Accident Prevention in Industry 4.0 Using Retrofit: A Proposal

Paulo Henrique Mariano[a], Bruno Pinho Campos[b], Frederico Augusto Cardozo Diniz[c], Carlos Frederico Cavalcanti[d] and Ricardo Augusto Rabelo Oliveira[e]

Computing Department (DECOM), Federal University of Ouro Preto (UFOP), Ouro Preto 35400-000, Brazil

Keywords: Safe, Artificial Intelligence, Pattern Recognition, Machine Learning, Retrofit, Industry 4.0.

Abstract: In this work, we present an Industry 4.0 retrofit solution to prevent accidents in industrial environments, specifically focusing on the operation of bandsaw machines. It examines a real-world scenario where a company aims to enhance worker safety by implementing an integrated solution. The proposed solution involves a pattern recognition system that monitors the work area and sends commands to stop the machine in case of dangerous movements near the bandsaw. This system adheres to Industry 4.0 principles, demonstrating how this methodology can create a safer industrial environment to connect information technology (IT) and operational technology (OT).

1 INTRODUCTION

Despite all the technology involved and discussed, factories still have a lot of manual labor, and some are dangerous tasks that can cause accidents. In 2020, the Ministry of Labor data showed 465.772 workplace accidents in Brazil. Of these, 47.293 involved wrist, hand, and finger injuries, representing 10.15 percent of total accidents, making them the most common type of workplace injury. These injuries include cuts, fractures, and amputations(Social and Security, 2023).

Serious workplace accidents have significant impacts on workers. These impacts include physical injuries, psychological distress, economic hardship, and social vulnerability. Workers often face long recovery periods, permanent disabilities, and associated psychological trauma. The financial impacts are severe, especially for informal workers who lack social protection and benefits. The burden on families is substantial, as they often need to provide care without adequate resources, exacerbating social and economic vulnerabilities(Menezes and Magro, 2023).

In recent times, Industry 4.0 concepts have become a focal point for many companies striving to remain competitive in the modern market. Industry 4.0, often referred to as the Fourth Industrial Revolution, encompasses the integration of Information Technology (IT) and Operational Technology (OT) to create smart factories(Dalenogare et al., 2018). This

Figure 1: Day of Work in an Band Saw Machine.

[a] https://orcid.org/0009-0001-0505-3879
[b] https://orcid.org/0009-0006-8124-0998
[c] https://orcid.org/0009-0007-6977-2411
[d] https://orcid.org/0000-0002-8522-6345
[e] https://orcid.org/0000-0001-5167-1523

Mariano, P. H., Campos, B. P., Diniz, F. A. C., Cavalcanti, C. F. and Oliveira, R. A. R.
Accident Prevention in Industry 4.0 Using Retrofit: A Proposal.
DOI: 10.5220/0013473900003929

In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 2*, pages 649-656
ISBN: 978-989-758-749-8; ISSN: 2184-4992

fusion aims to improve productivity, optimize processes, and significantly reduce waste of materials and money. Industry 4.0 leverages advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), machine learning, big data analytics, and blockchain to create interconnected systems. These technologies facilitate real-time monitoring, predictive maintenance, and seamless communication at all levels of production(Dalenogare et al., 2018). Beyond these capabilities, this technology can also improve the safety and security environment of industrial workers and employees.

This work proposes a novel Industry 4.0 retrofit solution specifically designed to prevent accidents in industrial environments by focusing on the operation of band saw machines. Unlike traditional safety mechanisms, which often rely solely on reactive measures, this solution introduces an integration of machine learning-based pattern recognition to monitor and enhance worker safety proactively. The machine learning system continuously monitors the work area for dangerous movements, sending immediate commands to stop the machine in case of a potential hazard, as depicted in Figure 9.

A key scientific contribution of this work is the proposal of integrating technology into mechanical machines to enhance not only productivity but also safety.

This combination of real-time hazard detection represents a significant advancement in industrial safety systems, ensuring immediate and long-term protection in high-risk environments. The proposed methodology contributes to the field by providing a scalable solution that can be adapted to various industrial processes, establishing a new standard for accident prevention and safety management in Industry 4.0 environments.

2 RELATED WORK

This section presents articles about using pattern recognition in an industry 4.0 context. The paper (Serey et al., 2023) highlights the growing importance of PR and DL due to the increased volume of data generated and the need for efficient management of this data in asymmetric structures. Examines recent PR and DL applications, evaluating their ability to process large volumes of data and identify significant patterns. The review includes an in-depth study of 186 references, selecting 120 for detailed analysis. PR and DL are integrated with various computational techniques, such as data mining, data science, and cognitive systems, to improve automation and reduce human intervention in decision-making processes. Challenges include the need for large datasets for practical training, the high cost of computational, and the difficulty in interpreting deeply nested models.

The article (Zouhal et al., 2021) Provides a comprehensive view of how industrial inspection can be transformed within the context of Industry 4.0. It focuses on adapting industrial inspection to new technologies that allow for an agile and personalized configuration and integrating cameras embedded in robots to automate visual inspection. Discusses the state of the art in visual computing applied to industrial inspection, highlighting how techniques Advanced image processing and machine learning are transforming inspection equipment and quality control. The case study shows how cameras embedded in industrial robots can perform image acquisitions and processing to detect wear on tools and other anomalies in a production environment. The suggestion is that data collected during automated inspections could be used to train predictive maintenance algorithms.

(Oborski and Wysocki, 2022) explores the use of convolutional neural networks (CNNs), to improve visual quality for control systems in Industry 4.0. Several AI algorithms were evaluated to identify those most suitable for visual quality control tasks in a real manufacturing environment. CNNs were chosen for further study because of their effectiveness in processing images and performing complex classification and detection tasks. CNN algorithms were tested with production line datasets and photos. In which he showed efficiency.

The paper (Sharma and Sethi, 2024)show highlights the importance of agriculture in the economy and how plant diseases can significantly affect agricultural production. Recognizes the need for effective methods to detect diseases early to reduce losses. Utilizes deep learning techniques, specifically CNNs, which are well suited to data processing tasks image processing due to its ability to extract essential features from large datasets of images. The model allows for a more precise identification of areas affected by wheat leaf diseases, separating them pixel by pixel. The article presents a significant improvement in classifying diseases in wheat leaves, achieving a 99.43 percent accuracy using Point Rend segmentation, compared to 96.08 percent without segmentation.

The article (Li et al., 2023) presents an improved YOLOv8-based algorithm for detecting glove use in industrial environments, addressing the specific challenges of detecting small objects like hands and

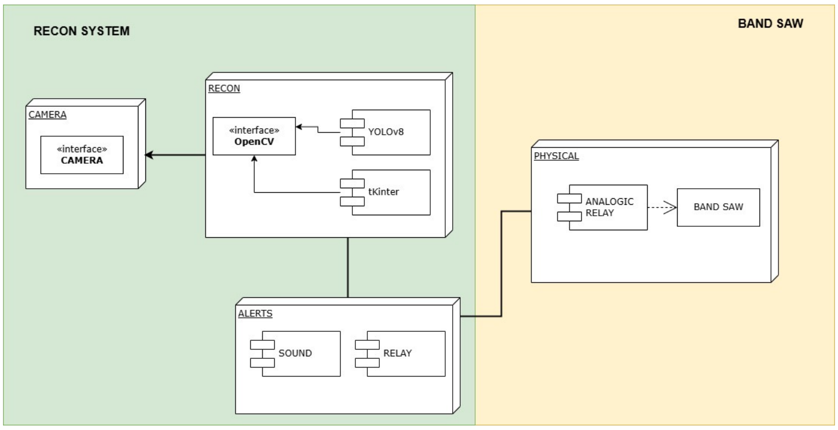

Figure 2: Industry 4.0 Retrofit proposal.

gloves in complex factory settings. The study introduces YOLOv8-AFPN-M-C2f, a modified version of YOLOv8 that replaces the standard PAFPN with the Asymptotic Feature Pyramid Network (AFPN) to improve the integration of non-adjacent feature layers and enhance the detection of smaller objects. The paper also incorporates a new shallow feature layer to increase sensitivity to detailed visual elements, such as gloves. Using a custom dataset of 2,695 high-resolution images from industrial environments, the study demonstrates a 2.6 percent increase in mAP@50 percent, a 63.8 percent improvement in FPS, and a 13 percent reduction in parameters compared to the original YOLOv8 model. These improvements highlight the potential for more accurate and efficient detection of protective gloves in industrial contexts, contributing to enhanced worker safety and automated monitoring of personal protective equipment (PPE) compliance.

Although these articles have focused on product quality and how various techniques can enhance it, this is well-established in the industry. The emphasis on applying these principles to human safety at work is sometimes neglected. Our study approaches this topic by examining the use of technology from a security perspective with integration in industrial machines.

2.1 Machine Learning

The first proposal, as described before, is a computer vision system based on machine learning to improve the recognition of hands with graphene gloves. This study presents three types of machine learning computer vision technologies(MediaPipe, Detectron2 and Yolov8) and compares them to identify which is better for this type of project.

The recognition time is crucial because it defines the speed of the signal sending. The initial algorithm is calculated based on this. The time of response of the algorithm when some point crosses the line is given by:

$$T_d = T_f - T_i \tag{1}$$

Where:

- T_d is the detection time,
- T_f is the final time (the moment when the point crosses the line),
- T_i is the initial time (the moment when the image processing starts).

3 DEVELOPMENT

A Proof-of-Concept (POC) was developed, consisting of a computer vision system and a communication interface with the machine, to validate the solution and integrate various technologies, as described below. OpenCV (Open Source Computer Vision Library) is a highly optimized open-source library that implements real-time computer vision and machine learning algorithms . It is used to capture video in real-time, process the images, and convert them into different color formats to facilitate the recognition (Bradski and Kaehler, 2000). In the project, it is used to video capture and draw the ”dangerous area”.

3.1 MediaPipe

MediaPipe, developed by Google, is a framework for building multimodal (e.g., video, audio, and sensor data) machine learning pipelines. It provides ready-to-use solutions for detecting and tracking body parts. In the algorithm, it is used to identify the hands and fingers. (Research, 2023)

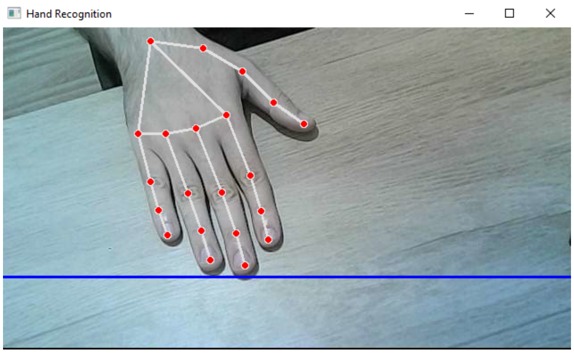

Figure 3: Hand Recognition With MediaPipe.

The algorithm works as follows:

- Image Capture: OpenCV captures real-time video from the camera. The captured images are flipped for a selfie view and converted from BGR to RGB format, which is the format expected by MediaPipe.
- Hand Detection and Tracking: MediaPipe is configured to detect hands with a minimum detection confidence threshold. MediaPipe processes the image for each captured frame to detect hands and identify specific landmarks on each hand.
- Safety Verification: The image includes a safety line. Each hand landmark is checked against this safety line, which covers an area of 8 centimeters and needs to be placed physically at the correct distance from the band saw. If any landmark crosses the safety line, a signal is sent to analogic relay thats stop the machine.

The MediaPipe function works well with bare hands, creating landmarks in points of hands, but it has recognition problems with gloves, resulting in intermittent performance. This lack of recognition poses a dangerous problem in this context. MediaPipe uses a dataset that does not include the specific gloves used in this work and in the factory, which may cause these recognition issues. We particularly notice the intermittence in areas where the gloves have a contrast of gray and black, like can be observe in the Figure 6.

3.2 Detectron2

To enhance recognition capability, we chose Detectron2, a library developed by Facebook AI Research (FAIR) for object detection, segmentation, and other visual recognition tasks (Wu et al., 2019). It's created and used our own dataset of 1200 images of hands with gloves (specially graphene gloves). The chose Detectron2 because of its robust resources for identifying and segmenting images in complex scenarios.

Figure 4: Hand Recognition With MediaPipe, issues in use of gloves.

However, this power requires a machine with high-level hardware.

In our project, after creating the dataset, its prepared the environment and trained the model based on our dataset. Its chose the Mask R-CNN-50 algorithm for training, aiming to segment the gloves in the images, based on this related good accuracy (Tahir et al., 2021). Image segmentation is necessary to recognize only the gloves and not other parts manipulated by the operator.

Figure 5: Hand Recognition(in Red) throw by segmentation with Detectron2.

The recognition results with Detectron2 were good, as it was able to recognize hands with different types of gloves (and most important to the project the graphene gloves). When an image is segmented and recognized, it is painted in red. We combined Detectron2 with OpenCV to provide the solution. OpenCV was used to capture real-time images and draw the "dangerous area," a rectangle to simulate the place of the saw. When the segmentation area crosses the dangerous area, the rectangle is also painted in red.

However, the performance of the solution was poor. The frames per second (fps) were less than one, compromising the effectiveness of the solution in real time. Other algorithms based on Detectron2 were tested, but we could not achieve success in improving performance. We reduced the size of the images to 640x480, but the fps did not improve. This indicates that more computing power is needed to use this solution effectively.

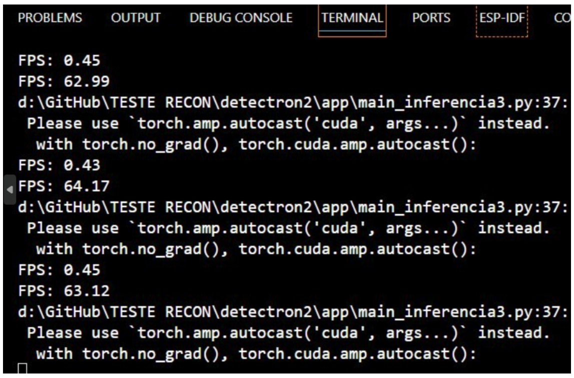

Figure 6: The output during the execution of Detectron2 showed intermittent FPS, highlighting the performance issues.

3.3 YOLOv8

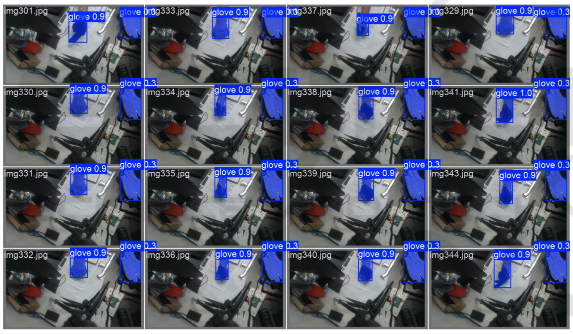

Figure 7: Result the training of model with YOLOv8.

Then we tried other techniques and decided to use YOLOv8 to recognize the hands. The approach with YOLOv8 was chosen due to its capability to classify and segment images, especially in real-time, with good FPS and accuracy (Jocher et al., 2023).

We started with training using a dataset of 1320 images of gloves in special graphene gloves in different positions, qualities, and lighting conditions. The algorithm performed well, with good accuracy and FPS, which led us to invest in this solution. The parameters to the training is describe in the table 1:

The graphs in the Figure 9, shows the results of metrics and losses during the training and validation of a model configured with 50 epochs, for the task of object detection and/or segmentation of one class (in this case, gloves).

Columns: Each column represents different metrics or types of loss. Rows (Train/Val):

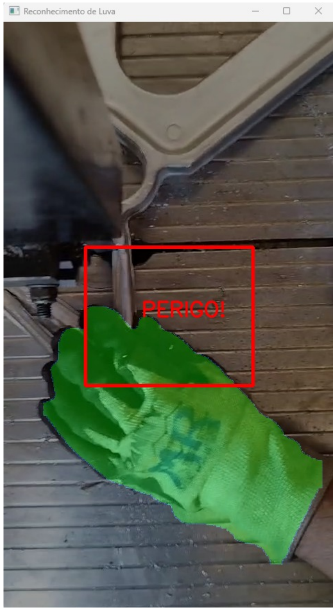

Figure 8: Solution Developed with YOLOv8.

Table 1: Training Parameters.

Parameter	Value
Model	yolov8n.pt
Epochs	50
Image Size	320
Batch Size	16
Workers	8

The top row refers to the training results. The bottom row refers to the validation results.

- The train/box-loss and val/box-loss represents the loss related to bounding boxes. Loss should decrease over epochs, indicating the model is learning to predict boxes correctly and int the model there is consistent reduction, which is a positive sign.

- The train/seg-loss and val/seg-loss: Related to segmentation loss (e.g., segmentation masks for gloves). The expected behavior is gradual decrease over epochs.Despite some initial fluctuations in validation loss, it shows convergence, suggesting improvement.

- The train/cls-loss and val/cls-loss refers to classification loss, indicating how well the model identifies objects as ”glove.”. The loss decreases, but more pronounced peaks in validation suggest potential data inconsistencies or overfitting.

- The train/dfl-loss and val/dfl-loss related to the prediction of distribution locations (distillation loss). The expected behavior is a continuous decrease. While loss decreases overall, peaks in validation indicate challenges with certain instances.

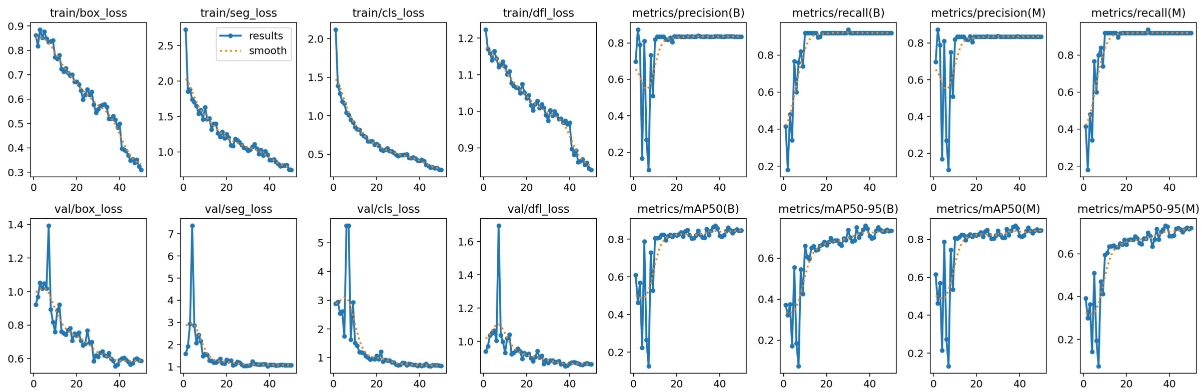

Figure 9: Results of training with YOlOv8.

- The metrics/precision(B) and metrics/ precision(M): Precision measures the proportion of correct detections relative to all detections. B and M might represent different configurations (e.g., bounding boxes vs. masks). The precision increases rapidly and stabilizes above 0.8, which is excellent.

- The metrics/recall(B) and metrics/recall(M): Recall measures the proportion of correctly identified objects relative to all existing objects. In th e training the recall improves and stabilizes above 0.8, indicating good coverage.

- metrics/mAP50(B) and metrics/mAP50(M): mAP50 (mean Average Precision) evaluates average precision with an IoU threshold of 50The results stabilize above 0.8, showing strong glove detection performance.

- The metrics/mAP50-95(B) and metrics/mAP50-95(M): mAP50-95 is a stricter metric considering multiple IoU thresholds (from 50 percent to 95 percent). While more challenging, values stabilize above 0.6, which is good for complex object detection tasks.

The results obtained over 50 epochs demonstrate consistent model performance. The training and validation losses (box-loss, cls-loss, seg-loss, and dfl-loss) gradually decreased, indicating effective learning. Despite some peaks in validation losses, the overall trend shows convergence, suggesting good generalization.

Performance metrics, such as precision and recall, stabilized above 0.8, indicating high reliability in glove detection. The mAP50 metric showed strong results (0.8), while the more stringent mAP50-95 remained around 0.6, reflecting the inherent challenges of the task.

The validation loss, although following a trajectory similar to the training loss, shows some spikes (particularly in val/cls-loss and val/dfl-loss), which may be related to outliers or specific characteristics of the validation set that make it difficult for the model to generalize. These spikes are expected in complex object detection problems and do not compromise overall performance, given that the global trend of the losses is towards convergence. In practice the results were observer, with a good balance between accuracy and performance.

After training the model, it was combined with OpenCV to capture and analyze video in real time for detection. OpenCV was also used to draw a "dangerous area," a rectangle around the saw of the band saw machine. When any part of the glove's mask segmentation enters the rectangle, the rectangle changes color to red, an alert sound is triggered, and a signal is sent to an analog relay. The Python library Tkinter was used to create controls for adjusting the width and height of the "dangerous area," allowing occupational safety engineers to define the dimensions of this safety zone. Additionally, the recognition box was removed, leaving only the mask segmentation to avoid cluttering the screen and prevent false alerts.

To ensure that the system runs smoothly without interruptions or delays, the sound signal processing was handled in a separate, independent process from the recognition process. This means that while one process is dedicated to detecting and segmenting the gloves in real time, another process is responsible for managing the sound alert system and sending the signal to the relay. This separation helps prevent any potential bottlenecks or slowdowns in the recognition process, ensuring that the emergency stop system can respond quickly and reliably.

The analog relay, activated by the signal from the independent process, triggers the existing physical emergency stop system of the machine.

4 RESULTS

The Proof-of-Concept (POC) assessed the system's ability to recognize hands and promptly send a stop signal when any landmark crossed the predefined safety line. The algorithms tested showed that the best combination of performance and accuracy was achieved using YOLOv8.

The YOLOv8 model demonstrated high reliability in detecting gloves, with performance metrics such as precision and recall stabilizing above 0.8. The mAP50 metric showed strong results (0.8), while the more stringent mAP50-95 remained around 0.6, reflecting the inherent challenges of the task.

The system was able to process real-time video and accurately detect when a hand with graphene gloves entered the dangerous area. The integration with OpenCV allowed for real-time video capture and the drawing of the "dangerous area," while the Python library Tkinter provided controls for adjusting the dimensions of this safety zone.

The separation of the sound signal processing into an independent process ensured that the system ran smoothly without interruptions or delays. This setup allowed the emergency stop system to respond quickly and reliably, enhancing the overall safety of the workplace.

In conclusion, the POC validated the effectiveness of the proposed solution, demonstrating that it is possible to integrate advanced computer vision and machine learning technologies into mechanical machines to enhance safety.

4.1 Experimental Setup

This initial phase validates the project's foundational concepts and technologies. These tests were not performed in the actual deployment scenario but in a controlled laboratory setting, denoted as the POC phase. A full HD 1080p webcam was used to connect to a PC via a USB 2.0 port in this setup. The application and the Modbus simulator ran on the same computer, which was equipped with an Intel 11th generation i5 processor at 3.2 GHz, 32 GB of RAM, an NVMe SSD of 512 GB, and a GeForce 4060 GPU with 8 GB. The solution was also tested with videos from the industrial process.

The project was further tested on a Raspberry Pi 5 (8 GB RAM, Quad-Core Arm Cortex-A76 processor at 2.4 GHz) with an AI Kit, and it demonstrated good performance as well.

4.2 Restrictions

The dataset for training the machine learning models was time-consuming and required capturing a wide variety of images of gloves in different positions, lighting conditions, and qualities, especially graphene gloves, the increase of the dataset can improve the accuracy with other types of gloves. Industrial environments are typically harsh on devices. The next step of the project is to implement the solution in the factory to observe its behavior, particularly with variations in lighting throughout the day and the presence of dirt. This aggressive environment can be a challenge for small devices like Raspberry Pi.

It is essential to raise awareness among workers about safety. The system does not work alone; adherence to safety rules is necessary to ensure the success of the project.

5 CONCLUSION

Retrofit is an effective solution for adapting existing equipment to meet the demands of modern industries. The project applied contemporary concepts such as machine learning to a standard industrial band saw machine. The initial test successfully integrated these technologies into the Industry 4.0 framework. In conclusion, the POC shows that this method can be extended to other machines and industrial processes, particularly to enhance the safety and security of employees in dangerous operations.

6 FUTURE WORKS

The work is currently in development, and the results from the Proof-of-Concept (POC) tests are promising, indicating a positive direction. The initial proof-of-concept phase involved controlled experiments in real-time recognition and videos of the workplace, which were sufficient to validate the system's core functionality and response times. Future work will include scaling these tests to real-world industrial environments and conditions to further refine the system's robustness. This includes integrating the application into small devices and observing the algorithm's behavior, speed, and accuracy in recognizing hands in a workplace environment. Additionally, an integration with a private blockchain solution will be developed to create immutable records of unsafe behaviors.

REFERENCES

Bradski, G. and Kaehler, A. (2000). Opencv: Open source computer vision library. https://opencv.org/. Accessed: 2025-01-19.

Dalenogare, L. S., Benitez, G. B., Ayala, N. F., and Frank, A. G. (2018). The expected contribution of industry 4.0 technologies for industrial performance. *International Journal of Production Economics*.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics yolov8.

Li, S., Huang, H., Meng, X., Wang, M., Li, Y., and Xie, L. (2023). A glove-wearing detection algorithm based on improved yolov8. *Sensors*, 23(24).

Menezes, M. N. and Magro, M. L. P. D. (2023). Psychosocial impacts of serious accidents at work: A look at the workers monitored by the worker's health reference center. *Revista Jurídica Trabalho e Desenvolvimento Humano*, 6:1–30.

Oborski, P. and Wysocki, P. (2022). Intelligent visual quality control system based on convolutional neural networks for holonic shop floor control of industry 4.0 manufacturing systems. *Advances in Science and Technology. Research Journal*, 16(2):89–98.

Research, G. (2023). Mediapipe: Cross-platform framework for building perception pipelines. https://ai.google.dev/edge/mediapipe/solutions/guide. Accessed: 2025-01-19.

Serey, J., Alfaro, M., Fuertes, G., Vargas, M., Durán, C., Ternero, R., Rivera, R., and Sabattin, J. (2023). Pattern recognition and deep learning technologies, enablers of industry 4.0, and their role in engineering research. *Symmetry*, 15(2):535.

Sharma, T. and Sethi, G. K. (2024). Improving wheat leaf disease image classification with point rend segmentation technique. *SN Computer Science*, 5:244.

Social and Security (2023). Work accident and disability. https://www.gov.br/previdencia/pt-br/assuntos/previdencia-social/saude-e-seguranca-do-trabalhador/acidente_trabalho_incapacidade. Accessed: June 23, 2024.

Tahir, H., Khan, M. S., and Tariq, M. O. (2021). Performance analysis and comparison of faster r-cnn, mask r-cnn and resnet50 for the detection and counting of vehicles. *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 587–594.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. https://github.com/facebookresearch/detectron2.

Zouhal, Z. et al. (2021). Approach for industrial inspection in the context of industry 4.0. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–5, Mauritius, Mauritius.