e-ISSN 2540-9840 p-ISSN 1412-3770 Jurnal Ilmiah Sains, October 2025, 25(2): 173-186
Accredited by Ministry of Education, Culture, DOI: https://doi.org/10.35799/jis.v25i2.64398
Research and Technology No: 105/E/KPT/2022 Available online at https://ejournal.unsrat.ac.id/v3/index.php/jis

Jamur pada Produk Pangan: Analisis Komparatif Varian YOLO untuk Deteksi *Rhizopus stolonifera* pada Roti

Vanny Hani Siwi[1], Jonathan Wuntu[2*], Norrytha Lineke Wuntu[3], Audy D. Wuntu[4]

[1] Faculty of Agriculture, Pembangunan University of Indonesia Manado, Indonesia
[2] Faculty of Engineering, Sam Ratulangi University, , Jl. Kampus Kleak, Manado, Indonesia
[3] Faculty of Animal Husbandry, Sam Ratulangi University, Jl. Kampus Kleak, Manado, Indonesia
4)Department of Chemistry Faculty of Mathematics and Natural Sciences Sam Ratulangi University, Jl. Kampus Kleak, Manado, Indonesia

*Corresponding author: jonathanwuntu026@student.unsrat.ac.id

ABSTRAK

Roti merupakan makanan pokok yang sangat rentan terhadap kontaminasi jamur, khususnya oleh Rhizopus stolonifer, yang dapat menimbulkan risiko serius terhadap kesehatan dan keamanan pangan. Deteksi dini dan akurat terhadap pertumbuhan jamur sangat penting untuk mencegah kerusakan serta menjaga keamanan konsumen. Penelitian ini menyajikan analisis komparatif terhadap beberapa varian terbaru YOLO (You Only Look Once), yaitu YOLOv8n, YOLOv10n, YOLO11n, dan YOLOv12n, dalam mendeteksi jamur Rhizopus stolonifer pada permukaan roti. Penelitian ini menggunakan dataset deteksi jamur yang bersumber dari platform Roboflow, yang berisi gambar roti beranotasi yang diambil dalam berbagai kondisi pencahayaan, tekstur, dan kontaminasi untuk mendukung pelatihan model yang optimal. Setiap varian YOLO dilatih dan dievaluasi dengan hiperparameter yang konsisten guna memastikan keadilan dalam perbandingan. Hasil eksperimen menunjukkan bahwa YOLOv8n memperoleh nilai mAP50 sebesar 0,472 dan mAP50:95 sebesar 0,203; YOLOv10n sebesar 0,474 dan 0,191; YOLO11n sebesar 0,504 dan 0,204; serta YOLOv12n sebesar 0,503 dan 0,224. Di antara varian tersebut, YOLO11n menunjukkan performa mAP50 tertinggi, sedangkan YOLOv12n mencapai nilai mAP50:95 terbaik, yang menandakan konsistensi deteksi yang lebih baik pada berbagai ambang IoU. Temuan ini menunjukkan bahwa arsitektur YOLO terbaru memiliki potensi yang menjanjikan untuk deteksi jamur Rhizopus stolonifer pada roti secara otomatis dan waktu nyata, sehingga dapat mendukung pengembangan sistem pemantauan keamanan pangan yang cerdas.

Kata kunci: *Deep learning;* deteksi roti jamur; *Rhizopus stolonifer*; *k*eselamatan pangan; pendeteksian objek YOLO

Mold on Food Product: Comparative Analysis of YOLO Variants for Detecting *Rhizopus stolonifer* on Bread

ABSTRACT

Bread is a staple food that is highly susceptible to fungal contamination, particularly by Rhizopus stolonifer, which poses significant health and food safety risks. Early and accurate detection of mold growth is essential to prevent spoilage and ensure consumer safety. This study presents a comparative analysis of recent YOLO (You Only Look Once) variants, YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n for detecting Rhizopus stolonifer mold on bread surfaces. This study utilized a mold detection dataset sourced from the Roboflow platform, which contains annotated bread images captured under diverse lighting, texture, and contamination conditions to support robust model training. Each YOLO variant was trained and evaluated under consistent hyperparameters to ensure fairness in comparison. Experimental results indicate that YOLOv8n achieved an mAP50 of 0.472 and mAP50:95 of 0.203; YOLOv10n achieved 0.474 and 0.191, respectively; YOLO11n achieved 0.504 and 0.204; and YOLOv12n achieved 0.503 and 0.224. Among these, YOLO11n demonstrated the highest mAP50 performance, while YOLOv12n attained the best mAP50:95 score, indicating superior detection consistency across varying IoU thresholds. These findings suggest that recent YOLO architectures offer promising potential for real-time and automated detection of Rhizopus stolonifer mold in bread, supporting advancements in intelligent food safety monitoring systems.

Keywords: Bread mold detection; deep learning; food safety monitoring; *Rhizopus stolonifer*; YOLO object detection

(Article History: Received 08-10-2025; Accepted 31-10-2025; Published 31-10-2025)

INTRODUCTION

Bread is a globally important staple whose physicochemical composition (moderate water activity, starches, and sugars) and processing make it vulnerable to microbial spoilage, particularly fungal colonization, which shortens its shelf life and creates economic losses for producers and retailers (Liu *et al.*, 2022). Bio-preservation and other mitigation strategies have been widely studied to reduce fungal spoilage while meeting consumer demand for fewer chemical preservatives (Rahman *et al.*, 2022).

Among fungi that spoil bakery products, Rhizopus species (commonly reported as *Rhizopus stolonifer*, also known as "black bread mold") are frequently observed. They can rapidly colonize bread surfaces under favorable humidity and temperature conditions. Such colonization can produce visible mycelia and, in some contexts, mycotoxins or secondary metabolites that raise concerns about food safety (Rahman *et al.*, 2022). Recent reviews synthesize the biology, detection challenges, and control strategies for *R. stolonifer* in postharvest and food contexts (Q. Liu *et al.*, 2024).

Conventional detection and control methods in bakeries, such as visual inspection, culture-based identification, and laboratory assays, remain important but are often slow, labor-intensive, and insensitive to very early contamination (Q. Liu *et al.*, 2024). This has motivated research into non-destructive, image-based screening and automation methods that can detect early surface anomalies before they lead to gross spoilage. Natural and biological control approaches are also being explored as complementary strategies to reduce fungal incidence (Ribes *et al.*, 2018).

In computer vision, deep convolutional neural networks (CNNs), particularly the YOLO family of one-stage detectors, have demonstrated real-time detection capabilities suitable for production-line inspection (Redmon *et al.*, 2015). Recent YOLO variants, YOLOv8, YOLOv10, YOLO11, and YOLOv12, incorporate architectural refinements specifically targeting small-object detection and inference efficiency. However, comparative evaluations of these modern YOLO variants for detecting early-stage fungal contamination on bread surfaces remain absent from the literature. While these models show promise for identifying subtle mold spots, no systematic study has benchmarked their relative performance on fungal-contamination datasets, leaving practitioners without empirical guidance for selecting optimal detection architectures for automated bakery inspection systems. This gap is critical because different YOLO versions may exhibit distinct trade-offs between detection accuracy and computational efficiency when applied to the specific visual characteristics of *R. stolonifer colonization*.

Recent literature has documented the rapid evolution of YOLO architectures, with comprehensive reviews tracing developments from YOLOv1 through YOLOv12 and benchmarking studies revealing distinctive performance trade-offs across versions (Jegham *et al*., 2024). In food quality monitoring, deep learning has shown promising results: YOLOv5 achieved 98.10% precision and 100% recall for mold detection across multiple food types, including bread (Jubayer *et al*., 2021). At the same time, thermal imaging combined with YOLO11n demonstrated an mAP50-95 of 0.607 for bread contamination detection (Madasamy Raja *et al*., 2025), and CNN-based transfer learning enabled early-stage microscopic mold detection suitable for smartphone deployment (Treepong & Theera-Ampornpunt, 2023). However, existing studies have focused on earlier YOLO versions (YOLOv3-YOLOv5), employed specialized imaging modalities (thermal, microscopic), or addressed general mold detection without species-specific focus on *Rhizopus stolonifer*. Furthermore, no systematic comparison of modern nano YOLO variants (YOLOv8n, YOLOv10n, YOLO11n, YOLOv12n) exists for conventional RGB-based bread mold detection, leaving a critical gap in understanding their relative performance and suitability for edge-based bakery inspection systems under practical deployment constraints.

Practical deployment of automated inspection systems in bakery environments necessitates consideration of computational constraints and infrastructure requirements. The nano (n) variants of YOLO models are engineered explicitly for resource-constrained environments, enabling deployment on edge devices such as embedded systems, industrial cameras, or mobile processors without dependence on cloud computing infrastructure. Edge-based detection offers critical advantages for bakery production lines: reduced latency for real-time decision-making, lower operational costs by eliminating cloud service fees, enhanced data privacy by keeping proprietary production data on-premises, and reliability in environments with limited or unstable internet connectivity. Despite these practical benefits, the comparative performance of nano YOLO variants (YOLOv8n, YOLOv10n, YOLO11n, YOLOv12n) for bread mold detection has not been systematically evaluated, leaving a knowledge gap regarding which architecture best balances detection accuracy with edge-deployment feasibility for automated food safety applications.

To fill this gap, the present study conducts a systematic comparative analysis of YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n for detecting *Rhizopus stolonifer* on bread images collected and annotated via the Roboflow platform. Each model was trained with consistent hyperparameters and evaluated using standard detection metrics (mAP50 and mAP50:95) to determine relative strengths in accuracy and robustness. The goal is to identify which modern YOLO variant best balances detection accuracy and inference practicality for real-time, automated bakery inspection systems, and to provide empirical guidance for applied food-safety deployments. This study makes three key contributions:1) First systematic benchmark: Provides the first comparative evaluation of modern nano YOLO variants (YOLOv8n, YOLOv10n, YOLO11n, YOLOv12n) specifically for early-stage fungal contamination detection on bread surfaces; 2) Edge-deployment feasibility assessment: Evaluates the accuracy-efficiency trade-offs of lightweight architectures suitable for resource-constrained bakery inspection systems; 3) Empirical guidance for practitioners: Offers data-driven recommendations for selecting optimal YOLO architectures in real-time, automated food safety applications.

RESEARCH METHOD

Research Procedure

This research employed a systematic methodology to ensure a comprehensive evaluation and implementation of four YOLO model variants, YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n, for bread mold detection applications. The investigation commenced with an extensive literature review and architectural analysis of the YOLO family to establish a robust theoretical foundation regarding one-stage object detection systems and their evolution toward real-time performance and enhanced accuracy. Subsequently, a dedicated mold detection dataset was collected and refined using the Roboflow platform, containing high-resolution images of bread samples with visible and early-stage fungal growth to represent diverse contamination patterns and lighting conditions.

The experimental framework involved a rigorous configuration of the hardware and software environments, optimized for deep learning model training, including GPU acceleration and efficient memory utilization, which are essential for high-throughput image processing. Following the infrastructure setup, a unified training pipeline was developed for all YOLO models, integrating robust data augmentation strategies and hyperparameter tuning to enhance model generalization and stability during the learning process.

Each YOLO variant was trained under consistent conditions to ensure fairness in comparison, with specific attention given to convergence behavior, loss patterns, and inference speed. After the training phase, the models were deployed and tested on unseen validation data that simulated realistic conditions for bakery inspections. This stage enabled an accurate performance evaluation using standardized metrics such as mAP50 and mAP50:95, allowing the assessment of each model's detection precision and robustness against visual variability. The methodological framework ensured that the comparative analysis reflected both practical feasibility and scientific rigor, ultimately identifying the YOLO model variant most effective for automated bread mold detection in food-quality control environments.

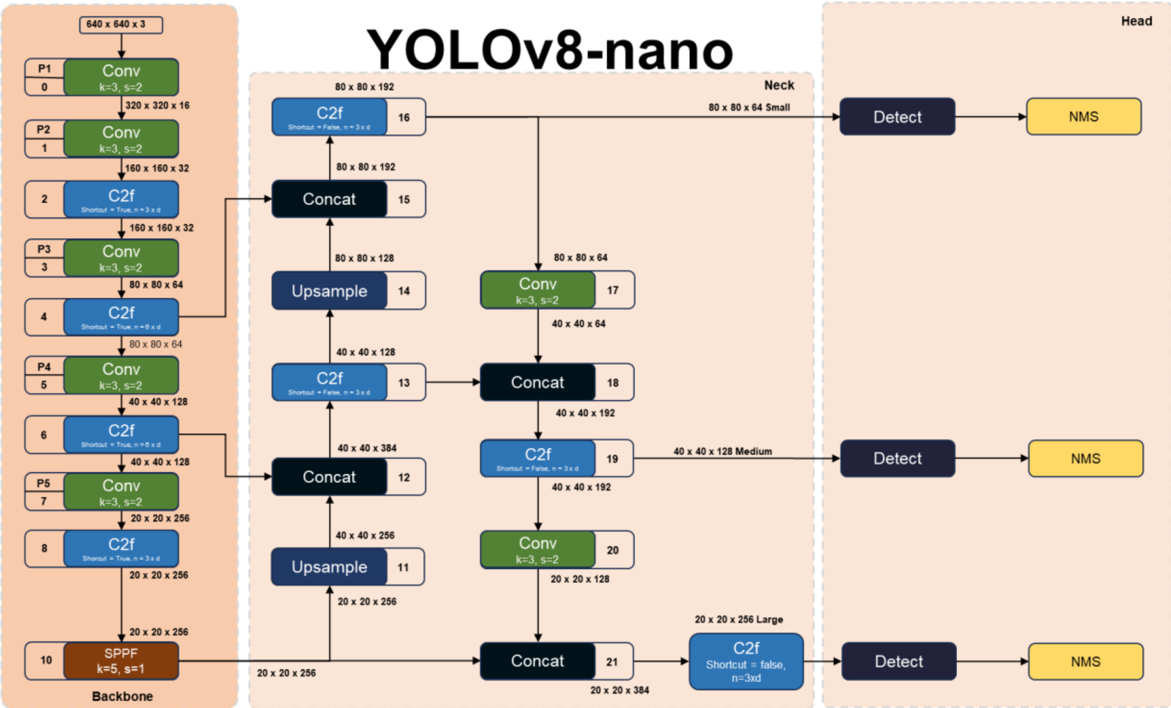

Figure 1. YOLOv8-nano architecture

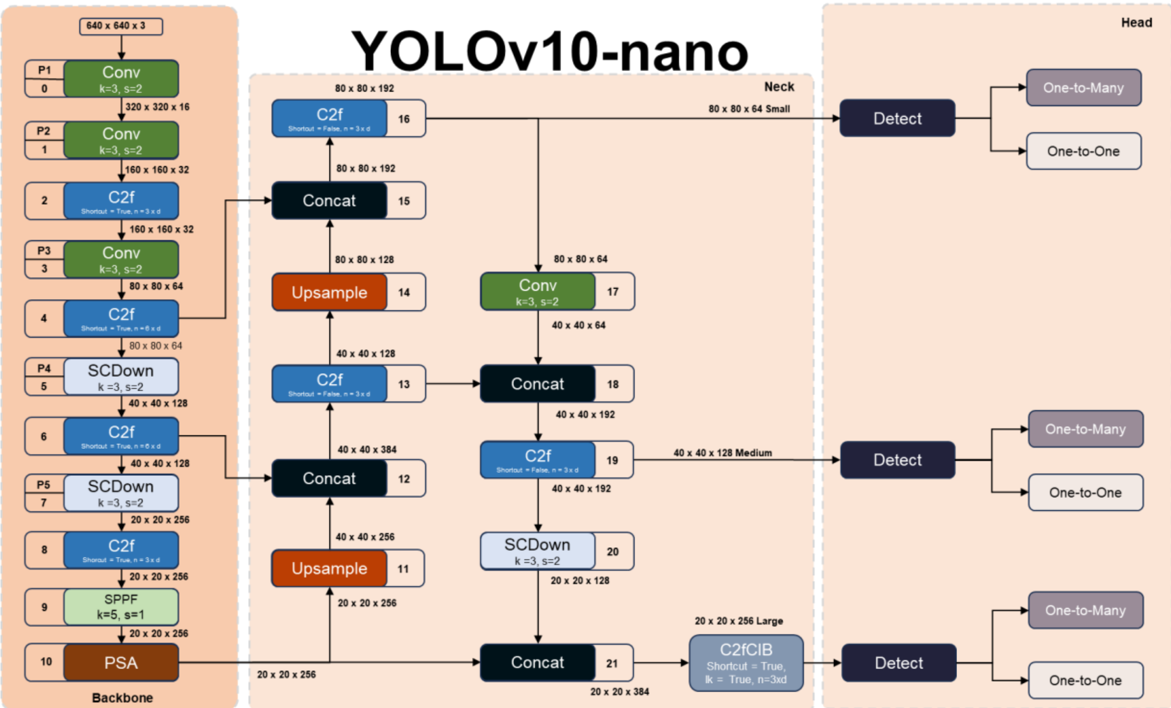

Figure 2. YOLOv10-nano architecture

YOLOv8

YOLOv8 was released in 10th of January 2023, by (*Explore Ultralytics YOLOv8 - Ultralytics YOLO Docs*, n.d.) Based on its predecessor, YOLOv4 (Bochkovskiy *et al*., 2020), it incorporates the use of PAN (Path Aggregation Network) architecture (S. Liu *et al*., 2018) for its neck architecture to merge as a single-stage detector from its backbone (central extractor), as shown in Figure 1. The introduction of the C2f (Convolution 2 fast) module is based on the CSP module (C. Y. Wang *et al*., 2019), which adds a bottleneck and two fast connections (or residuals within the C2f) module (He *et al.*, 2015), improving its computational efficiency while maintaining the accuracy of its overall model. Furthermore, the main difference between YOLOv8 and its predecessor is the use of anchor-free detection on its head. Anchor-free means it eliminates the reliance on pre-determined anchor points for each cell on the prediction image. The loss function that was used on YOLOv8 is CIoU (Complete Intersection over Union) Loss (Zheng *et al.*, 2020) for the IoU bounding boxes loss, DFL (Distributional Focal Loss) (Li *et al.*, 2020) for the distributional bounding boxes loss within an instance (one's image predictions), and BCE (Binary Cross Entropy) for its objectness loss. The CIoU loss is expressed as follows:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}, \tag{1}$$

$$L_{CIoU} = 1 - IoU + \frac{\rho(B, B^{gt})}{c^2} + \alpha v, \tag{2}$$

$$\alpha = \begin{cases} 0, & IoU < 0.5 \\ \frac{v}{(1 - IoU) + v}, & IoU \geq 0.5 \end{cases}, \tag{3}$$

$$v = \frac{4}{\pi^2}\left(arctan\,\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{4}$$

where, B is the bounding boxes that were generated by the predictions of YOLO, B^{gt} is the ground truth bounding boxes, $\frac{\rho(B,B^{gt})}{c^2}$ is the Euclidean distance between the ground truth bounding boxes and the prediction bounding boxes, v is the aspect ratio of the width and height for the bounding box of ground truth and prediction, and α is the balancing parameter to keep v normalize between 0 and 1. The calculation of the total loss of YOLOv8, so it can do a backward pass to update its weights, is formulated as follows:

$$L_{total} = \lambda_{CIoU}L_{CIoU} + \lambda_{DFL}L_{DFL} + \lambda_{BCE}L_{BCE}, \tag{5}$$

where, λ is the balance parameter that automatically updates after every epoch for each loss.

YOLOv10

YOLOv10 was released on the 23rd of May, 2024 (A. Wang *et al.*, 2024). It introduces a new attention mechanism called the PSA (Partial Self-Attention) module, which is inspired by Vision Transformer (Dosovitskiy *et al.*, 2020), as well as C2fCIB, which combines the C2f module with a new module called CIB (Compact Inverted Bottleneck), inspired by the Inverted Bottleneck in MobileNetv2 (Sandler *et al.*, 2018). Furthermore, it also added a new method to eliminate discrepancies in bounding box predictions without the use of NMS (Non-Maximum Suppression) (Hosang *et al.*, 2017) or NMS-free predictions. It utilizes a novel module, known as the One-to-One and One-to-Many prediction method, to replace the primary function of NMS, which is shown in the head section of Figure 2. The Loss function in YOLOv10 is the same as that in YOLOv8, other than the novel Loss function for calculating the loss between the One-to-One and One-to-Many (A. Wang *et al.*, 2024).

YOLO11

YOLO11 was released on the 27th of May, 2024. YOLO11, developed by Ultralytics, is based on YOLOv8, with the novelty of replacing the C2f module within YOLOv8 with a novel C3k2 module, which further improves its efficiency and increases performance accuracy. Furthermore, YOLO11 incorporates a novel attention mechanism called C2PSA (2 Convolution Partial Self-Attention) placed after the SPPF module, similar to that of YOLOv10, to further improve the extraction capabilities of the primary extractor (backbone). The Loss function in YOLO11 is the same as that in YOLOv8.

YOLOv12

YOLOv12 was released on the 18th of February, 2025 (Tian *et al*., 2025). YOLOv12 is based on YOLO11, with the novelty of replacing the existing C3k2 module with a novel A2C2f (Attention Area C2f) module, which incorporates an attention mechanism. Then, it was placed in a few locations inside the backbone and neck, replacing C3k2 and completely removing the attention mechanism of C2PSA, as well as the SPPF module from YOLO11. The incorporation of A2C2f aims to further enhance the feature map at the end of its primary extractor (backbone), extending all the way to the neck, thereby improving all missing features that have been affected by vanishing gradients in deep neural networks. The Loss function in YOLOv12 is the same as that in YOLOv8

Dataset

The dataset used in this research is publicly available on the Roboflow platform (gopletzzz, 2025). The train set comprises approximately 764 images, the validation set contains 87 images, and the test set comprises 23 images, totaling 874 images. The mold dataset contains only one class, called "bread_mold". Furthermore, when trained on YOLOv8, YOLOv10, YOLO11, and YOLOv12, the pre-existing architecture of Ultralytics on YOLOv8 already has a built-in feature that augments data to improve variation for better training.

Training Configuration

Table 1. Training Configuration

Parameters	Configuration
Platform	Kaggle
GPU	NVIDIA Tesla P100
Image Size	640 x 640
Epochs	300
Batch Size	32
Optimizer	Stochastic Gradient Descent
Learning Rate	0.01

The training configuration for this research was conducted on Kaggle's cloud platform. The configurations are running on the NVIDIA Tesla P100 GPU, with an input image size of 640 × 640 and a batch size of 32. Running for 300 epochs with an optimizer of Stochastic Gradient Descent (SGD) with the default learning rate of YOLO 0.01, as shown in Table 1.

Evaluation Metrics

In this study, the evaluation metrics used include F1 score, precision (P), recall (R), mean average precision (mAP). Additionally, the number of parameters (Parameters) was also considered. The formulas for these metrics are as follows:

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \, , \tag{6}$$

where n is the number of classes, in this research, there is only one, and AP_k is the mean of k classes, so mAP is the overall mean of the average of all classes' score on the dataset.

$$Precision = \frac{TP}{TP + FP}, \tag{7}$$

$$Recall = \frac{TP}{TP + FN}, \tag{8}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{9}$$

where TP is True Positive, in this research, it means the model prediction bounding box is inside the ground truth, FP is False Positives, which means the model prediction's bounding box is within the image but not inside accurately enough, and FN is False Negatives, which means the model accurately predicts that there are no objects to be predicted.

RESULTS AND DISCUSSION

This section presents the experimental results and analysis of four YOLO variants, YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n, applied for detecting *Rhizopus stolonifer* mold on bread. The discussion encompasses model performance evaluation, comparative analysis across the four architectures, assessment of the custom mold detection dataset, and analysis of each model's inference speed. These components collectively aim to validate the effectiveness, accuracy, and practical applicability of the proposed detection framework in real-world food inspection and safety monitoring contexts.

Model Comparison

The establishment of performance benchmarks and computational efficiency metrics through systematic comparison among multiple YOLO architectures represents a fundamental step in validating the scientific contribution and practical feasibility of deep learning-based mold detection on bread. This comparative analysis serves several critical objectives in assessing model performance, including demonstrating algorithmic advancements across YOLO generations, quantifying improvements in computational efficiency for potential real-time deployment, and providing empirical evidence of the trade-off between detection accuracy and processing speed for each model. Such a comprehensive evaluation is essential in the evolving field of computer vision applications for food safety, where the adoption of efficient and reliable object detection systems can significantly enhance automated quality control and contamination prevention efforts.

To ensure the integrity and reproducibility of the comparative process, this research employed a standardized benchmarking protocol using a custom bread mold dataset curated via the Roboflow platform. The dataset features high-resolution images of bread under various environmental conditions, differing in lighting, background texture, and degrees of mold growth, to capture realistic scenarios encountered in bakery inspection environments. The benchmarking framework maintained consistent preprocessing routines, uniform training parameters, identical evaluation metrics, and equivalent hardware configurations across all YOLO models. This standardization ensured that observed performance differences could be attributed solely to architectural innovations rather than inconsistencies in experimental setup.

The comparative methodology focused on two primary evaluation aspects relevant to practical food safety applications: detection performance and computational efficiency. Detection performance was assessed using standard computer vision metrics, including mean Average Precision (mAP) at IoU thresholds of 0.50 (mAP50) and 0.50:0.95 (mAP50:95), as well as precision-recall analysis and F1-score evaluation to balance detection sensitivity and specificity. These metrics collectively provided a comprehensive understanding of each YOLO model's capacity to accurately localize and identify *Rhizopus stolonifer* mold across diverse visual conditions. Meanwhile, computational efficiency was evaluated in terms of model size, inference time, and hardware utilization, factors critical for real-time implementation within automated bakery monitoring systems or embedded inspection devices.

Table 2. Model Comparison with Mold on Bread Dataset

Model	Parameters	GFLOPs	Test mAP (%)		F1 (%)
			mAP50	mAP50:95	
YOLOv8n	3,011,043	8.2	47.2	20.3	38.0
YOLOv10n	2,707,430	8.4	47.4	19.1	34.0
YOLO11n	2,590,019	6.4	50.4	20.4	39.0
YOLOv12n	2,568,227	6.5	50.3	22.4	39.0

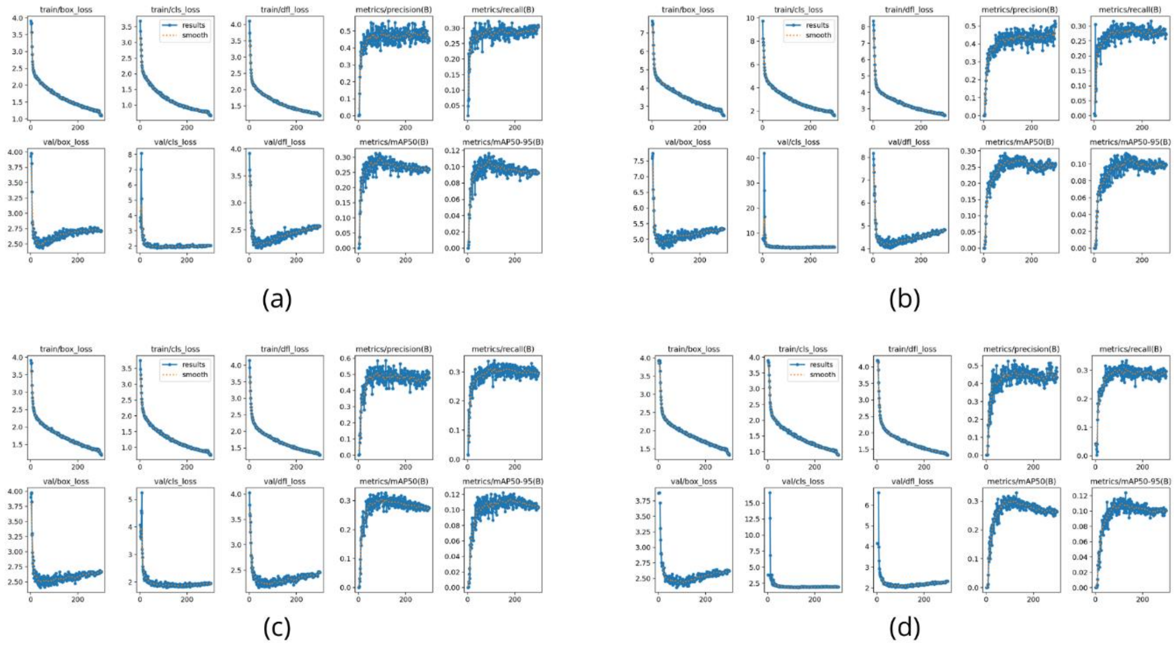

Figure 3. The results that have been generated by YOLOv8n (a), YOLOv10n (b), YOLO11n (c), and YOLOv12n (d), at the end of training of each model.

Table 2 and Figure 3 present the comparative analysis results for the four YOLO variants: YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n, highlighting their respective accuracy and computational performance characteristics. The results indicate that YOLO11n achieved the highest mAP50 score of 0.504, while YOLOv12n achieved the highest mAP50:95 of 0.224, signifying superior detection consistency across varying intersection-over-union thresholds. YOLOv8n and YOLOv10n demonstrated competitive yet slightly lower performance metrics, indicating that architectural enhancements in later YOLO versions contributed to improved model generalization and detection precision.

A comparative analysis was also performed on the Precision and Recall of each YOLO variant used in this research. Precision refers to the model's accuracy in predicting, which is the percentage of all true positives among all positive predictions. Recall measures how well the model identifies all the correct instances in an image. The Precision and Recall can be expressed mathematically as follows:

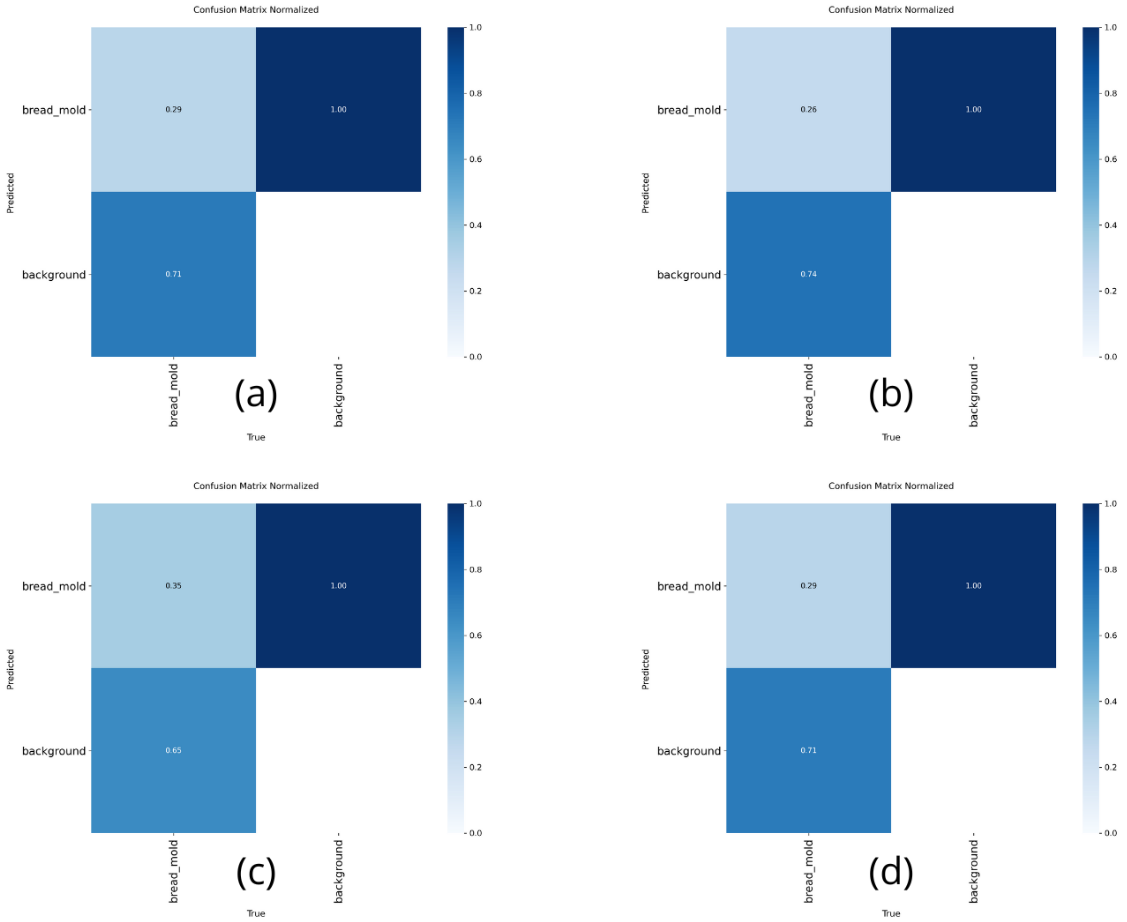

Figure 4. Normalized confusion matrix for models of YOLOv8n (a), YOLOv10n (b), YOLO11n (c), and YOLOv12n (d).

Figure 4 highlights the precision and recall of each YOLO variant, where, although YOLOv12n in Figure 4(d) achieves high accuracy in terms of mAP50:95, the percentage of True positives for detecting bread mold is the same as in YOLOv8n, as shown in Figure 4(a). Furthermore, YOLO11n achieves the highest True Positives percentage of 35%, as shown in Figure 4(c).

Beyond numerical comparison, the analysis underscores the practical implications of model selection for industrial deployment. Although minor differences exist in detection accuracy, YOLO11n and YOLOv12n provide an advantageous balance between detection precision and computational demand, making them suitable for real-time inspection scenarios. Their compact model structures and high inference speeds allow deployment on low-power or edge devices typically found in bakery production lines, supporting continuous contamination monitoring without the need for high-end hardware infrastructure. These characteristics enhance operational scalability, reduce inspection latency, and promote the adoption of intelligent food safety technologies across various production settings.

In summary, the comparative evaluation demonstrates that recent YOLO architectures deliver notable advancements in both accuracy and computational efficiency for mold detection tasks. The results confirm the robustness and adaptability of the YOLO framework in ensuring food quality and consumer safety through automated visual inspection, laying the groundwork for future integration into innovative food monitoring systems.

Qualitative Results Comparison

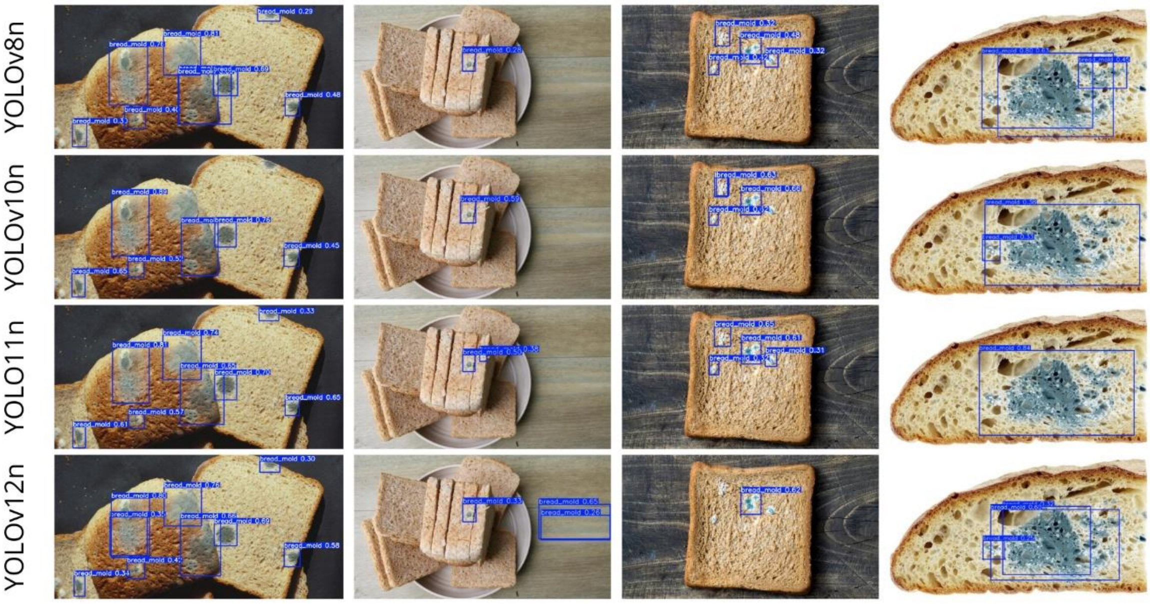

Figure 5. Qualitative results of each YOLO variant trained on the Mold on Bread dataset

This section compares the qualitative results between YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n. The findings show that YOLO11n provides better accuracy detection compared with the other YOLO variants in several cases, as shown in Figure 5. In the first column of the figure, YOLOv8n, YOLO11n, and YOLOv12n successfully detect all the molds on bread; however, YOLOv8n and YOLOv12n create double bounding boxes, consequently detecting more than they should. In the second column of the figure, only YOLO11n successfully detects two molds on bread compared with the other YOLO variants, while also interestingly, even though YOLOv12n achieves the highest mAP50:95, YOLOv12n manages to misdetect the background as the mold. In contrast, YOLO11n is more precise and generates a correct prediction (generating only one bounding box), as shown in Figure 5.

CONCLUSION

The comparative analysis of YOLOv8n, YOLOv10n, YOLO11n, and YOLOv12n for detecting *Rhizopus stolonifer* mold on bread demonstrated that YOLO11n achieved the best overall performance, with the highest mAP50 of 0.504 and consistent qualitative accuracy across various test cases. While YOLOv12n attained the highest mAP50:95 of 0.224, qualitative observations revealed that it occasionally produced redundant bounding boxes or false detections, whereas YOLO11n consistently generated precise and stable predictions. These results confirm YOLO11n's superior balance between detection accuracy and computational efficiency, making it most suitable for real-time bread mold detection in food inspection systems. This study demonstrates that modern nano YOLO architectures can effectively automate early-stage detection of fungal contamination, offering a practical alternative to labor-intensive visual inspection and culture-based methods in bakery quality control. Future work should focus on expanding the dataset to include more diverse bread types and contamination conditions, integrating multispectral imaging for early-stage mold detection, and testing real-time deployment on embedded edge devices to enhance practical implementation in automated food safety monitoring environments.

REFERENCES

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. https://arxiv.org/abs/2004.10934v1

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR 2021 - 9th International Conference on Learning Representations*. https://arxiv.org/abs/2010.11929v2

Explore Ultralytics YOLOv8 - Ultralytics YOLO Docs. (n.d.). Retrieved October 8, 2025, from https://docs.ultralytics.com/models/yolov8/

gopletzzz. (2025). Bread Mold Detection Dataset. In *Roboflow Universe*. Roboflow. https://universe.roboflow.com/gopletzzz/bread-mold-detection-55dam

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hosang, J., Benenson, R., & Schiele, B. (2017). Learning non-maximum suppression. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*, 6469–6477. https://doi.org/10.1109/CVPR.2017.685

Jegham, N., Koh, C.Y., Abdelatti, M., & Hendawi, A. (2024). YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions. https://arxiv.org/abs/2411.00201v2

Jubayer, F., Soeb, J. A., Mojumder, A. N., Paul, M. K., Barua, P., Kayshar, S., Akter, S. S., Rahman, M., & Islam, A. (2021). Detection of mold on the food surface using YOLOv5. *Current Research in Food Science*, *4*, 724. https://doi.org/10.1016/J.CRFS.2021.10.003

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., & Yang, J. (2020). Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Advances in Neural Information Processing Systems*, *2020-December*. https://arxiv.org/abs/2006.04388v1

Liu, A., Xu, R., Zhang, S., Wang, Y., Hu, B., Ao, X., Li, Q., Li, J., Hu, K., Yang, Y., & Liu, S. (2022). Antifungal Mechanisms and Application of Lactic Acid Bacteria in Bakery Products: A Review. *Frontiers in Microbiology*, *13*, 924398. https://doi.org/10.3389/FMICB.2022.924398/XML

Liu, Q., Chen, Q., Liu, H., Du, Y., Jiao, W., Sun, F., & Fu, M. (2024). Rhizopus stolonifer and related control strategies in postharvest fruit: A review. *Heliyon*, *10*(8). https://doi.org/10.1016/J.HELIYON.2024.E29522

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path Aggregation Network for Instance Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8759–8768. https://doi.org/10.1109/CVPR.2018.00913

Madasamy Raja, G., Pathmanaban, P., Selvaraju, P., & Vanaja, S. (2025). Bread contamination detection using deep learning and thermal imaging. *Journal of Food Engineering*, *400*, 112639. https://doi.org/10.1016/J.JFOODENG.2025.112639

Rahman, M., Islam, R., Hasan, S., Zzaman, W., Rana, M. R., Ahmed, S., Roy, M., Sayem, A., Matin, A., Raposo, A., Zandonadi, R. P., Botelho, R. B. A., & Sunny, A. R. (2022). A Comprehensive Review on Bio-Preservation of Bread: An Approach to Adopt Wholesome Strategies. *Foods 2022, Vol. 11, Page 319*, *11*(3), 319. https://doi.org/10.3390/FOODS11030319

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 779–788. https://doi.org/10.1109/CVPR.2016.91

Ribes, S., Fuentes, A., Talens, P., & Barat, J. M. (2018). Prevention of fungal spoilage in food products using natural compounds: A review. *Critical Reviews in Food Science and Nutrition*, *58*(12), 2002–2016. https://doi.org/10.1080/10408398.2017.1295017

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Tian, Y., Ye, Q., & Doermann, D. (2025). *YOLOv12: Attention-Centric Real-Time Object Detectors*. https://doi.org/10.0

Treepong, P., & Theera-Ampornpunt, N. (2023). Early bread mold detection through microscopic images using convolutional neural network. *Current Research in Food Science*, *7*, 100574. https://doi.org/10.1016/J.CRFS.2023.100574

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-Time End-to-End Object Detection*. *NeurIPS*, 1–21. http://arxiv.org/abs/2405.14458

Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2019). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2020-June*, 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203

Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., & Zuo, W. (2020). Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Transactions on Cybernetics*, *52*(8), 8574–8586. https://doi.org/10.1109/TCYB.2021.3095305