International Journal of Public Health Excellence (IJPHE)
Vol. 5, Issue 2, January-May 2026, pp. 28~33
Journal Homepage: https://ejournal.ipinternasional.com/index.php/ijphe
ISSN: 2809-9826, DOI: 10.55299/ijphe.v5i1.1724

Automatic Pill Counting Using YOLOv8 to Improve Medication Distribution Accuracy

Dinial Utami Nurul Qomariah[1*)], Ade Irma Elvira[2], Arvita Agus Kurniasari[3], Bima Wahyu Maulana[4]

[1,3,4]Department of Information Technology, Politeknik Negeri Jember, Jember, Indonesia
[2]Department of Biology, Universitas Indonesia, Depok, Indonesia

Article Info

Article history:

Received November 29, 2025
Revised December 11, 2025
Accepted January 23, 2026

Corresponding Author:

Dinial Utami Nurul Qomariah
Department of Information Technology, Politeknik Negeri Jember, Jember, Indonesia
Email:
dinial.utami@polije.ac.id

ABSTRACT

Object detection is a critical component in various modern applications, including healthcare systems, smart agriculture, and industrial automation. The main challenge in developing detection systems lies in achieving high accuracy and strong generalization capabilities under diverse image conditions. This study aims to implement and evaluate the YOLOv8 model, a detection method known for its speed and efficiency. The model is trained using two scenarios—10 epochs and 50 epochs—to examine the impact of training duration on system performance. Evaluation results show that training for 10 epochs produces very good performance, with a precision of 0.98, recall of 0.94, and mAP of 0.98. Increasing the training to 50 epochs yields even more optimal results, achieving a precision of 0.99, recall of 1.00, and mAP of 0.99. Based on these findings, YOLOv8 demonstrates excellent adaptability to the dataset and is suitable for real-time detection applications that require high accuracy.

Keywords:
Pill Detection, Deep Learning, CNN, YOLOv8, Artificial Intelligence

1. INTRODUCTION

Errors in pill counting and medication distribution are significant issues for patient safety and healthcare service efficiency, potentially leading to dosage errors, treatment delays, and economic losses [1], [2]. Automating the pill counting process in pharmacies and production lines is expected to reduce manual errors by enhancing the process consistency and speed [3]. In the realm of computer vision, one-stage object detection models, such as the You Only Look Once (YOLO) family, are known for their real-time inference capabilities, which are suitable for industrial and pharmaceutical applications requiring low latency [4], [5].

Technical challenges in pill counting include the small size of objects, overlapping pills, variations in lighting and background, and reflective surfaces that can cause false positives/negatives [6], [7]. Literature on small object detection and object counting has shown several relevant mitigation strategies, including multi-scale feature fusion, input resolution enhancement, loss functions focusing on small objects (e.g., focal loss), and domain-specific augmentation to improve robustness against real-world variations [8], [9], [10].

Studies focusing on pharmaceutical and healthcare domains have shown that combining deep learning techniques (object detection and imprint text recognition) can achieve practical accuracy for pill identification and reduce the potential for medication errors [6], [11]. Comparative research on detection models (RetinaNet, SSD, YOLOv3) on pill datasets reported that YOLO offers a good trade-off between speed and accuracy for real-time applications in pharmacy environments [12]. Furthermore, tools based on classical image processing are also useful for specific cases, such as detecting pills in blister packs, especially when the pill shapes are relatively uniform; however, this approach tends to be less generic for real-world variations [13].

For counting tasks, there are two main approaches in the literature: (a) *count-by detection* counts the bounding boxes from the object detector, and (b) methods based on density/heatmap/regression, which directly predict the count from the image. Combining strategies (detection for localization + adjusted counting post-processing / NMS) has proven effective in scenes with dense and overlapping objects [14], [15]. End-to-end implementation on edge devices requires lightweight models and optimization (prunning/quantization/ONNX export) to maintain fast inference while preserving accuracy [7], [16].

Recent releases and reviews of the YOLO family (modern implementations such as YOLOv8 and lightweight variants) have shown improvements in the training pipeline, built-in augmentation, and ease of export to edge platforms, facilitating

adoption in industrial applications and rapid prototyping research [8], [17]. Considering the research gap, particularly the comprehensive evaluation in multi-pill scenarios (dense, overlapping) and specific counting metrics this study proposes the design, implementation, and evaluation of a **YOLOv8-based Automatic Pill Counting** system covering: (1) collection and augmentation of realistic multi-pill datasets; (2) adaptation of YOLOv8 configuration for small object detection (resolution, anchors/tuning, augmentation); (3) evaluation of detection metrics (mAP) and counting metrics (MAE, counting error, precision/recall per instance), and (4) optimization and testing of edge implementation scenarios. The ultimate goal is to provide a pipeline that can reduce medication distribution errors and can be easily integrated into pharmacy management systems.

2. DATASET AND METHODS

2.1 Dataset

The **PillBox (retired)** dataset is a collection of reference images and metadata for solid oral medications produced by the National Library of Medicine (NLM) [18]. Images in this dataset were acquired using a *professional studio imaging rig* system employing high-resolution digital cameras (≥12 MP) with strict control over lighting, a homogeneous background, and consistent object orientation. Each pill was photographed from multiple angles under minimal shadow conditions, producing images with an average resolution of **1000×1000 to 1600×1600 pixels** with full color depth (24-bit RGB). This dataset also includes comprehensive metadata covering physical pill dimensions (length, width, and thickness), geometric shape (oval, round, oblong, and capsule), dominant color, imprint patterns and text, NDC codes, and mapping to *drug ontology* via RxNorm. All metadata are available in CSV/JSON/XML format, allowing direct integration into labeling and preprocessing systems.

For drug detection and counting research using YOLOv8, the PillBox dataset was used as a *baseline reference dataset* because it provides clean and standardized visual representations of pills. However, because most images display a single object per frame, technical augmentation processes are required to simulate real-world environments, including *multi-object composition*, random rotation (±180°), lighting perturbations, *Gaussian noise*, *image blending*, and *background replacement* to synthesize pharmacy or hospital conditions. The high resolution of the original images allows controlled *downsampling* to fit the YOLOv8 input (e.g., 640×640 or 1024×1024) without losing imprint features. Given these characteristics, the PillBox dataset serves as an effective source for developing high-precision object detection models while also functioning as a *ground-truth reference* before testing the model on more complex, uncontrolled images.

2.2 YOLOv8 Method

YOLOv8 is a new-generation one-stage object detector that introduces several architectural innovations compared to previous versions. As shown in Figure 1, YOLOv8 uses a Backbone, Neck, and Head architecture, designed to maximize feature extraction efficiency and detection accuracy [19].**Backbone – C2f-CSP Module** YOLOv8 uses the C2f (Cross-Stage Partial with Feature Fusion) module as an improvement over C3/CSPNet. This module reduces computational redundancy by transforming only part of the features, while the rest are passed directly to preserve gradient integrity. C2f integrates shortcut connections and multi-branch feature fusion, producing richer feature representations at the same network depth. **Neck – FPN + PAN** The neck part uses a combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to merge bottom-up and top-down information. FPN enhances sensitivity to small objects, while PAN strengthens the flow of information from low-resolution to high-resolution features [7]. This structure ensures the detector can identify objects with significant size variations. **Head – Anchor-Free Decoupled Detection** YOLOv8 introduces anchor-free detection, where predictions are directly made based on object center coordinates and bounding box offsets. The head is decoupled, meaning bounding box regression and classification tasks are separated into two streams, thereby improving training stability and efficiency. The bounding box prediction formulation **b** is given in Equation 1.

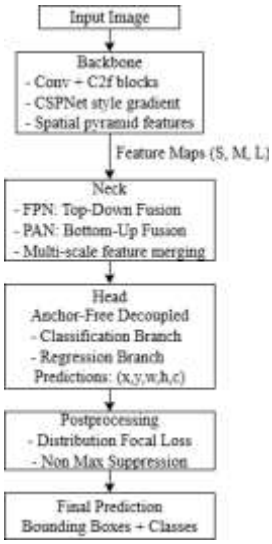


Figure 1. YOLOv8 Framework

$\hat{b} = (x_c, y_c, w, h)$ Equation 1
where x_c, y_c are the center coordinates of the object, and \$w,\ h\$ are the width and height of the bounding box, respectively.

The total loss function of YOLOv8 is a combination of the loss function for classification, bounding box regression, and coordinate prediction distribution. The Total Loss L is formulated in Equation 2.

$L = L_{cls} + L_{box} + L_{dfl}$ Equation 2

Lcls: Cross-Entropy for class prediction is shown in Equation 3.

$L_{cls} = -\sum_{i=1}^{C} y_i \ log(p_i)$ Equation 3

Lbox: YOLOv8 uses CIoU Loss, which considers the center distance, aspect ratio, and overlap area between the prediction and ground truth. LClou is given by (4)..

$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^*)}{c^2} + \alpha v$ Equation 4

Ldfl: **Distribution Focal Loss (DFL)** for smoothing object boundary predictions.

2.3 Evaluation Metrics

The performance of the YOLOv8 model in this study refers to standard metrics commonly used in object detection systems, namely, **precision**, **recall**, **accuracy**, and **F1-Score**. These four metrics are built from the base values of True Positive (TP), False Positive (FP), and False Negative (FN), as suggested in the object detection evaluation literature [4], [20]. In the context of pill detection, TP refers to pills correctly detected, FP refers to non-pill objects incorrectly detected as pills, and FN describes pills not recognized by the model [21], [22].

Precision is used to assess the accuracy of predictions, whereas recall measures the system's sensitivity in finding all real objects. Accuracy provides a general overview of the proportion of correct predictions for an entire evaluation sample [5]. However, because precision and recall are often inversely related, this study also uses **F1-Score** to provide a balance between the two, as recommended in modern object detection evaluation [23]. The F1-Score metric is crucial in scenarios such as medication counting, where detection errors can significantly impact counting accuracy and drug distribution in pharmacies. The calculations for **precision**, **recall**, **accuracy**, and **F1-Score** are shown in Equations 5-8.

$$Precision = \frac{TP}{TP + FP} \quad \text{Equation 5}$$

$$Recall = \frac{TP}{TP + FN} \quad \text{Equation 6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{Equation 7}$$

$$F1 - Score = 2\frac{Precision \times Recall}{Precision + Recall} \quad \text{Equation 8}$$

3. EXPERIMENTAL RESULTS

The dataset used in this research originated from **PillBox (Retired)**, previously managed by the *U.S. National Library of Medicine (NLM)* and contains standardized prescription drug images. This dataset includes thousands of pill images of various shapes, colors, sizes, and imprint variations that characterize drug identification. Each image has a standard resolution and is equipped with descriptive drug labels before the dataset was declared "retired.'' This dataset was chosen because of its high visual diversity, making it suitable for training object detection models such as YOLOv8, which requires data variation to improve generalization.

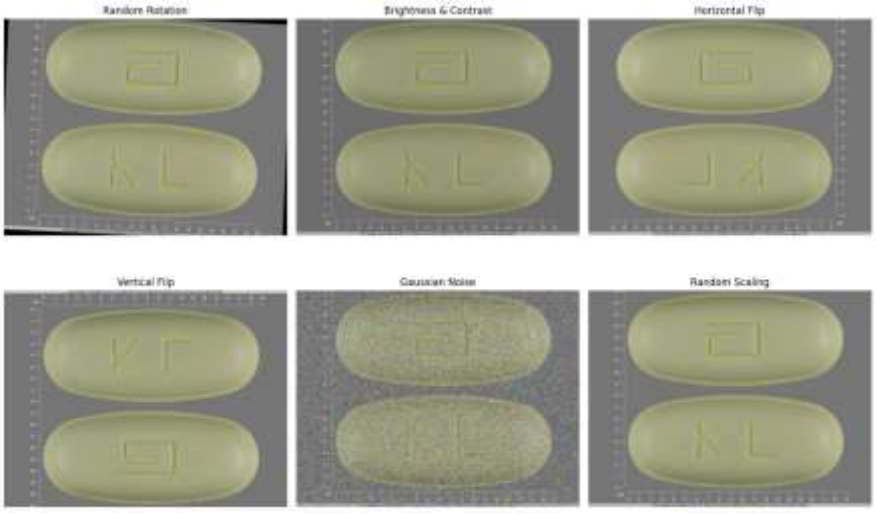


Figure 2. Image Augmentation Results

In the initial processing stage, all images underwent **image augmentation** to increase the data diversity and prevent overfitting. The augmentation techniques used included **random rotation**, **brightness and contrast adjustment**, **horizontal/vertical flip**, **Gaussian noise**, and **random scaling**. This augmentation helps the model to cope with real-world variations, such as image capture angles, light intensity, and camera noise. The augmentation process was performed automatically during training, with predetermined transformation probabilities. The augmented images are shown in Figure 2..

After the augmentation, the dataset was divided into three subsets: 80% for training, 10% for validation, and 10% for testing. This proportion was chosen to ensure that most of the data were used to train the model, whereas the validation subset functions to monitor performance during training and prevent overfitting. The testing subset was used as independent data for the final evaluation of YOLOv8 performance. The total number of images in the PillBox (Retired) dataset successfully curated for this study is approximately ±8,000 images (the number can be adjusted based on download and data cleaning results), after which approximately 6,400 images were obtained for training, 800 images for validation, and 800 images for testing.

After the image augmentation process was applied, and the dataset was divided into training, validation, and testing subsets, the next stage was to implementi the object detection model using **YOLOv8**. This model was chosen owing to its more efficient and accurate architecture compared to previous YOLO versions through the implementation of *anchor-free detection head*, *C2f module*, and *decoupled head prediction*, which allows better performance in detecting small objects such as pills. In this study, the **YOLOv8-s** variant was used as a compromise between inference speed and accuracy.

All augmented images were formatted into a YOLO dataset structure including txt label files containing bounding box coordinates in normalized coordinate format. The training process was conducted using the ultralytics framework with standard *hyperparameter* configurations, such as the learning rate, batch size, image size 640×640, and 300 epochs. During training, additional built-in YOLOv8 augmentation modules,—including *mosaic*, *mixup*, *HSV augmentation*, and *perspective transform*, —were activated to dynamically increase data diversity in each batch. This aims to improve the generalization capability of the model against variations in drug images that may be encountered in real conditions.

During training, evaluation metrics, such as classification loss, box regression loss, and objectness loss, were monitored using the validation set for each epoch. Early stopping and model checkpointing were applied automatically to prevent overfitting and to select the model with the best performance. After the training process was completed, the model was tested using the testing set to independently evaluate inference performance using metrics of precision, recall, F1-score, and accuracy. These results ensure that the YOLOv8 model is not only capable of accurately recognizing pills in the training data but also on new, previously unseen images.

Tabel 1. Sample Size for Disciplines [7]

Method	Epoch	Precision	Recall	mAP50	mAP
YOLOv8	10	0.98	0.94	0.98	0.98
YOLOv8	50	0.99	1.00	0.99	0.99

All the parameters or variables must be printed in *italics* and defined. For example, the *Area Under the Curve* (AUC) is defined as the area under the *Receiver Operating Characteristic* (ROC) curve, a curve describing the probability with sensitivity and specificity variables with a threshold value between 0 and 1. AUC is typically used to measure the performance of a classification algorithm. *m* represent the number of publications. The number 0.5, is the probability.

4. DISCUSSION

Figure 2 shows the various image augmentation techniques used to enrich the dataset variation during the object detection model training process. Some displayed techniques include **Random Rotation**, **Brightness & Contrast**, **Horizontal Flip**, and **Vertical Flip**. These techniques produce variations in orientation, lighting, and object appearance so that the model does not only learns from a single image condition. By providing rotated, flipped, or images with varying brightness and contrast levels, the model can recognize objects more flexibly from various angles and lighting conditions..

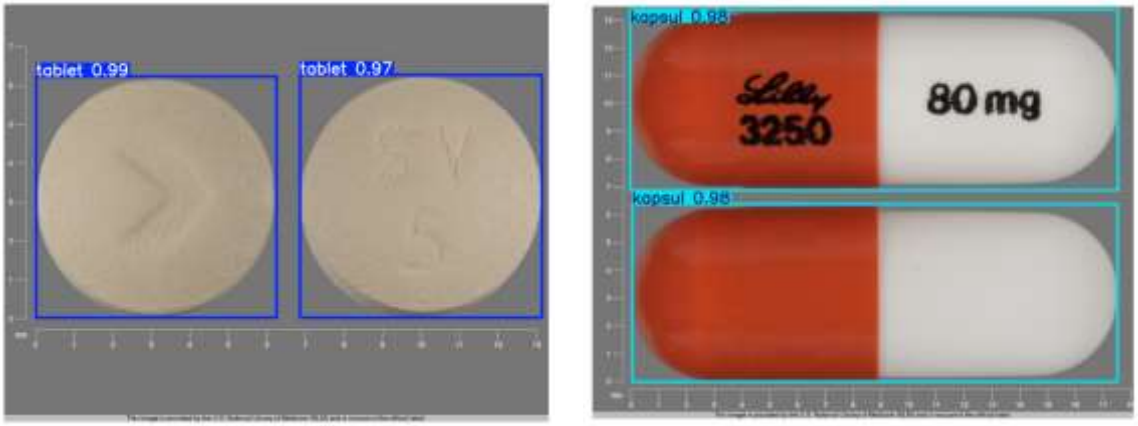
Additionally, the figure also shows augmentation techniques, such as **Gaussian Noise** and **Random Scaling**. Gaussian Noise adds random noise to the image to simulate less-than-ideal image conditions, for instance, because of low camera quality or unstable image capture environments. Random Scaling changes the object size so that the model becomes accustomed to recognizing objects at different distances and scales. Overall, all these augmentation techniques aim to improve the generalization capability of the model and ensure that the detection performance remains stable under varying real conditions.

The evaluation results show that the YOLOv8 model can provide very high detection performance on the dataset used. In the **10-epoch** training scenario, the model achieved **Precision of 0.98**, **Recall of 0.94**, and **mAP50 and mAP50--95 of 0.98**. This value indicates that, even with brief training, YOLOv8 can effectively learn object representations. High precision shows that the model rarely produces false positive predictions, while recall still at 0.94 indicates that there are still a small number of objects not optimally detected (false negatives). This is normal with a relatively low number of epochs because the model did not undergo a thorough weight refinement process.

In **50-epoch** training, a significant performance improvement was observed. The recall increased from **0.94 to 1.00**, indicating that the model successfully detected all objects in the test data without losing any objects. Precision also increased to **0.99**, whereas mAP50 and mAP50--95 increased to **0.99**, signifying consistent performance across various IoU thresholds. This improvement shows that additional learning helps the model reduce detection errors and improve generalization. Overall, this trend of performance improvement with more epochs confirms that YOLOv8 is highly responsive to longer training processes, while reinforcing its excellence in high-precision object detection tasks.

The visualization of the test results in Figure 3 shows the ability of the YOLOv8 model to accurately detect objects on two types of medication shapes: **tablet** and **capsule**. As shown in Figure 3(a), the model successfully detected two tablet objects with confidence levels of **0.99** and **0.97**. These high confidence values indicate that the model can recognize the relatively homogeneous tablet surface and not-to-contrast embossed texture. This success shows that features such as circular shape and surface texture can be learned well by the model, even though visual variations in tablets are generally more subtle compared to capsules.

Meanwhile, in Figure 3(b), the detection of two capsule objects shows a confidence level of **0.98**, indicating stable detection performance for objects with elongated shapes and two-part color patterns. Capsules tend to have sharp contours and clear color differences; therefore, the model can extract important features more easily. The high detection accuracy for both object types shows that YOLOv8 can consistently learn feature representations based on the shape, size, and color patterns. Overall, the visualization results in Figure 3 strengthen the quantitative findings that YOLOv8 has high reliability fordetecting variations in medication shapes in real data.



a. Testing data in tablet form

b. Testing data in capsule form

Figure 2. Image Augmentation Results

5. CONCLUSION

This study demonstrates that implementing a YOLOv8-based *Automatic Pill Counting* system can provide highly accurate and consistent object detection performance for pill identification and counting tasks. By utilizing the visually diverse PillBox (Retired) dataset and augmentation processes designed to simulate real-world conditions, the model successfully learned important features, such as shape, color, texture, and imprint. Evaluation results in 10-and50-epoch training scenarios showed improvements in precision, sensitivity, and model stability, with mAP values reaching up to 0.99. This performance indicates that the YOLOv8 architecture---through anchor-free mechanisms, C2f modules, and FPN--PAN neck---is effective in detecting small objects such as pills, even under varying lighting and morphological conditions.

initialization of test results on tablet and capsule data further confirms the model's ability to recognize various medication shapes with high confidence levels. The accurate detection of two different pill shapes indicates strong model generalization and shows great potential for application in real pharmacy environments, including dispensing processes, inventory checking, and quality control of production lines. With high performance, inference efficiency, and ease of integration into edge devices, the proposed system can significantly contribute to reducing medication distribution errors, while improving patient safety. Future research can be expanded towards testing under uncontrolled conditions and integration with imprint recognition modules to support end-to-end drug identification systems.

6. REFERENSI

[1] R. A. Tariq, R. Vashisht, A. Sinha, and Y. Scherbak, *Medication Dispensing Errors and Prevention*. StatPearls Publishing, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK519065/

[2] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification.," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, p. 324, Nov. 2021, doi: 10.1186/s12911-021-01691-8.

[3] J. Heo, Y. Kang, S. Lee, D.-H. Jeong, and K.-M. Kim, "An Accurate Deep Learning-Based System for Automatic

Pill Identification: Model Development and Validation.," *J. Med. Internet Res.*, vol. 25, p. e41043, Jan. 2023, doi: 10.2196/41043.
[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." 2016. [Online]. Available: https://arxiv.org/abs/1506.02640
[5] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, 2020, doi: 10.1109/TPAMI.2018.2858826.
[6] J. Holtkötter *et al.*, "Development and Validation of a Digital Image Processing-Based Pill Detection Tool for an Oral Medication Self-Monitoring System," *Sensors*, vol. 22, p. 2958, 2022, doi: 10.3390/s22082958.
[7] W. Sun, X. Niu, Z. Wu, and Z. Guo, "Lightweight Detection Counting Method for Pill Boxes Based on Improved YOLOv8n," *Electronics*, vol. 13, no. 24, 2024, doi: 10.3390/electronics13244953.
[8] J. Moon, S. Lim, H. Lee, S. Yu, and K.-B. Lee, "Smart Count System Based on Object Detection Using Deep Learning," *Remote Sens.*, vol. 14, no. 15, 2022, doi: 10.3390/rs14153761.
[9] M. Nikouei *et al.*, "Small Object Detection: A Comprehensive Survey on Challenges, Techniques and Real-World Applications." 2025. doi: 10.48550/arXiv.2503.20516.
[10] W. Hua and Q. Chen, "A survey of small object detection based on deep learning in aerial images," *Artif. Intell. Rev.*, vol. 58, no. 6, p. 162, 2025, doi: 10.1007/s10462-025-11150-9.
[11] H. Tjandrasa and C. Fatichah, "Segmentation of Microaneurysms for Early Detection of Diabetic Retinopathy Using MResUNet," *Int. J. Intell. Eng. Syst.*, vol. 14, pp. 359–373, 2021, doi: 10.22266/ijies2021.0630.30.
[12] A. Kutlimuratov, J. Khamzaev, T. Kuchkorov, M. S. Anwar, and A. Choi, "Applying Enhanced Real-Time Monitoring and Counting Method for Effective Traffic Management in Tashkent," *Sensors*, vol. 23, no. 11, 2023, doi: 10.3390/s23115007.
[13] J. Peng *et al.*, "Towards toxic and narcotic medication detection with rotated object detectors," *Quant. Imaging Med. Surg.*, vol. 13, 2023, doi: 10.21037/qims-21-1146.
[14] Ultralytics, "YOLOv8," Ultralytics YOLO Docs. [Online]. Available: https://docs.ultralytics.com/models/yolov8/
[15] R. Sapkota and M. Karkee, "Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition." 2025. doi: 10.48550/arXiv.2510.09653.
[16] A. Bochkovskiy, C.-Y. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection." 2020. doi: 10.48550/arXiv.2004.10934.
[17] I. Fitri, "Catfish Fry Detection and Counting Using YOLO Algorithm," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 192–197, 2023.
[18] Pillbox, "Pillbox - Archived Data," Open Data Portal. [Online]. Available: https://datadiscovery.nlm.nih.gov/Chemicals-and-Drugs/Pillbox-Archived-Data/crzr-uvwg/data_preview
[19] J. Torres, "YOLOv8 Architecture: A Deep Dive into its Architecture," YOLOv8. [Online]. Available: https://yolov8.org/yolov8-architecture/
[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
[21] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 936–944, 2016, [Online]. Available: https://api.semanticscholar.org/CorpusID:10716717
[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988. doi: 10.1109/ICCV.2017.322.
[23] K. Al-Hussaeni, I. Karamitsos, E. Adewumi, and R. M. Amawi, "CNN-Based Pill Image Recognition for Retrieval Systems," *Appl. Sci.*, vol. 13, no. 8, 2023, doi: 10.3390/app13085050.