using image processing in Python with the OpenCV library, concluding that two factors mainly determine the efficiency of the system, the software used and the ambient lighting [11].

"Design of a classifier system for apples by colour, using artificial vision, for the company Fresh & Natural C.I." from Andrea Aguilar, developed an apple classifier using artificial vision, capturing images from a webcam and using MATLAB as a processing interface, obtaining an effectiveness of 100% for a number of 18 apples [12].

III. Proposed System

The following is a general description of the network's operation, an explanation of the dataset used, the criteria used to select the neural network, the hardware and software used during this work, and considerations that must be considered.

A. System Overview

As illustrated in Fig. 1, the system is founded upon the state-of-the-art Yolov5s architecture to train a neural network that can accurately identify personal protective equipment in real time. To ensure optimal performance, a custom dataset is prepared, employing detailed image segmentation and data augmentation techniques to triple the number of images. The trained model is then deployed to evaluate frames captured by a webcam, seamlessly detecting PPEs worn by individuals in real time.

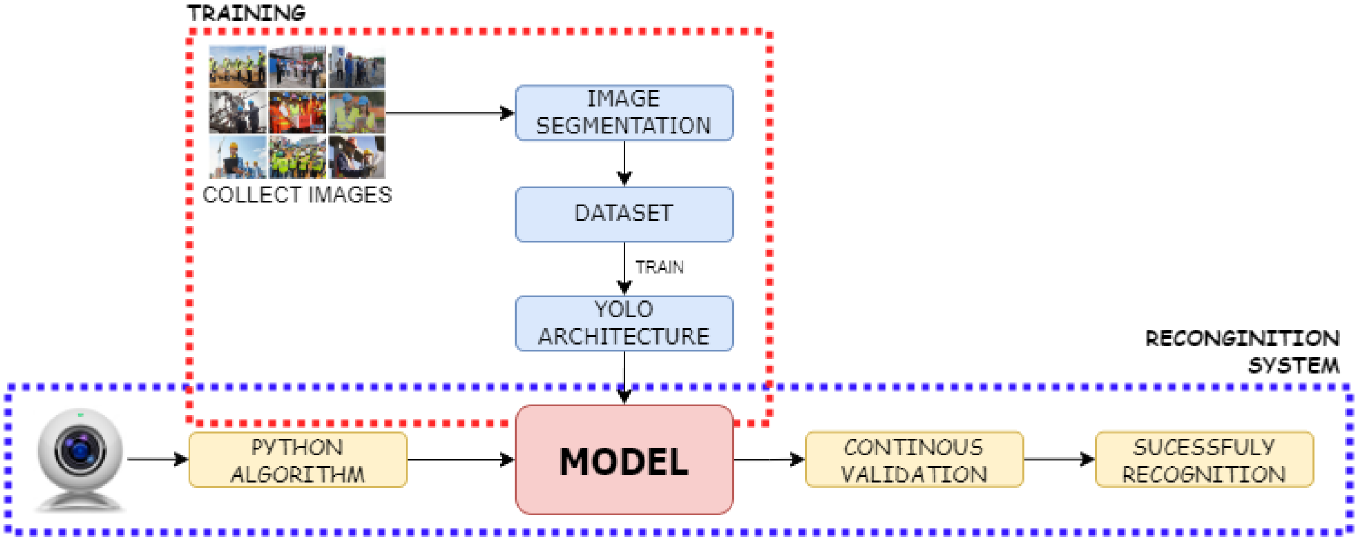

Fig. 1. Flowchart proposed system.

B. EPP Dataset

In this study, the classes to be recognized in the system were defined according to the Peruvian technical health standard for personal protective equipment, which is a ministerial resolution [13]. The classes were also defined based on the body parts involved in accidents. The classes to be detected were as follows: helmet, vest, goggles, gloves, shoes, and the person itself, resulting in six classes. This classification system was crucial for the training and testing of the deep learning model used in this study, as it allowed for accurate and reliable detection of the relevant personal protective equipment in real-time; in Fig. 2, the total object per class in the entire dataset is represented.

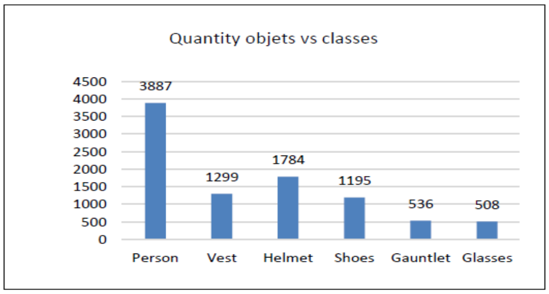

Fig. 2. Quantity objects per class.

C. Software and Hardware Seleccion

After a long comparative investigation of different open-access neural networks[14], it was chosen to use the YOLO v5 neural network, because it is easy to train, it can be used directly with video and webcam source and is lightweight than older YOLO's version. But in YOLOv5 there are some variants, because of that, two parameters were considered for selecting the neural network variant: the accuracy concerning a dataset called Common Objects in Context (COCO) and the processing speed in images per unit of time [15].

Below, in Table I, we show the different YOLOv5 variants with the established parameter data for comparison.

With a simple calculation, we can obtain ratings for each of the YOLO variants, shown in the following Table II, where a lower rating indicates a better relationship between processing time and system accuracy.

It is observed that v5n is the best option for a system where an optimal relationship between processing time and accuracy is desired, but in the present case, one used v5s because one wanted more accuracy, sacrificing some performance.

TABLE I. PROCESS TIME PER IMAGE YOLO

YOLO variant	COCO accuracy (%)	Process time (ms/image)
v5n	27.6	62.7
v5s	37.6	173.3
v5m,	45.0	427.0
v5l	49.0	857.4
v5x	50.7	1579.2

TABLE II. PROCESS TIME PER ACCURACY YOLO

YOLO variant	Process time (ms) / accuracy
v5n	2.27
v5s	4.61
v5m,	9.49
v5l	17.50
v5x	31.15

Python 3.9.7 was used for training and implementing the neural network, as it has active community support and is compatible with a wide range of existing Python libraries. A cloud service called Google Colab is used for training due to the intensive GPU usage required, which can be expensive. Google Colab offers free access to GPUs for a few hours per week, and it also allows users to connect with Google Drive to save progress and resume when more hours are available on Google Colab, all Software and Hardware used is shown in Table III.

TABLE III. EXPERIMENTAL SETUP

Hardware / Software	Description
Lenovo T430 i5 3320M	Computer specification
Microsoft Windows 10 Pro	Operating system
Google Colab	Web App for Neural Network training
Python 3.9.7	Python's version used
Roboflow	Web App for dataset creation
Generic Webcam	1080p 60fps

D. PPE Detector System

The YOLO source code has been analyzed, and it has been determined that the input image is resized to a resolution of 240 x 240 pixels, as per the information gleaned from the neural network's source code. Moreover, it has been observed that YOLO employs a 6x6 initial kernel, and uses a 3x3 kernel in the deeper layers, every kernel means a convolution and every convolution makes the image smaller. In the final stage of training, the layers were modified to predict six classes that align with the prepared dataset. A detailed breakdown of the neural network layers can be found in Fig. 3, providing further insights into its inner workings.

Conv means a convolutional layer, C3 means three convolutional layers with same parameters, Up Sampling is used to resize the image to its original size, NMS is to keep the bounding box of each class with the highest confidence score.

Using YOLOv5s, training the neural network can be accomplished through simple steps. Firstly, the YOLOv5 GitHub repository must be cloned. Next, the necessary libraries located in requirements.txt must be installed. Finally, the training steps can be followed using the following arguments: --data data.yaml --epochs 200 --weights yolov5s.pt --batch-size 40.

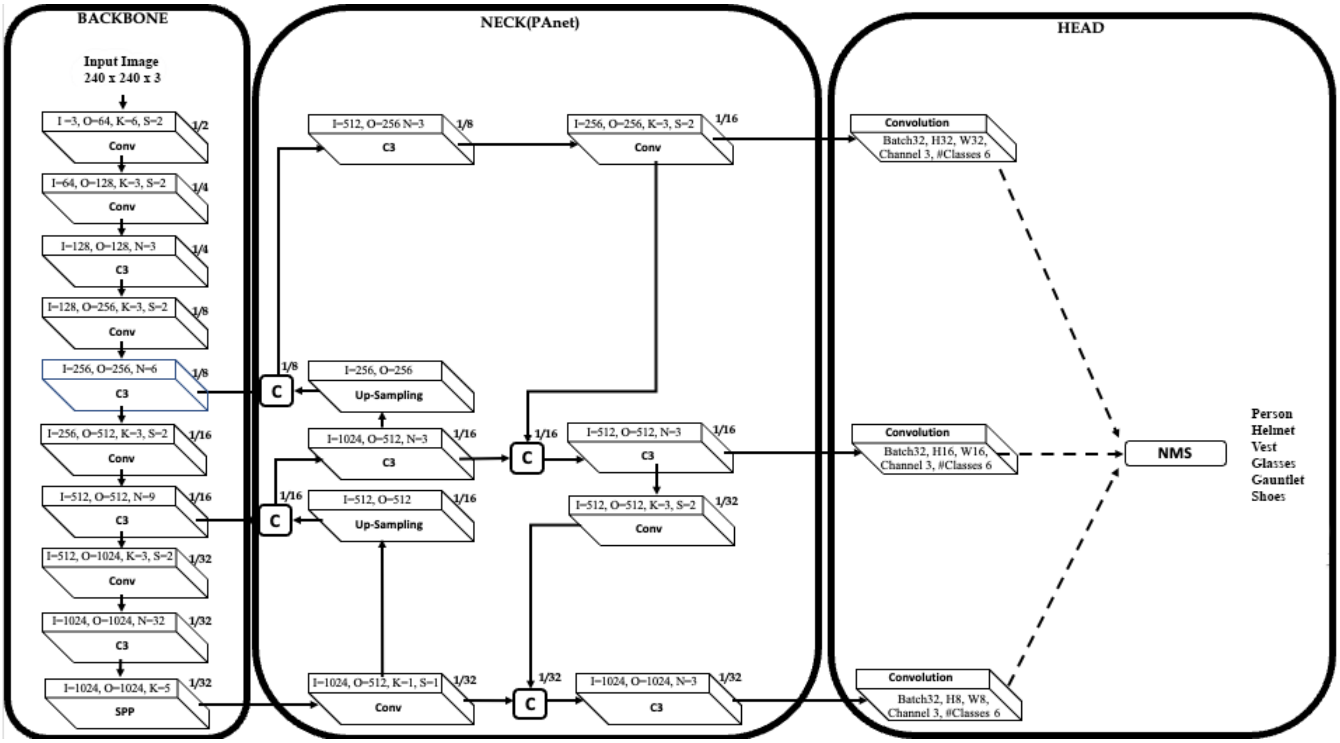

Fig. 3. YOLOv5s architecture.

IV. Results and Evaluation

The results of a neural network system can be divided into two categories: training results and recognition results. Training results are observed during the training process and serve as an indicator of whether the neural network is performing well or not. Recognition results are custom metrics that allow us to evaluate the performance of the neural network.

A. *Training Results*

The system's accuracy and losses should be examined first when training a neural network. While these are not directly related, they provide insight into the system's overall behaviour shown in Fig. 4. System's global accuracy is 80%, or 0.8, this can be interpreted as image recognition, but since the system will be used in video, in section C, experimental results are shown, due multiple images in a second can be processed in video. This situation indicates that many PPEs are being recognized. On the other hand, losses are less than 10%. This result suggests that the system only misclassifies a small fraction of the PPEs it encounters, and confusion with other categories is minimal. While this indicates good system performance, it should be noted that these are general observations.

B. *Evaluation Metrics*

The widely adopted mean average precision will be used (mAP) metric to evaluate the object detection model. The mAP[16] metric is a comprehensive measure of the effectiveness of object detection models, which considers both the precision (1) and recall (2) of the model across different levels of confidence. It is calculated by averaging the precision values at different levels of recall, where precision is the proportion of correctly classified objects among all objects classified at a certain level of confidence, and recall is the proportion of correctly classified objects among all objects that should have been classified:

$$\text{Precision} = TP / (TP + FP) \qquad (1)$$

$$\text{Recall} = TP / (TP + FN) \qquad (2)$$

Here, *TP* represents the number of true positives, which indicates the number of PPE objects correctly detected and classified by the model. *FP* represents the number of false positives, corresponding to the number of non-PPE objects incorrectly classified as PPE objects. *FN* represents the number of false negatives, which indicates the number of PPE objects that have been missed by the model and not detected.

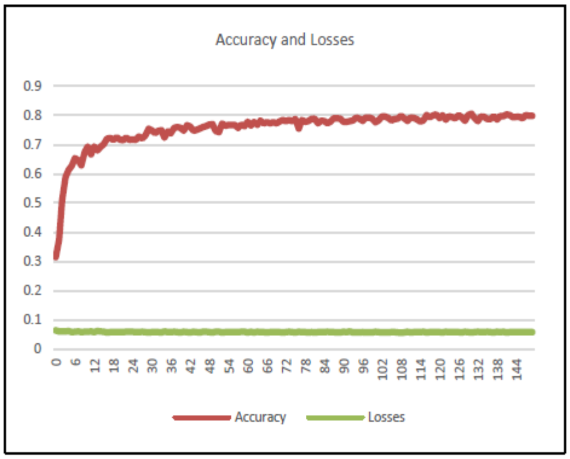

Fig. 4. Accuracy and losses of the proposed system per epoch trained.

The mAP metric is a comprehensive measure of the effectiveness of object detection models, which considers both the precision and recall of the model across different levels of confidence. It is calculated by averaging the precision values at different levels of recall, where precision is the proportion of correctly classified objects among all objects classified at a certain level of confidence, and recall is the proportion of correctly classified objects among all objects that should have been classified.

The mAP metric is widely used in many applications, including workplace safety monitoring and quality control in manufacturing, where the reliable detection and classification of PPE objects is crucial for ensuring worker safety and product quality. By utilizing the mAP metric, we can obtain a comprehensive and quantitative assessment of the performance of the object detection model, which is essential for ensuring the effectiveness and reliability of the system in real-world applications.

C. Result Analysis

Using a threshold of 0.5 for each class identification, using 1080p webcam input source and 600 x 400 output video, the object detection model achieved a precision of 0.98130841 and a recall of 0.8677686 in identifying PPE objects, including helmets, gloves, vests, shoes, safety glasses, and people. The precision value indicates that the model accurately classified 98.13% of the detected PPE objects, while the recall value indicates that 86.78% of all PPE objects in the images were successfully detected and classified. The model's ability to identify multiple types of PPE objects suggests various features for object recognition, although gloves were the most challenging object to recognize. A frame is processed in around 0.075 seconds that means 13 frames per second using YOLO V5s (refer Fig. 5).

The closest better version of YOLO is 2x slower according to Table II, it means 7 frames per second, it wouldn't have been optimal for real-time applications.

These results suggest that the model performs well in identifying PPE objects, but further optimization may be needed to improve its accuracy in detecting and classifying gloves. Monitoring and evaluating the model's performance will be necessary for ensuring its reliability and effectiveness in real-world applications.

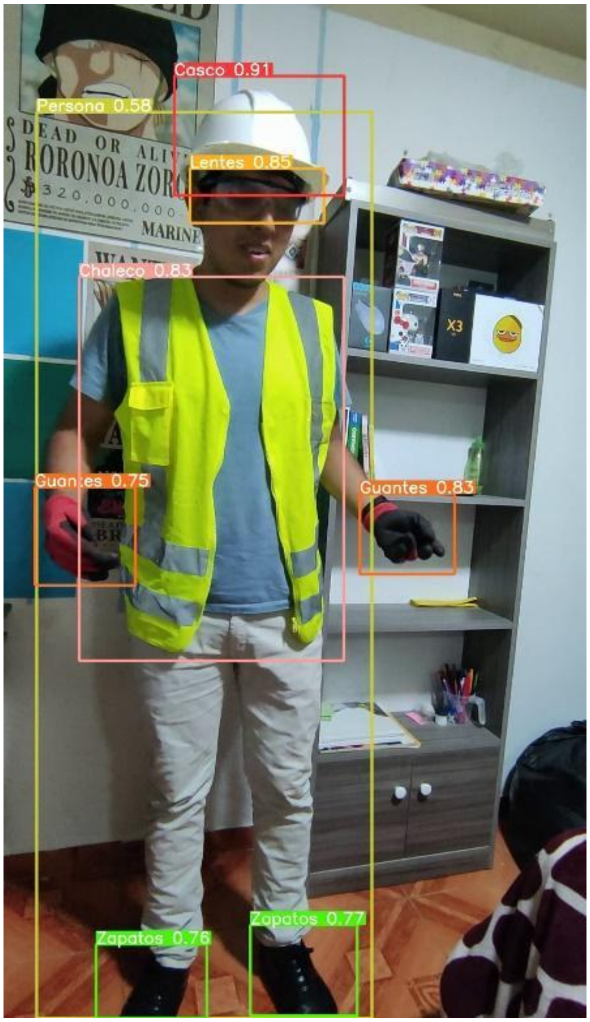

Fig. 5. Result of system's prediction.

V. Discussion

To improve the performance of the model, it is recommended to experiment with different threshold values for detecting personal protective equipment, as this can help fine-tune the algorithm for optimal sensitivity and specificity. Additionally, using metrics such as precision and recall can provide a more comprehensive evaluation of the model's performance and highlight areas for improvement. Expanding the dataset to include more diverse and challenging scenarios can enhance the model's robustness and generalizability. By implementing these recommendations, the model can be optimized for more accurate and reliable detection of personal protective equipment in various real-world applications.

VI. Conclusion

The YOLOv5s neural network is lightweight enough to be used in real-time scenarios, as demonstrated by its successful performance on relatively old hardware while achieving efficient results. However, to further improve the present work, it is suggested to have a better distribution of classes in the dataset, with roughly the same number of objects per class, to enhance recall, especially for challenging objects.

Using CNN with a video source or webcam offers advantages compared to its application on static images. This is because it is not necessary to recognize the object immediately or in the first video frame. Instead, the object can be recognized within a specific period, even if it only appears in a few frames, and it will still be considered valid. This temporal recognition capability allows for capturing and recognizing objects in a video sequence, which is beneficial for applications such as object tracking or real-time object analysis. By leveraging the contextual information provided by consecutive frames, the CNN can improve recognition accuracy and provide more robust results in dynamic environments.

By augmenting the dataset, the model is exposed to a greater variety of images, which can improve its ability to generalize to new, unseen data. When working with limited data, these results in a better-trained model with higher precision and lower losses are often necessary. Furthermore, data augmentation can also help to reduce overfitting, a common problem in deep learning where the model becomes too specialized to the training data and performs poorly on new data. Therefore, performing data augmentation on the dataset is crucial in developing robust and accurate computer vision models.

References

[1] MTPE, “ Notifications of work accidents, hazardous incidents, and occupational diseases”, OGETIC, pp 5-9, April 2022.

[2] S. Macalopu and S. Guzman. “Work accidents and personal protective equipment among public cleaning workers in the José Leonardo Ortiz district”, ACC CIETNA: Journal of the School of Nursing, vol 1(2), pp. 14-23, https://doi.org/10.35383/cietna.v1i2.153.

[3] El Peruano. “The Occupational Safety and Health Act”, pp 1-13, August 2011.

[4] F. Joiya, “Obtec detection: yolo vs faster r-ccn”, IRJMETS, vol 4, pp. 1-5, September 2022.

[5] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, “Fast personal protective equipment detection for real construction sites using deep learning approaches”, Sensors, vol 21, 3478, May 2021.

[6] S. Barro, T. Fernandez, H. Perez and C. Escudero. “Real-time personal protective equipment monitoring system”, Computer Communications, chapter 36, pp 42-50, January 2012.

[7] A. Gramoral, “Design of an artificial vision system to determine the quality of mandarins”, Lima: UTP, 2020, pp. 219-223.

[8] L. Machaca, “Recognition of anomalous events in videos obtained from surveillance cameras using convolutional networks”, Arequipa: UNSA ,2019 ,pp. 45-53.

[9] E. Castillo, “Development of an artificial vision system to perform a uniform classification of lemons”, Lima: UPN, 2018, pp 46-52-

[10] A. A. Protik, A. H. Rafi and S. Siddique, "Real-time personal protective equipment (PPE) detection using YOLOv4 and tensorFlow," 2021 IEEE Region 10 Symposium (TENSYMP), Jeju, Korea, Republic of, 2021, pp. 1-6, doi: 10.1109/TENSYMP52854.2021.9550808

[11] M. Jurado and A. Fernandez, “Electronic system for quality control of chicken eggs through image processing”, Ambato: UTA, 2018, pp. 52-56.

[12] A. Aguilar, “Design of a color-based apple sorting system using artificial vision for the company Fresh & Natural C.I.”, Quito: UTE, 2017, pp. 40-46.

[13] MINSA, “Technical Health Standard for the use of Personal Protective Equipment”,pp. 15-32, July 2020.

[14] O. Ezekiel, M. Ekata and A. Eseoghene. “A comparative study of YOLOv5 and YOLOv7 object detection algorithms”, Journal of Computing and Social Informatics, vol 2(1), pp 1-12, February 2023.

[15] G. Jocher, “Ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements”, Github repository, October 2020.

[16] J. Davis and M. Goadrich, “The relationship between Precision-recall and ROC curves”, Proceedings of the 23rd international conference on Machine learning, pp. 233-240, June 2006.