Predictive Maintenance Based on Deep Learning: Early Identification of Failures in Heavy Machinery Components

Pablo Cabrera Melgar, Luis Hilasaca Chambi, Raúl Sulla Torres
Universidad Nacional De San Agustín De Arequipa, Arequipa, Perú

***Abstract*—Deep learning-based predictive maintenance is a key strategy in industry to prevent unexpected failures, reduce downtime, and improve operational safety. This study presents an advanced approach for early fault detection in heavy machinery components using image analysis, focusing on four critical defect types: hose wear, piston failure, corrosion, and moisture. To this end, three state-of-the-art object detection models were implemented and compared: YOLOv11, RT-DETR, and YOLO-World. The dataset consists of images captured in real-life industrial environments exhibiting variations in lighting, texture, and material degradation. A manual preprocessing and annotation process was applied to improve training quality. Model performance was evaluated using key metrics such as the precision-recall (PR) curve and the confusion matrix to determine the most efficient technique for real-time fault detection. Experimental results show that YOLOv11 achieves the highest overall accuracy, with an mAP@0.5 of 83.8%, followed by YOLO-World at 82.4% and RT-DETR at 80.3%. In terms of efficiency, YOLO-World offers a balance between accuracy and detection speed, while RT-DETR shows stable performance but lower accuracy for certain defect types. These findings confirm that deep learning-based detection models enable the rapid and accurate identification of industrial defects, facilitating the implementation of predictive maintenance strategies.**

Keywords*—*Predictive maintenance; deep learning; fault detection; artificial intelligence

I. Introduction

Maintenance of heavy machinery is a critical issue in industry, as unexpected failures can lead to high operating costs, downtime and safety risks. Traditionally, corrective and preventive maintenance strategies have been used. Corrective maintenance consists of repairing a machine only when it fails, which can cause costly interruptions. Preventive maintenance, on the other hand, involves scheduled inspections and repairs, even if they are not always necessary, which can increase costs without ensuring efficiency [1]. In this context, deep learning-based predictive maintenance has emerged as an innovative solution that enables early detection of failures through real-time data analysis [2]. This methodology not only optimizes intervention planning, but also reduces operating costs, minimizes downtime, and improves the efficiency of industrial inspections, contributing to greater productivity and safety in working environments [3].

Recent advances in computer vision have facilitated the automation of fault detection in mechanical components, enabling accurate defect identification through image analysis [4]. In particular, object detection models have demonstrated high performance in visual inspection tasks, providing scalable and efficient solutions for industrial maintenance [5]. However, most existing studies focus on general-purpose datasets, operate under controlled laboratory conditions, or focus on isolated defect types, which limits their applicability in real-life industrial settings. Furthermore, many models prioritize accuracy over inference speed, making them unsuitable for real-time applications where immediate fault detection is crucial.

Therefore, there is a clear gap in the literature regarding the implementation of fast, accurate, and generalizable object detection models in real-life industrial settings, specifically for the detection of multiple simultaneous faults in heavy machinery components. This work addresses this gap by integrating state-of-the-art object detection architectures capable of managing visual variability, complex backgrounds, and diverse fault types under operating conditions.

In this study, three state-of-the-art models are evaluated: YOLOv11, RT-DETR, and YOLO-World. These were selected for their advanced capabilities in balancing speed and accuracy, their adaptability to diverse visual inputs, and their proven performance in object detection benchmarks. YOLOv11 offers a strong balance between real-time performance and detection accuracy. RT-DETR incorporates transformer-based attention mechanisms that improve the recognition of small or occluded defects. YOLO-World offers greater flexibility in managing open vocabulary detection, which is essential when defect categories evolve or are refined over time.

The primary objective of this research is to develop an automatic visual inspection system for detecting defects in heavy machinery components based on deep learning. The system focuses on four recurring industrial defects: hose wear, piston failure, moisture, and corrosion. To achieve this goal, a specific dataset of high-resolution images captured under real-world working conditions was created and manually annotated. The models are trained and evaluated to compare their accuracy, processing speed, and generalization capabilities in realistic scenarios.

The paper is organized as follows: Section II provides a review of related work, Section III details the methodology, Section IV presents the experiments and results, and Section V presents conclusions.

II. Related Work

This study [6] proposes an improved version of the VGG19 convolutional neural network, named Multipath VGG19

(MVGG19), for the detection of defects and the recognition of industrial objects. Six public data sets with images of mechanical parts and defective materials were used. MVGG19 improves feature extraction using a multi-path scheme and concatenation fusion. The experiments showed that MVGG19 outperforms VGG19 in five of the six datasets, with an average improvement of 6.95% in the classification accuracy.

This paper in [7] presents a deep learning-based approach to identify defects in images, exploring segmentation and unsupervised detection methods. For evaluation, the MVTec Anomaly Detection (MVTec AD) dataset is used, which provides images with more than 70 types of anomalies and accurate pixel-level annotations. Different approaches based on deep neural networks, such as autoencoders and generative models, are compared with traditional computer vision techniques. The results obtained show the performance of the evaluated methods and highlight improvement opportunities for applications in real environments.

The objective of the research [8] is the development of a multi-phase Convolutional Neural Network (CNN) model to detect and analyze corrosion in metallic materials. The model employs binary classification, multiclass classification and patch distribution to identify affected areas. It was trained with 600 images, achieving 94.87% accuracy in binary classification, 92.1% in multiclass classification and up to 96.5% in patch distribution. In addition, it achieved 91.5% accuracy in region segmentation at the image level and 89.2% at the pixel level. This approach is useful for experts in critical industries such as aerospace and manufacturing and can be applied in other areas beyond corrosion.

In this paper [9], PatchCore is presented for anomaly detection in industrial manufacturing using a representative memory pool of nominal local features, achieving a balance between inference times and performance. In the MVTec AD benchmark, it achieves an AUROC of up to 99.6%, halving the error rate compared to the best competitor. Furthermore, it offers competitive results on other datasets and in scenarios with few samples.

In this study [10], a comprehensive review of deep learning-based anomaly detection techniques is presented, analyzing neural network architectures, supervision levels, loss functions, metrics and datasets. In addition, a framework based on industrial environments is proposed and current approaches are evaluated under this context. Open challenges in image anomaly detection are also highlighted and the advantages and limitations of various architectures depending on their supervision level are analyzed.

The objective of this research [11] is to analyze the use of convolutional neural networks (CNN) for the automated detection of corrosion on metallic surfaces. For this purpose, different CNN architectures are compared, including pre-trained models and specific designs adapted to this problem. The results show that CNNs outperform traditional methods based on texture and color analysis, improving both the accuracy and efficiency of the inspection process. In addition, one of the proposed architectures significantly optimizes the computational time, maintaining a performance comparable to that of the most advanced models.

This study [12] proposes a method based on artificial neural networks to detect internal leakage in hydraulic cylinders by analyzing pressure signals. Key features such as location, height and width of the peaks are extracted, reducing dimensionality and optimizing processing. The neural network classifies the system into three states: optimal, mild failure and severe failure. This approach improves the detection of leaks caused by wear and seal damage, increasing the reliability and efficiency of hydraulic systems in heavy machinery, reducing costs and maintenance times in industrial environments.

This research [13] analyzes the current challenges and provides a review of the most recent unsupervised approaches, organized into five categories. In addition, public datasets used in this area are presented and different methods are compared to identify their advantages and disadvantages. Finally, unsolved problems are highlighted and future lines of research are proposed to foster the development of more efficient and applicable solutions in different industrial sectors.

This study [14] presents a semi-orthogonal embedding-based approach for unsupervised anomaly segmentation by optimizing the use of multi-scale features of pre-trained CNNs along with Mahalanobis distance. It aims to mitigate the high computational cost associated with multidimensional covariance tensor inversion, a key limitation for scalability in deep networks. To this end, random feature selection is generalized using semi-orthogonal embedding, which allows for a more efficient and robust approach, cubically reducing the computational cost without affecting performance. Experiments on standard datasets, such as MVTec AD, KolektorSDD, KolektorSDD2 and mSTC, show that this method outperforms the state of the art, achieving significant improvements in accuracy and efficiency. These results validate its applicability in large-scale anomaly detection.

The research proposes [15] a new framework called PaDiM for image anomaly detection and localization within a single-class learning environment. PaDiM employs a pre-trained convolutional neural network (CNN) for patch-level feature extraction and models the normal class distribution using multivariate Gaussian distributions. In addition, it takes advantage of correlations between different semantic levels of the CNN to improve anomaly localization. The proposal outperforms current methods on MVTec AD and STC datasets, and extends the evaluation protocol to measure its performance on unaligned datasets, getting closer to real industrial inspection scenarios. Thanks to its low computational complexity and high performance, PaDiM is presented as a viable alternative for various industrial applications.

The study [16] proposes Abyss Fabric, an automated system for corrosion detection and monitoring on offshore platforms, improving maintenance efficiency. Using computer vision and a Convolutional Neural Network (CNN), it segments inspection images and integrates the results into a digital twin to identify corrosion and its severity. Evaluated on an oil platform, it achieves 91.83% accuracy, processing large volumes of data automatically and optimizing maintenance planning, reducing costs and operational risks.

The research proposes [17] an unsupervised deep learning model for anomaly detection in temporal data of manufacturing processes, with the aim of improving the interpretability and scalability of these systems in industrial environments. Its application in the assembly tightening process in the automotive industry demonstrates a significant improvement in anomaly identification, facilitating its implementation and overcoming the limitations of conventional approaches.

This study [18] proposes a crash response testing method and frequency-domain analysis to detect defects in shock absorber rods and steering racks on an automotive production line. Machine and deep learning were used to build a discrimination model based on features extracted from the measured signals. The results indicate that frequency analysis accurately identifies the location and presence of defects, improving quality control and facilitating the implementation of smart factories.

III. Methodology

The proposed methodology, illustrated in Fig. 1, focuses on detecting faults in heavy machinery components using deep learning. The objective is to identify signs of deterioration in real time, enabling the implementation of predictive maintenance strategies. The methodology consists of four main phases: data collection, preprocessing, model training, and model evaluation.

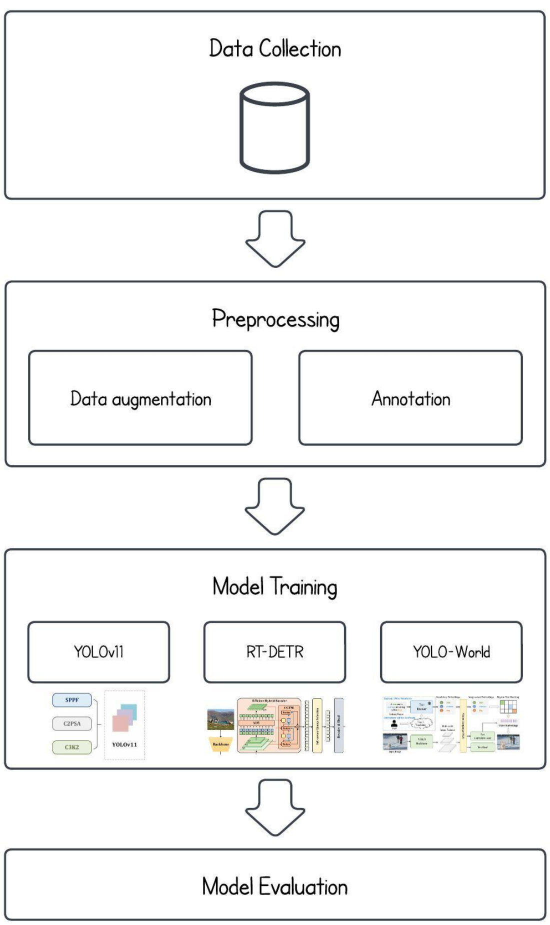

Fig. 1. Proposed methodology.

A. Data Collection

The dataset used in this study was developed from scratch and consists of images captured in real industrial environments, with the goal of ensuring representative and relevant samples. The images were taken with a high-resolution camera, allowing for an adequate level of detail for visual fault identification. During the capture process, specific criteria related to image quality are determined, such as good resolution and lighting conditions that ensure clear visibility of critical areas of the machinery. Furthermore, efforts were made to include images under different operating conditions to increase the variability and robustness of the dataset.

The dataset covers four main types of failures in heavy machinery: hose wear, piston failure, moisture, and corrosion. Hose wear includes cracks, abrasions, and deformations in hydraulic and pneumatic systems due to prolonged use or extreme conditions. Piston failure manifests as oil leaks, cracks, or loss of displacement efficiency. Moisture refers to the presence of water or oil, which can indicate leaks or condensation. Finally, corrosion refers to the deterioration of metallic components due to exposure to moisture, chemicals, or aggressive environments. This unique dataset forms the basis for the development of an automated fault detection system, focused on improving predictive maintenance strategies in industrial settings. Representative images of each type of fault are presented in the Fig. 2.

Fig. 2. Component failure dataset.

Table I presents the four defect classes used for training, along with the number of labeled instances in the component failure dataset. These classes represent common failure types in heavy machinery, where accurate detection is essential for predictive maintenance.

B. Preprocessing

To optimize the performance of the fault detection system for heavy machinery components, the dataset underwent a preprocessing process that included two key stages: data augmentation and detection annotation. These techniques improved the model's ability to identify faults under a variety of conditions,

TABLE I. DATASET CLASSES AND LABELS

Classes	Labels
Hose wear	324
Piston failure	268
Moisture	234
Corrosion	638

ensuring better generalization and reducing the impact of a limited dataset.

1) Data augmentation: Given the challenges of collecting large quantities of defect images in industrial environments, common data augmentation techniques, such as rotation, flipping, brightness adjustments, and contrast modifications, are employed to simulate various real-world conditions. These transformations were applied using standard libraries such as `Albumentations` and `OpenCV`, widely used in computer vision. This approach improves the robustness and generalization capabilities of deep learning models, increasing their accuracy under different lighting, orientation, and perspective conditions [19].

2) Annotation: Each image in the dataset was manually annotated using bounding boxes to accurately identify the faults present. To ensure labeling consistency and quality, an annotation protocol was applied that included: a clear definition of each fault type supported by visual examples, precise delineation of the affected areas, and expert cross-review to validate each annotation.

This rigorous annotation process is crucial, as label accuracy and consistency directly affect the model's ability to learn useful representations. In industrial applications, where fault detection can entail significant costs, labeling quality is a determining factor for the effectiveness of predictive maintenance systems [20].

C. Model Training

For early detection of faults in heavy machinery components, three advanced models are selected: YOLOv11, optimized for real-time detection and capable of rapid response to anomalies during operation; RT-DETR, based on transformer architectures, which stands out for its high accuracy in identifying complex defects through contextual and spatial analysis of images; and YOLO-World, which integrates vision-language modeling for open vocabulary detection. This combination ensures an optimal balance between speed, accuracy, and flexibility for continuous industrial monitoring.

1) YOLOv11: It is an advanced real-time detection model that improves feature extraction using C3k2, SPPF, and C2PSA blocks, achieving greater accuracy (mAP) and computational efficiency [21]. Its high speed and ability to detect small defects make it ideal for predictive maintenance, allowing anomalies in industrial components to be identified before critical failures occur. Its scalable design facilitates deployment on both edge devices and high-performance environments, optimizing early detection and reducing downtime. Fig. 3 illustrates the architecture of YOLOv11.

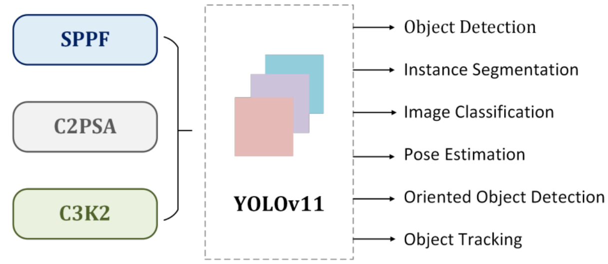

Fig. 3. Architecture of YOLOv11.

2) RT-DETR: This is a Vision Transformer-based detection model designed to operate in real time without the need for non-maximum suppression (NMS), which improves its efficiency [22]. Its architecture, illustrated in Fig. 4, optimizes multi-scale feature fusion using a hybrid encoder and IoU-based query selection, enabling higher detection accuracy. This balance of speed and accuracy makes it particularly suitable for applications such as surveillance, autonomous driving, and predictive maintenance in industrial environments.

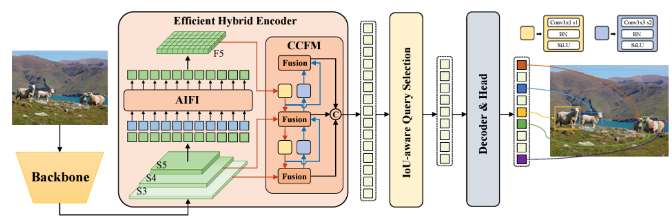

Fig. 4. Architecture of RT-DETR.

3) YOLO-World: It is an extension of the YOLO series that introduces open vocabulary detection capabilities through vision-language modeling and pre-training on large datasets [23]. To this end, it incorporates the Re-parameterizable Vision-Language Path Aggregation Network (RepVL-PAN) and a region-text contrastive loss, which improves the interaction between visual and linguistic information. This approach enables object detection in a zero-shot scenario with high efficiency. The architecture of the model is illustrated in Fig. 5. In addition, its tuned version exhibits outstanding performance in tasks such as object detection and instance segmentation with open vocabulary.

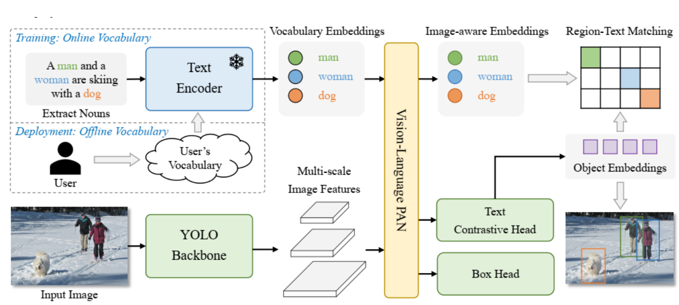

Fig. 5. Architecture of YOLO-World.

IV. Experiments and Results

To evaluate the performance of the selected object detection models YOLOv11, RT-DETR, and YOLO-World a comprehensive set of experiments was conducted using the proprietary Component Failure Dataset, specifically designed for detecting faults in heavy machinery components. This section details the evaluation methodology, performance metrics, and experimental results, providing a thorough assessment of each model's effectiveness in real-world industrial applications.

A. Model Performance Evaluation

The selected models were trained and validated under industry-relevant conditions, ensuring a realistic assessment of their detection accuracy, computational efficiency, and real-time applicability. The evaluation framework incorporated key performance metrics to provide a holistic analysis of each model's strengths and limitations:

1) Precision-Recall (PR) curve: Provides insights into the trade-off between precision and recall, helping assess detection reliability across different defect types.

2) Confusion matrix: Analyzes classification accuracy, highlighting correct detections and common misclassifications to identify areas for improvement.

3) Inference speed (FPS - Frames Per Second): Determines the efficiency of the model processing, which is critical for real-time fault detection in industrial environments.

This structured evaluation ensures that models are evaluated not only in terms of accuracy, but also in terms of deployment feasibility, computational efficiency, and their ability to minimize false detections in an real setting.

B. YOLOv11

To evaluate YOLOv11 performance in detecting component failures, we present the precision-recall (PR) curve and the confusion matrix, which provide insight into its detection capabilities.

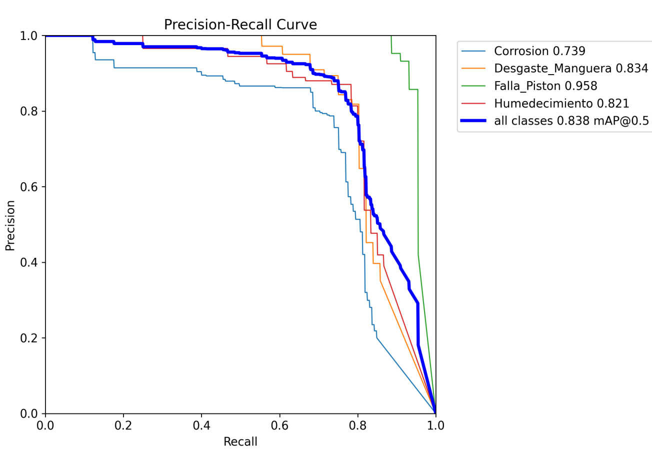

Fig. 6. PR_curve YOLOv11.

Fig. 6 shows the precision-recall (PR) curve of YOLOv11, highlighting its performance in detecting four types of defects: corrosion, hose wear, piston failure, and moisture. The model achieves an overall average precision (mAP@0.5) of 83.8%, with its best performance in detecting piston failures 95.8%, followed by hose wear 83.4% and moisture 82.1%. However, corrosion has the lowest performance 73.9%, suggesting potential challenges in its identification.

Together, these results demonstrate the robustness of YOLOv11 for industrial fault detection, while highlighting opportunities for improvement, particularly in corrosion detection.

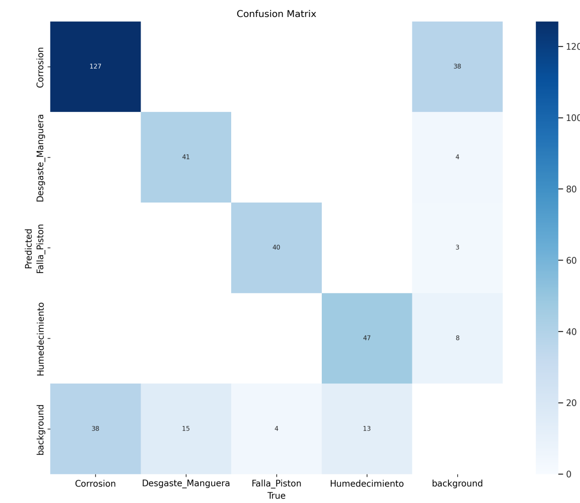

Fig. 7. Confusion matrix YOLOv11.

In Fig. 7, the confusion matrix provides a detailed evaluation of YOLOv11 classification performance in detecting defects in heavy machinery. The correct predictions are concentrated on the main diagonal, demonstrating high precision for most categories, including corrosion, hose wear, piston failures, and moisture. The model achieves excellent performance in identifying piston failures and hose wear, with minimal misclassifications.

However, some misclassifications are observed, especially in corrosion, where they are incorrectly classified as background. Similarly, some moisture cases are also misclassified as background, suggesting that the model may have difficulty distinguishing these defects under certain conditions. These results highlight the robustness of YOLOv11 in defect detection while also indicating potential areas for improvement, especially in reducing misclassifications with the background category.

C. RT-DETR

The performance of the RT-DETR model in detecting defects in heavy machinery is evaluated using the PR curve and the confusion matrix, allowing a detailed analysis of its detection accuracy.

Fig. 8 shows the precision-recall (PR) curve obtained with the RT-DETR model for the detection of different types of faults in heavy machinery components. Individual curves are presented for each evaluated class: Corrosion 73.4%, Hose Wear 84.4%, Piston Failure 91.8% and Moisture 71.6%. In addition, the overall performance curve of the model for all classes is included, obtaining an (mAP@0.5) of 80.3%. These results indicate good performance of the model in fault detection, with the "Piston Failure" class presenting the highest precision compared to the others.

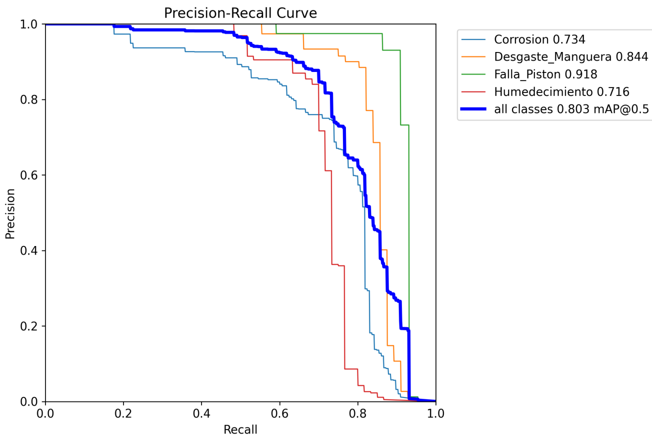

Fig. 8. PR_curve RT-DETR.

Fig. 9 presents the confusion matrix, which allows a detailed analysis of the RT-DETR model performance in identifying faults in heavy machinery.The model performs solidly in detecting piston faults and hose wear, with a low number of misclassifications. However, some confusions were identified, especially in the corrosion category. Similarly, certain examples of moisture were misclassified, suggesting that the model may have difficulty differentiating these faults under specific conditions.

These findings demonstrate the effectiveness of the RT-DETR model in detecting industrial defects, although they also highlight areas for improvement, primarily in reducing false negatives in the classification of corrosion and moisture.

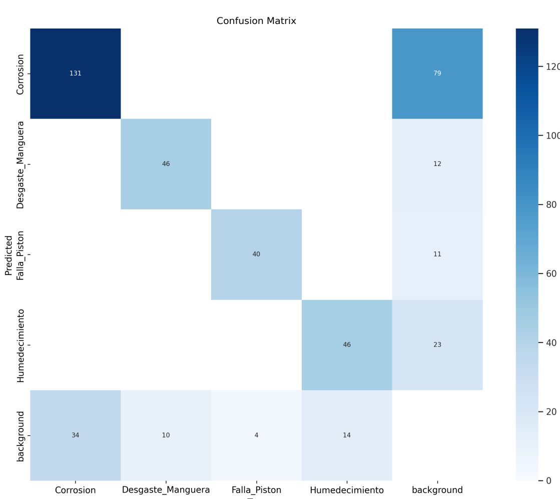

Fig. 9. Confusion matrix RT-DETR.

D. YOLO-World

To evaluate the performance of the YOLO-World model in industrial fault detection. In particular, the Precision-Recall (PR) curve and the confusion matrix allow analyzing the model's ability to differentiate between different classes of defects.

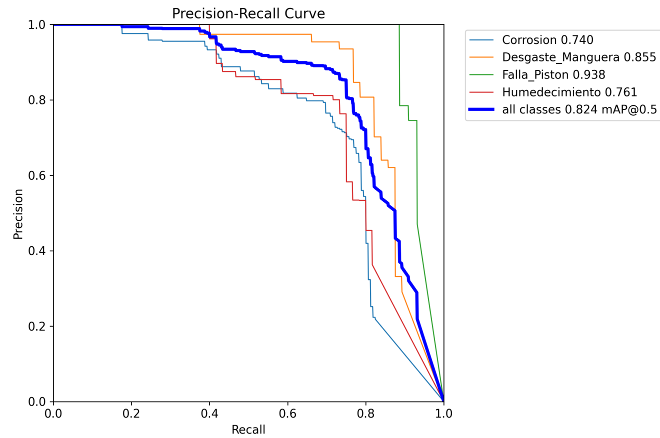

Fig. 10. PR_curve YOLO-World.

Fig. 10 presents YOLO-World PR curve, with an overall mAP@0.5 of 82.4%. The highest detection accuracy is achieved for piston failure 93.8%, followed by hose wear 85.5%, moisture 76.1%, and corrosion 74.0%. While the model performs well, corrosion detection remains the most challenging category.

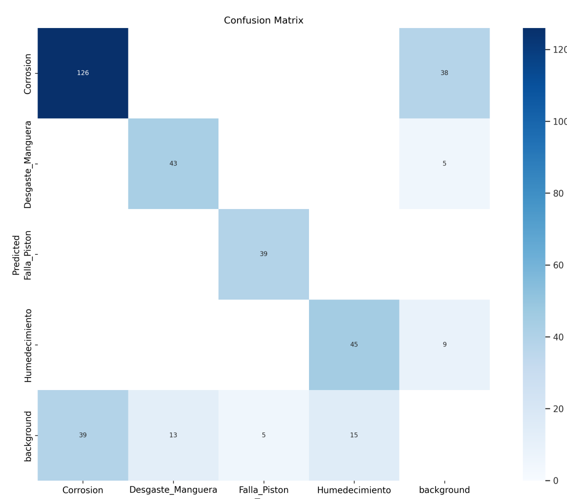

Fig. 11. Confusion matrix YOLO-World.

The confusion matrix Fig. 11 shows that YOLO-World achieves high accuracy, with most correct predictions aligned on the diagonal. However, corrosion and moisture exhibit higher misclassification rates, indicating potential difficulties in distinguishing these defects from background noise. Further refinement in feature extraction could improve the model's accuracy in these categories.

In general, YOLO-World demonstrates competitive performance, although improvements in corrosion and wetting detection could further improve its reliability.

E. Comparison and Discussion

The performance of the three selected object detection models, YOLOv11, RT-DETR, and YOLO-World, was analyzed based on their detection accuracy and computational efficiency. This section provides a comparative discussion of the models' strengths, limitations, and potential improvements in detecting faults in heavy machinery components.

Table II shows a comparison of the performance of the YOLOv11, RT-DETR, and YOLO-World models in fault detection, evaluated by overall accuracy, mAP@0.5, and inference time. Among them, YOLOv11 stands out as the most accurate model, achieving an mAP@0.5 of 83.4%, indicating its high ability to accurately identify defects. Furthermore, its inference time of 32.6 ms positions it as an efficient option for real-time applications [21].

On the other hand, RT-DETR achieved the highest overall accuracy 94.6%, but its mAP@0.5 of 80.3% was the lowest, suggesting that its detection may be less reliable compared to the other models [22]. Furthermore, its high inference time 45.7 ms makes it less suitable for speed-critical environments. In contrast, YOLO-World offers the best balance between accuracy and efficiency [23], with a mAP@0.5 of 82.4% and the lowest inference time 29.7 ms, making it the best alternative for real-time detection tasks, although with a slight reduction in accuracy compared to YOLOv11.

TABLE II. COMPARISON OF RESULTS

Model Results			
Model	**Accuracy(%)**	**mAP@0.5(%)**	**infer(ms)**
YOLOv11	89.1%	83.4%	32.6
RT-DETR	94.6%	80.3%	45.7
YOLO-World	90.2%	82.4%	29.7

The images in Fig. 12 illustrate detection examples for the three models. These images show how each model identifies component defects, highlighting the differences in accuracy and misdetection. This visual comparison reinforces the findings in the table, providing a clear representation of each model's strengths and weaknesses in detecting faults in heavy machinery components.

V. CONCLUSION

The study demonstrates that computer vision and deep learning models, such as YOLO-World, YOLOv11, and RT-DETR, are highly effective for fault detection in industrial environments, combining accuracy and operational efficiency. Among them, YOLOv11 achieved the highest overall accuracy, with an (mAP@0.5) of 83.8%, outperforming YOLO-World 82.4% and RT-DETR 80.3%. However, YOLO-World stood out for its balance between accuracy and inference speed, making it particularly suitable for real-time applications. These results underscore the importance of selecting models that are not only accurate but also adaptable to real-world conditions, ensuring efficient and reliable performance.

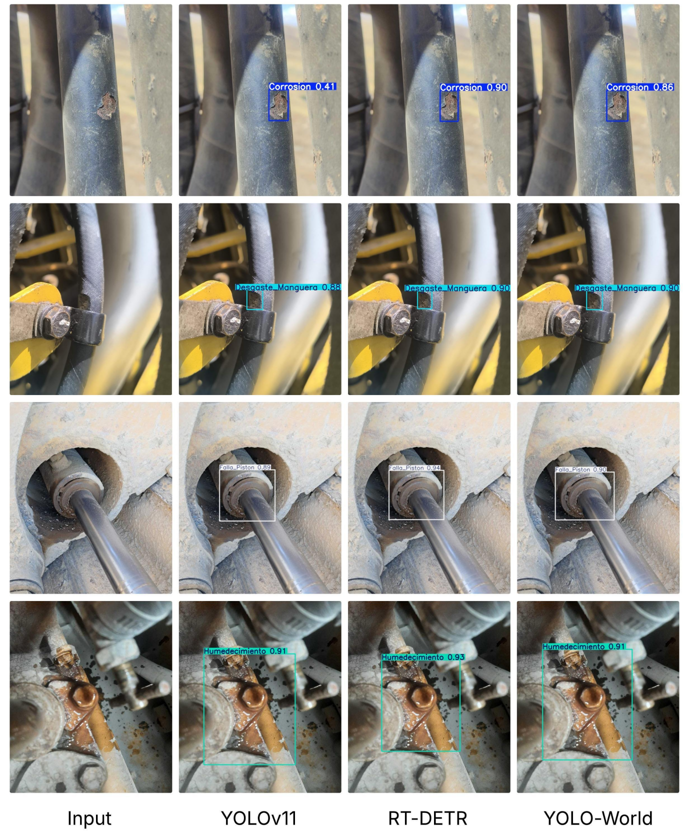

Fig. 12. Detection result of the three models.

Furthermore, the study identifies that certain classes of defects, such as corrosion and moisture, exhibit lower accuracy compared to other detected faults. To address this limitation, it is proposed that future work include a larger number of images of these defects, thereby improving the representation of these classes in the dataset. This strategy, along with other techniques such as data augmentation and model fine-tuning, will contribute to increasing the system's accuracy and robustness, optimizing its performance in predictive maintenance applications.

REFERENCES

[1] M. Pacchiotti and P. Paletto, "Deep learning based software prototype for predictive maintenance in industry 4.0 - 9th national congress of informatics engineering and information systems 2021(conaiisi)," 11 2021.

[2] O. Serradilla, E. Zugasti, and U. Zurutuza, "Deep learning models for predictive maintenance: a survey, comparison, challenges and prospect," 2020. [Online]. Available: https://arxiv.org/abs/2010.03207

[3] Abdeldjalil Latrach, "Application of deep learning for predictive maintenance of oilfield equipment," 2020. [Online]. Available: https://rgdoi.net/10.13140/RG.2.2.12595.09762

[4] J. Hurtado, D. Salvati, R. Semola, M. Bosio, and V. Lomonaco, "Continual learning for predictive maintenance: Overview and challenges," *Intelligent Systems with Applications*, vol. 19, p. 200251, Sep. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.iswa.2023.200251

[5] S. Kang, Z. Hu, L. Liu, K. Zhang, and Z. Cao, “Object detection yolo algorithms and their industrial applications: Overview and comparative analysis,” *Electronics*, vol. 14, no. 6, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/14/6/1104

[6] I. D. Apostolopoulos and M. Tzani, “Industrial object, machine part and defect recognition towards fully automated industrial monitoring employing deep learning. the case of multilevel VGG19,” *CoRR*, vol. abs/2011.11305, 2020. [Online]. Available: https://arxiv.org/abs/2011.11305

[7] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, “Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9584–9592.

[8] O. A. Oyedeji, S. Khan, and J. A. Erkoyuncu, “Application of cnn for multiple phase corrosion identification and region detection,” *Applied Soft Computing*, vol. 164, p. 112008, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494624007828

[9] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, “Towards total recall in industrial anomaly detection,” 2022. [Online]. Available: https://arxiv.org/abs/2106.08265

[10] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, and Y. Jin, “Deep industrial image anomaly detection: A survey,” *Machine Intelligence Research*, vol. 21, no. 1, p. 104–135, Jan. 2024. [Online]. Available: http://dx.doi.org/10.1007/s11633-023-1459-z

[11] D. J. Atha and M. R. Jahanshahi, “Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection,” *Structural Health Monitoring*, vol. 17, no. 5, pp. 1110–1128, 2018. [Online]. Available: https://doi.org/10.1177/1475921717737051

[12] G. Wrat, P. Ranjan, S. K. Mishra, J. T. Jose, and J. Das, “Neural network-enhanced internal leakage analysis for efficient fault detection in heavy machinery hydraulic actuator cylinders,” *Proceedings of the Institution of Mechanical Engineers, Part C*, vol. 239, no. 3, pp. 1021–1031, 2025. [Online]. Available: https://doi.org/10.1177/09544062241289309

[13] Y. Cui, Z. Liu, and S. Lian, “A survey on unsupervised anomaly detection algorithms for industrial images,” *IEEE Access*, vol. 11, p. 55297–55315, 2023. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2023.3282993

[14] J.-H. Kim, D.-H. Kim, S. Yi, and T. Lee, “Semi-orthogonal embedding for efficient unsupervised anomaly segmentation,” 2021. [Online]. Available: https://arxiv.org/abs/2105.14737

[15] T. Defard, A. Setkov, A. Loesch, and R. Audigier, “Padim: a patch distribution modeling framework for anomaly detection and localization,” 2020. [Online]. Available: https://arxiv.org/abs/2011.08785

[16] E. Ferguson, T. Dunne, S. Potiris, V. Vlaskine, J. Mohammed, S. Bargoti, N. Ahsan, and M. Naqshbandi, “Automatic detection and classification of corrosion with convolutional neural networks,” 10 2020.

[17] T. Schlegl, S. Schlegl, N. West, and J. Deuse, “Scalable anomaly detection in manufacturing systems using an interpretable deep learning approach,” *Procedia CIRP*, vol. 104, pp. 1547–1552, 2021, 54th CIRP CMS 2021 - Towards Digitalized Manufacturing 4.0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827121011598

[18] Y.-G. Yoon, J.-H. Woo, and T.-K. Oh, “A study on the application of machine and deep learning using the impact response test to detect defects on the piston rod and steering rack of automobiles,” *Sensors*, vol. 22, no. 24, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/24/9623

[19] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, pp. 1–48, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:195811894

[20] Y. Haobo, “A study on industrial surface defect detection based on deep learning,” *International Conference on Cyberphysical Social Intelligence 2024(ICCSI)*, pp. 1–6, 2024.

[21] R. Khanam and M. Hussain, “Yolov11: An overview of the key architectural enhancements,” 2024. [Online]. Available: https://arxiv.org/abs/2410.17725

[22] W. Lv, Y. Zhao, Q. Chang, K. Huang, G. Wang, and Y. Liu, “Rt-detrv2: Improved baseline with bag-of-freebies for real-time detection transformer,” 2024. [Online]. Available: https://arxiv.org/abs/2407.17140

[23] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, “Yolo-world: Real-time open-vocabulary object detection,” 2024. [Online]. Available: https://arxiv.org/abs/2401.17270