

Computer Vision Real-Time Impurity Detection in Sour Cassava Starch Production

Giovanna C. R. Mesquita, Layla V. S. Teles, Mateus R. Cruz, Samuel B. Mafra , and William O. F. Carvalho 

Abstract— This paper presents a computer vision-based framework for real-time impurity detection in cassava conveyor belts used in starch production. A YOLOv11m model was trained using transfer learning on a dataset collected from artisanal processing units, incorporating data augmentation to enhance generalization. The model achieved 80% precision and over 80% recall, demonstrating reliable detection performance despite class imbalance and visual similarity between cassava and impurities. Real-time validation using a mobile-based workflow confirmed robustness under varying conditions. The proposed approach shows potential for improving process efficiency, product quality, and automation in small- and medium-scale cassava starch industries.

Keywords— Computer vision, YOLOv11m, Real-time detection, Impurity classification.

I. INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is a plant native to South America and is widely cultivated throughout Brazil, ranking among the most consumed food crops worldwide due to its high adaptability to diverse climatic conditions and its broad range of applications [1]. It can be consumed *in natura* or used as a raw material for the production of starches and flours, with applications in the paper, adhesive, paint, meat-processing, and wholesale sectors, including the manufacture of sweet and sour cassava starch and related products [2].

Sour cassava starch is obtained through the natural fermentation of cassava starch, a process that imparts distinctive characteristics such as characteristic flavor and high expansion capacity during baking [3], [4]. Its production relies on starch extraction through unit operations including washing, peeling, grinding, centrifugation, fermentation, and drying, predominantly carried out in small- and medium-scale artisanal processing units [5]. In Brazil, sour cassava starch production is strongly associated with traditional processing systems, with the municipality of Conceição dos Ouros standing out as a consolidated and representative production hub. In this context,

the activity directly and indirectly employs approximately 25% of the local population, and an annual output estimated to be between 13,000 and 15,000 tons has led to the municipality's recognition as the "National Capital of Sour Cassava Starch" [6].

However, among the unit operations involved in the process, grinding constitutes a critical stage, as its efficiency depends directly on adequate raw material selection. The presence of foreign materials such as branches, leaves, soil particles, and, in more severe cases, metallic fragments originating from transportation and unloading can compromise both equipment performance and final product quality. These impurities pose recurrent operational challenges, particularly in artisanal production systems where sorting and discard procedures are largely manual, making grinding a sensitive and failure-prone stage of the industrial process.

In light of these challenges, the application of Computer Vision (CV) techniques emerges as a promising approach to support sour cassava starch production. CV is a subfield of artificial intelligence (AI) that addresses the realization of vision using computers, enabling machines and systems to interpret, analyze, and extract meaningful information from images, videos, and other visual data to support automated actions and decision-making [7]. Recent advances in deep learning have expanded the application of object detection models in agriculture, food inspection, and industrial quality control. A recent systematic review highlights the effectiveness of the You Only Look Once (YOLO) family in agricultural applications, including crop monitoring, disease identification, and automated quality assessment [8]. Likewise, CV has become an important tool for food inspection, enabling the automatic detection of foreign and small objects and impurities, thereby improving product quality and safety with more accurate and reliable quality inspection systems [9], [10], [11].

The automatic, real-time detection of impurities enables the activation of visual or auditory alerts, acting as a trigger for operator intervention prior to the grinder inlet. Within this framework, this work proposes and evaluates a CV-based system for real-time impurity detection in cassava grinding processing, specifically designed for artisanal environments. Unlike previous works, which are mainly focused on laboratory datasets or generic food inspection tasks, this work addresses impurity detection in sour cassava starch production under real industrial operating conditions. The main contribution is the development and validation of a complete CV framework, from the creation of the first dataset for this application collected in an industrial production environment to model training using YOLOv11 transfer learning and deployment through a mobile-based real-time inspection system. CV is capable of identifying heterogeneous foreign materials under industrial operating

G. C. R. Mesquita, L. V. S. Teles, M. R. Cruz, S. B. Mafra, and W. O. F. Carvalho are with National Institute of Telecommunications (Inatel), 37536-001, Santa Rita do Sapucaí, Brazil, e-mail: giovanna.cabral@ges.inatel.br, layla.sousa@ges.inatel.br, mateus.cruz@inatel.br, samuelbmafra@inatel.br, william.orivaldo@inatel.br. This work has been funded by the following research projects: Brasil 6G Project with support from RNP/MCTI (Grant 01245.010604/2020-14), xGMobile Project codes XGM-AFCCT-2024-5-1-2 and XGM-AFCCT-2026-6-18-1 with resources from EMBRAPA/MCTI (Grant 052/2023 PPI IoT/Manufatura 4.0) and FAPEMIG (Grant PPE-00124-23), Ciência por Elas with resources from FAPEMIG (Grant APQ-04523-23), Fomento à Internacionalização das ICTMGs with resources from FAPEMIG (Grant APQ-05305-23), Programa de Apoio a Instalações Multiusuários with resources from FAPEMIG (Grant APQ-01558-24), Redes Estruturantes, de Pesquisa Científica ou de Desenvolvimento Tecnológico with resources from FAPEMIG (Grant RED-00194-23), and Centro de Competência em Redes 5G e 6G Ref. Finep nº 1060/2 contract nº 0-1-25-0883-00. This work has also been supported by undergraduate research scholarships from FAPEMIG.

conditions, supporting operator decision-making during the grinding stage, and demonstrating its potential to enhance equipment protection, process robustness, and final product quality.

II. METHODOLOGY

This section presents the methodology adopted for the collection, preparation, and organization of the dataset used to train the YOLO medium (YOLOv11m) model for impurity detection. YOLOv11m was selected due to its balance between detection accuracy, inference speed, and computational efficiency [8]. Unlike two-stage object detectors, which generally achieve higher detection accuracy at the expense of increased computational cost and longer inference time, the YOLO family performs object detection in a single-stage pipeline, making it more suitable for low-latency scenarios [12]. Compared with other single-stage detectors, YOLOv11m provides improved feature extraction and more robust detection while maintaining a favorable trade-off between speed and accuracy. These characteristics make it suitable for real-time impurity detection system on cassava grinder inlet.

A. Data Collection and Preparation

Technical visits were conducted at cassava starch factories located in Conceição dos Ouros - MG, where footage of the cassava selection conveyor belts was recorded. A total of approximately 200 minutes of video at 60 fps was obtained, however, only those segments featuring an adequate quantity and diversity of samples were selected for the dataset. The recordings were captured using a mobile device camera from different angles and under varying lighting conditions to ensure greater visual diversity for the dataset. This variation enhanced the model's generalization capability, allowing it to operate more robustly across different industrial scenarios. Following the filming, the videos were segmented using the Roboflow platform and converted into frames to extract the images that comprise the dataset used in this study.

B. Dataset

Image annotation was performed by labeling the cassavas and residues using the bounding box tool, as shown in Fig. 1. Two classes were defined for labeling: (i) wood, representing the primary residue identified on the conveyor belts and (ii) cassava, included to enable the model to correctly distinguish the food from impurities. The background, in turn, does not correspond to an actual class in the dataset. Instead, it is used during model evaluation to represent undetected objects or detections that do not match any valid annotation. When an object present in an image is not detected by the model, or when the model detects a non-existent object that does not correspond to any class annotation, these occurrences are counted as false negatives and associated with the background category.

Following the labeling stage, 1,094 annotated images were obtained, and data augmentation techniques were applied, such as 90° rotation, resizing, and brightness variation, to increase dataset diversity and mitigate potential overfitting issues. As

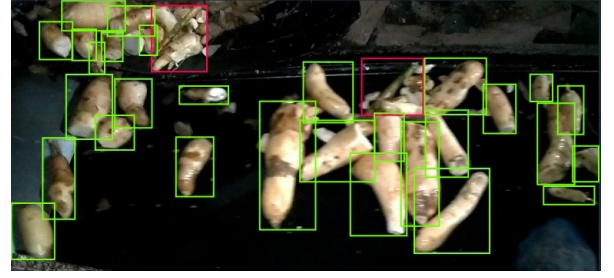


Fig. 1: Frame for labeling bounding for cassava (green boxes) and wood (red boxes) for CV model training.

a result, a total of 1,685 images were generated and split into proportions of 70% for training, 20% for validation, and 10% for testing. This distribution follows established practices in the literature and allows for a proper evaluation of the model's generalization capability. Subsequently, the dataset was used to train the YOLOv11 model, enabling the adaptation of pre-trained weights to the specific classes of this study. It is noteworthy that the classes exhibit distinct sample quantities, as shown in Fig. 2. It is noteworthy that the cassava class contains approximately 12.55 times more samples than the wood class. This class imbalance is an inherent characteristic of the proposed application rather than a limitation of the dataset. In the sour cassava starch production process, the conveyor belt is expected to transport predominantly cassava, while wood fragments and other impurities occur only sporadically. Therefore, the larger number of cassava samples accurately reflects the real industrial operating conditions and was intentionally preserved during dataset construction. As a consequence better representation of cassava samples is evidenced, whereas the reduced number of wood samples makes impurity detection more challenging and contributes to the greater number of misclassifications.

C. Model Training and Validation

After creating the dataset on the Roboflow platform, the data were uploaded to Kaggle for model training. Due to the availability of established libraries specifically for CV applications, the Python programming language was chosen. The necessary libraries for model development were imported, including the Ultralytics library, which is responsible for implementing the YOLO architecture, specifically version 11.

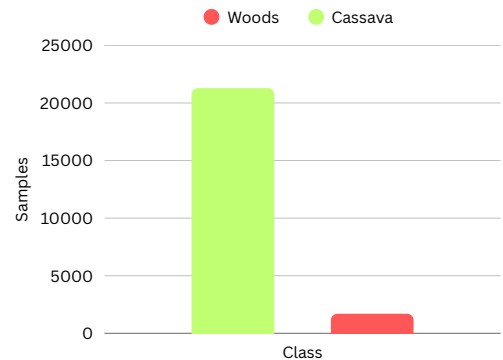


Fig. 2: Number of samples by class.

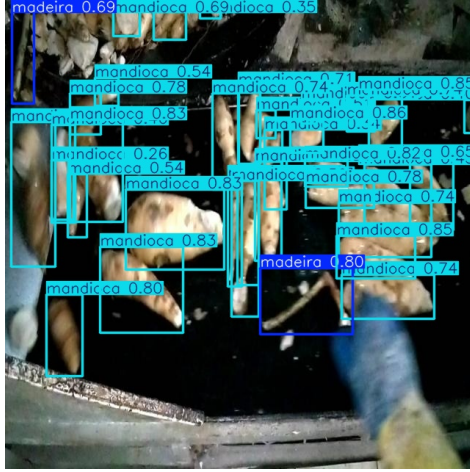


Fig. 3: Prediction Image.

Model training was conducted using the transfer learning technique [13]. In this context, a YOLOv11 model previously trained on a large, generic dataset was used as a starting point. Starting from this pre-trained model, fine-tuning was performed using specific images from the cassava selection process. This approach allows for leveraging visual features learned in previous training stages, reducing the need for large volumes of specific data and making the process computationally more efficient [14]. The training was configured for 150 epochs, and no significant overfitting was observed in this setup. In each epoch, the model fully traverses the training dataset, performing relevant feature extraction, loss function calculation, and iterative neural network weight adjustment through optimization techniques. At the end of each epoch, the model's performance is evaluated using the validation set, which does not directly participate in the weight adjustment stage. During this phase, metrics such as mean Average Precision (mAP), precision, and recall are calculated for monitoring the evolution of learning and verifying the model's generalization capability, with the goal of minimizing detection errors for impurities present on the cassava conveyor belt.

III. RESULTS AND DISCUSSION

Following the completion of the training and validation stages, the best-performing model corresponding to the *best.pt* file generated at the end of the process was selected for the prediction phase. In this stage, the trained model is applied to images from the test set, which were not used during training, to evaluate its performance with unseen data. Fig. 3 presents an example of this inference stage, where the model automatically identifies regions of interest in the images, generating bounding boxes, classifying the detected objects, and assigning a confidence score to each prediction. These annotations allow for both quantitative analysis through performance metrics and qualitative visual analysis of the results obtained by the proposed system.

Among the performance metrics for the YOLOv11 model, we utilized the confusion matrix and the performance graph. The confusion matrix enables the analysis of characteristics related to both the correct localization of residues in the image and their classification by the model, while the performance graph

allows for monitoring the evolution of the model's learning throughout the training process.

In the confusion matrix illustrated in Fig. 4, it is observed that the model shows high performance in detecting the cassava class (mandioca), which was expected given the large number of samples for this class in the dataset. However, in some instances, cassava is confused with the background, indicating detection failures possibly associated with poor lighting or low image contrast. Regarding the wood class (madeira), although the model can identify it, a higher number of errors is observed compared to the cassava class. Since the model was exposed to substantially more cassava samples during training, it learned more representative visual features for this class. Conversely, the limited number of wood samples reduced the variability of impurity patterns available during training, contributing to the higher number of misclassifications, particularly when wood fragments exhibit visual characteristics similar to cassava. Furthermore, the model tends to classify background areas as cassava, indicating higher sensitivity in detecting this class, which results in the occurrence of false positives.

Fig. 5 presents the performance metric graphs obtained after the model training. Each graph allows for an analysis of whether the learning process occurred appropriately throughout the 150 epochs. In the first column of graphs, the *train/box_loss* and *val/box_loss* curves (Figs. 5(a)-(b)) can be observed, which measure the error in bounding box prediction; the slope and progressive reduction of these curves indicate that the model is improving object localization in the images. Similarly, the *train/cls_loss* and *val/cls_loss* curves (Figs. 5(c)-(d)), responsible for measuring object classification error, also show a decreasing behavior, evidencing that the model is learning to correctly classify impurities. The *train/dfl_loss* and *val/dfl_loss* metrics (Figs. 5(e)-(f)), used by YOLO to refine bounding box positioning relative to objects, likewise decrease throughout the training, indicating a continuous improvement in box fitting precision. It is also noted that the training and validation curves exhibit similar behaviors, suggesting good model generalization capability and the absence of significant overfitting.

Furthermore, the performance metrics show a positive evolution. The *precision* metric (Fig. 5(g)) ranges from 0 to 1, where higher values indicate a greater proportion of correct detections relative to the total detections made by the model. Throughout the training, this metric shows progressive

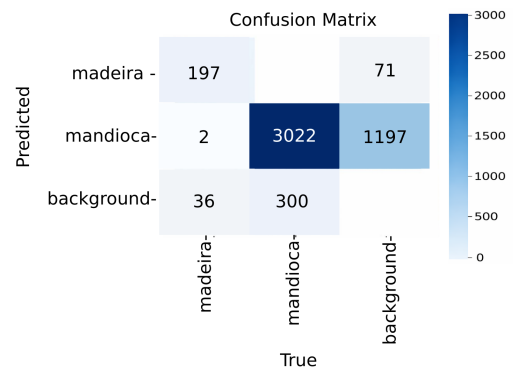


Fig. 4: Confusion Matrix.

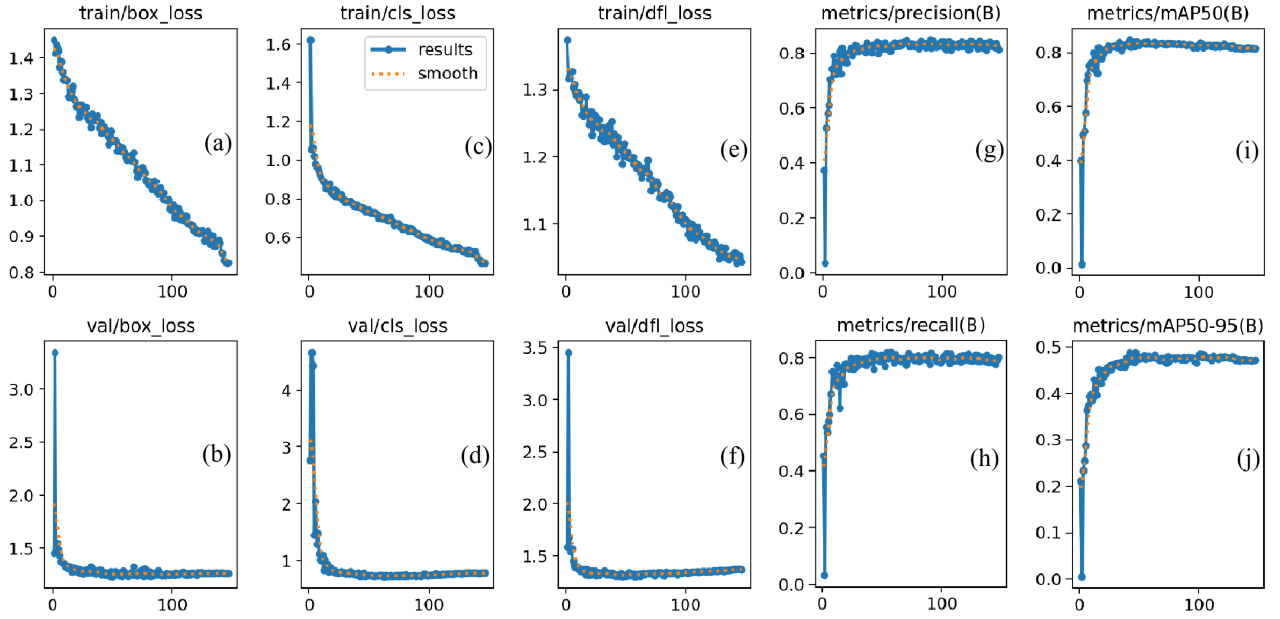


Fig. 5: Performance metrics as a function of the number of epochs.

growth, indicating a reduction in incorrect detections. After a certain number of epochs, the curve reaches values close to 0.8, meaning approximately 80% of the detections made are correct. In addition, the metric stabilizes at this level without significant drops, indicating consistency and stability in the model's predictions over the epochs. The *recall* metric (Fig. 5(h)) also ranges from 0 to 1 and represents the proportion of real objects correctly detected. It is noteworthy that this metric increases over the epochs, reaching high values and remaining stable to 0.8, which indicates that the model is capable of identifying most of the objects present in the images.

Finally, the *mAP50* and *mAP50-95* metrics (Figs. 5(i)-(j)) indicate improvements in detection quality. The first one relates to the correct delimitation of at least 50% of the object by the bounding box, while the latter evaluates box precision across different rigor levels, from 50% to 95%. Their increasing behavior and subsequent stabilization indicate training convergence and suggest that the number of epochs adopted was appropriate for the studied problem.

Given the quantitative performance observed throughout the training, it is also necessary to evaluate the model's behavior under conditions closer to a real-world scenario. Thus, beyond the analysis of learning metrics, a practical validation stage was conducted to assess the model's capability for real-time detection using images captured directly by a mobile device. This validation was performed using the Roboflow Workflow tool, which enables real-time model deployment through a mobile device's camera. Initially, the dataset was trained on the Roboflow platform to obtain a model hosted within its environment. Subsequently, a Workflow was created to integrate the trained model with a real-time image acquisition system.

Fig. 6 illustrates the workflow adopted for real-time object detection in the production environment. Initially, images are continuously acquired from a camera positioned over the conveyor belt, capturing the operating conditions and the presence of cassava alongside potential impurities. These images are subsequently processed by a detection model based

on the YOLO architecture, trained using a labeled dataset via Roboflow, which enables the extraction of discriminative visual features and the identification of impurity patterns. For each input frame, the model outputs the predicted classes, corresponding bounding box coordinates, and associated confidence scores. Based on these predictions, bounding boxes are superimposed onto the original image to localize the detected objects and indicate their spatial extent within the field of view. In parallel, textual annotations are added to each detection, indicating the class label and confidence score, thereby enhancing interpretability and supporting decision-making. Finally, the system generates a comprehensive output comprising the total count of detected impurities, the annotated images, and the structured detection data, including class labels and confidence metrics, which can be further utilized for monitoring and analysis of the production process.

After assembling the flowchart, the "Run From Anywhere" functionality generates a QR code for application access. By scanning this QR code with a mobile device, it executes the detection model through the smartphone camera, allowing for real-time testing and validation. This approach facilitates verifying the model's performance in different environments without the need for additional local integration or specific software installation. Fig. 7 presents representative results obtained during real-time detection. Fig. 7(a) illustrates the operational concept of the system, in which input frames captured by a mobile device are processed via a cloud-based workflow, and the corresponding annotated output frames are returned to the user interface. Figs. 7(b)-(d) show examples of detection outcomes in the production environment, where the model

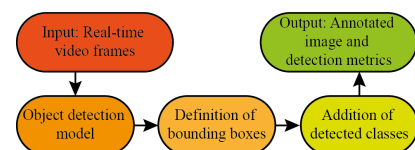


Fig. 6: Workflow Flowchart.

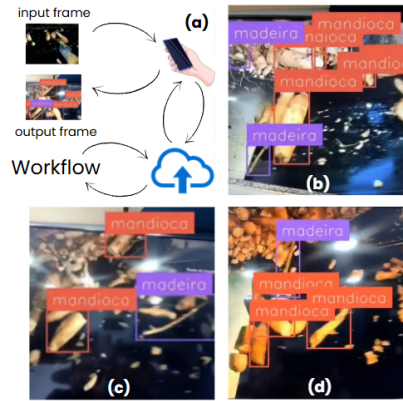


Fig. 7: Real Time Detection Images.

successfully identifies cassava (“mandioca”) and impurities (“madeira”). In these images, bounding boxes are overlaid to indicate the spatial localization of each detected object, while color-coded labels provide the predicted class for each instance. The results demonstrate the model’s capability to operate under practical and real-time conditions by accurately detecting and classifying objects despite variations in illumination, object positioning, and scene complexity.

While the real-time laboratory experiments demonstrate the feasibility and robustness of the proposed approach, further developments are envisaged to transition the system toward a fully deployable solution in industrial environments. In particular, the trained detection model is intended to be embedded into a compact edge-computing platform, such as a Raspberry Pi, directly interfaced with a camera mounted above a conveyor belt. This configuration will enable on-site low-latency inference without reliance on external processing infrastructure. As illustrated in Fig. 8, the system will be designed to operate continuously, automatically identifying cassava and residual impurities (wood, clod of clay, steel fragments, etc.) and triggering an immediate auditory alert whenever contamination is detected, in order to assist operators in real-time decision-making and removal actions.

IV. CONCLUSIONS

This study presented an automated framework for real-time impurity detection based on CV applied to cassava conveyor



Fig. 8: Model Implementation Image.

belts in the starch production process. The proposed YOLO-based model, trained with a transfer learning strategy, demonstrated reduced training time and improved generalization. Although limitations were observed in distinguishing the “wood” class, mainly due to visual similarity to cassava and dataset imbalance, the system achieved a detection precision of 80% and a recall above 80%, indicating reliable identification of most objects in the scene. In addition, stable performance across training epochs confirmed the robustness of the model. The system also demonstrated effective real-time operation under practical conditions, supporting its applicability in artisanal and small-scale industrial environments. Future work includes dataset expansion, improved class balance, and deployment on embedded edge-computing platforms for low-latency on-site operation. Overall, the proposed approach represents a relevant step toward the integration of CV in the starch industry, with potential benefits in automation, quality control, and operational efficiency.

REFERENCES

- [1] D. G. Martinez, “Produção de etanol de segunda geração a partir de resíduos do processamento da mandioca,” M.Sc. dissertation, Universidade Estadual do Oeste do Parana, Cascavel, SC, Brazil, 2016.
- [2] V. A. Balbino and E. M. Lima, “The strategic cost management in small rural agricultural family industries: a case study in a cassava agribusiness in Caarapó/MS,” *Custos e agronegocio on line*, vol. 11, p. 203–223, 2015.
- [3] M. C. R. d. Silva, “Pão de queijo: fundamentos práticos para produção,” M.Sc. dissertation, Universidade Federal de Uberlândia, 2021.
- [4] M. P. Cereda and V. H. Brito, “Modificação de amido por fermentação – polvilho azedo,” in *Coleção Biotecnologia Industrial. Volume 3: Processos Fermentativos e Enzimáticos*, 2nd ed. São Paulo: Blucher, 2017, ch. 16, pp. 509–546.
- [5] A. C. A. Carvalho and F. A. Bastos, “Análise do processo de fabricação do polvilho azedo em Conceição dos Ouros–Minas Gerais,” *Revista de Ciências Agro-Ambientais*, vol. 22, no. 1, pp. 27–38, 2024.
- [6] Prefeitura Municipal de Conceição dos Ouros, “Conheça conceição dos ouros,” <https://conceicaodosouros.mg.gov.br/conceicao-dos-ouros/>, Prefeitura Municipal de Conceição dos Ouros, 2025, access in : 17 dez. 2025.
- [7] Y. Matsuzaka and R. Yashiro, “AI-based computer vision techniques and expert systems,” *AI*, vol. 4, no. 1, pp. 289–302, 2023.
- [8] C. M. Badgujar, A. Poulouse, and H. Gan, “Agricultural object detection with You Only Look Once (YOLO) algorithm: A bibliometric and systematic literature review,” *Comput. Electron. Agric.*, vol. 223, p. 109090, 2024.
- [9] H. Bagherpour and N. F. Peyruo, “Enhanced YOLO-based framework for accurate detection and identification of common wheat impurities with distinct objects,” *Sci. Rep.*, 2025.
- [10] J. Ge, Q. Sun, XinZhou, Z. Zhang, Z. Tang, H. Wang, Z. Zhu, S. Fan, M. Guo, and J. Xian, “Research on foreign object detection of overhead transmission lines based on optimized YOLO11,” in *2025 12th International Forum on Electrical Engineering and Automation (IFEAA)*, 2025, pp. 743–747.
- [11] C. S. Rao, C. M. Rao, R. Amardeep, K. V. N., and P. B. Edwin Prabhakar, “Hybrid multi-scale attention fusion network for vision-based food quality inspection in smart processing industries,” in *2026 4th International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2026, pp. 1–8.
- [12] G. Jocher and J. Qiu, “Ultralytics YOLO11,” 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [13] Ultralytics, “Transfer learning,” <https://www.ultralytics.com/glossary/transfer-learning>, Ultralytics, access in : 09 Mar 2026.
- [14] Dave Bergmann, “O que é ajuste fino?” <https://www.ibm.com/br-pt/think/topics/fine-tuning>, IBM, access in : 09 Mar 2026.