Journal of Robotics and Control (JRC)
Volume 5, Issue 2, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i2.20900

Unveiling the Advancements: YOLOv7 vs YOLOv8 in Pulmonary Carcinoma Detection

Moulieswaran Elavarasu [1], Kalpana Govindaraju [2*]
[1, 2] Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203
Email: [1] emoulieswaran@gmail.com, [2] kalpanag@srmist.edu.in
*Corresponding Author

***Abstract*—In this work, precision and recall measures are used to assess the performance of YOLOv7 and YOLOv8 models in identifying pulmonary carcinoma on a distinct collection of 700 photos. The necessity of early disease detection is increasing, thus choosing a reliable object detection model is essential. The goal of the research is to determine which model works best for this purpose, taking into account the unique difficulties that pulmonary cancer presents. The work makes a contribution to the field by showcasing the improvements made to YOLOv8 and underlining how well it detects both benign and malignant. YOLOv7 and YOLOv8 were used to independently train custom models using the pulmonary carcinoma dataset. The models' performance was measured using precision, recall, and mean average precision measures, which allowed for a comprehensive comparison examination. When it came to precision (58.2%), recall (61.2%), and mean average precision at both the 0.5:0.95 (33.3%) and 0.5 (53.3%) criteria, YOLOv8 outperformed YOLOv7. The 3.0% accuracy gain highlights YOLOv8's improved capabilities, especially in identifying small objects. YOLOv8's enhanced accuracy can be attributed to the optimisation of the detection process through its anchor-free design. According to this study, YOLOv8 is a more reliable model for pulmonary carcinoma identification than YOLOv7. The results indicate that YOLOv8 is the better option because of its higher recall, precision, and enhanced capacity to detect smaller objects—all of which are critical for early illness detection in medical imaging.**

Keywords*—YOLOv7; YOLOv8; Object Detection; Computer Vision; Pulmonary Carcinoma Detection; Medical Image Analysis.

I. INTRODUCTION

The shape and color of objects used as the basis for recognition by traditional image detection algorithms [1]. Due to a lack of appropriate resilience and error detection, these sophisticated algorithms have some application restrictions [2]. Due to their ability to get beyond the drawbacks of conventional image identification algorithms and efficiently extract object attributes from complicated images, deep-learning-based object recognition algorithms have gained popularity in current research and applications.

Accuracy, speed, and environmental variance are traditional object detection algorithms [3]. There are several approaches to object detection [4], and they can broadly be categorized into 2 main types: one-stage detectors and two-stage detectors. Most object detection techniques may be separated into one of two types: the initial method based on candidate areas, are Faster RCNN (region-based convolutional neural network) and Fast RCNN.

1. Two-Stage Algorithm:

- Faster R-CNN: Building on Faster R-CNN, Fast R-CNN integrates the region proposal network (RPN) into the detection pipeline, faster inference and allowing for end-to-end training [5] , [6].
- Fast R-CNN: An improvement upon Fast R-CNN, R-CNN uses the entire image for feature extraction and introduces a region of interest (RoI) pooling layer to efficiently classify object proposals [7] , [8].
- R-CNN (Regions with Convolutional Neural Networks): R-CNN was one of the pioneering approaches that uses a two-stage process. It involves generating region proposals using a selective search algorithm and then classifying these proposed regions using a convolutional neural network (CNN) [9].

2. One-Stage Algorithm[10]:

- YOLO (You Only Look Once): YOLO is a one-stage detector[11] that divides the input image into a grid and predicts bounding boxes and class probabilities directly from this grid [12]. Mean average precision (mAP), a scaled-down version called Fast YOLO beats existing real-time detectors while processing data at an incredible 155 frames per second [13].
- SSD (Single Shot Multibox Detector): Similar to YOLO, SSD is a one-stage detector that generates multiple bounding box predictions at different scales and aspect ratios in a single pass through the network [14].
- EfficientDet: This is an efficient and scalable object detection model that balances accuracy and computational efficiency. It uses a compound scaling method to optimize model parameters for different resolutions [15].
- RetinaNet: RetinaNet introduces the focal loss to address the class imbalance problem in one-stage detectors, resulting in improved performance, especially for detecting rare objects [16].

These methods produce the class probability and object location coordinate values directly, yielding the final detection outcome after only one inspection [17], [18], [19], [20]. The YOLO algorithm is applicable to a variety of computer vision (CV) tasks, including those involving hospitals, autonomous vehicles, drones, the military, wildlife,

and others [21]. To identify which YOLO algorithm performs better than the others. According to past studies, In terms of speed and accuracy, Compared with some newer versions, YOLOv8 outperforms YOLOv3 [22], YOLOv4, YOLOv5, YOLOv6 and YOLOv7 [23], [24], [25]; additionally, it is necessary to determine whether YOLOv8 surpasses YOLOv7. This study evaluated YOLOv7 and YOLOv8.

This paper ordered as Section 2 represents the YOLO background; Section 3 reviews of literature. Section 4 methodology. Section 5 discuss about results, and at last, Section 6 we conclude paper.

II. BACKGROUND OF YOLO

YOLO is a unique method for detecting objects. YOLO permits real-time speeds and end-to-end training while keeping a high average level of precision. The method creates S×S grid from the provided image. A grid cell is identified an object if its center within the cell. Expected confidence scores and B-bounding boxes for those boxes will be present for each grid cell. If there is no object present in that cell, the confidence ratings should be 0. If not, the intersection over union (IOU) between the projected box and the five predictions—x, y, w, h, and confidence—that make up each bounding box should be used as the confidence score. The center of the box with respect to the grid cell's boundaries is represented by the (x, y) coordinates. W and H stand for the expected width and height in entire image. The confidence prediction represents the IOU between the actual box and the predicted box. The network design based on the GoogLeNet image categorization model. In the network, 24 convolutional layers come before 2 completely coupled layers. Instead of using GoogleNet's inception modules, 3×3 convolutional layers and 1x1 reduction layers are integrated. Fig. 1 shows the YOLO evolution [13]. In Fig. 1, it shows the different versions of YOLO and when it was introduced.

A. YOLO Versions improvements

Each unique iteration of YOLO provides more beneficial upgrades and enhancements. Compared with some newer versions, Table I provides a summary of the conclusions drawn from YOLO versions.

B. YOLOv7 Model

Focus on Efficiency: YOLOv7 prioritizes reduced model complexity and increased inference speed while maintaining good accuracy [27]. Scholarly literature references YOLOv7's basic position within the YOLO series [28]. It achieves this through:

- Slimmed-down backbone network: Utilizes EfficientNet, which combines depthwise convolutions and squeeze-and-excitation modules for efficient feature extraction [29].
- Simplified head structure: Employs a single-stage detection framework with a single head for both classification and regression, reducing computational overhead [18].
- Improved training methods: Introduces Focal Loss and CIoU Loss for better bounding box prediction and handling of hard examples [16], [30].
- Enhanced Focus on Small Objects: Implements techniques to improve small object detection:
- Multi-scale training: Trains the model on images resized to different scales, improving feature learning for objects of various sizes [31].
- Anchor-free prediction: Eliminates pre-defined anchors, allowing the model to dynamically predict bounding boxes of any size, better suited for small objects [32].

With its amazing features, YOLOv7 is a real-time object detector. It has better accuracy of all real-time object detectors with 30 FPS, with a 56.8% mAP, and outperforms all object detectors in the range of 5 to 160 FPS. YOLOv7 has expanded efficient layer aggregation networks (E-ELAN). The optimal structure of the model and its original attributes both preserved using the suggested compound scaling strategy [13].

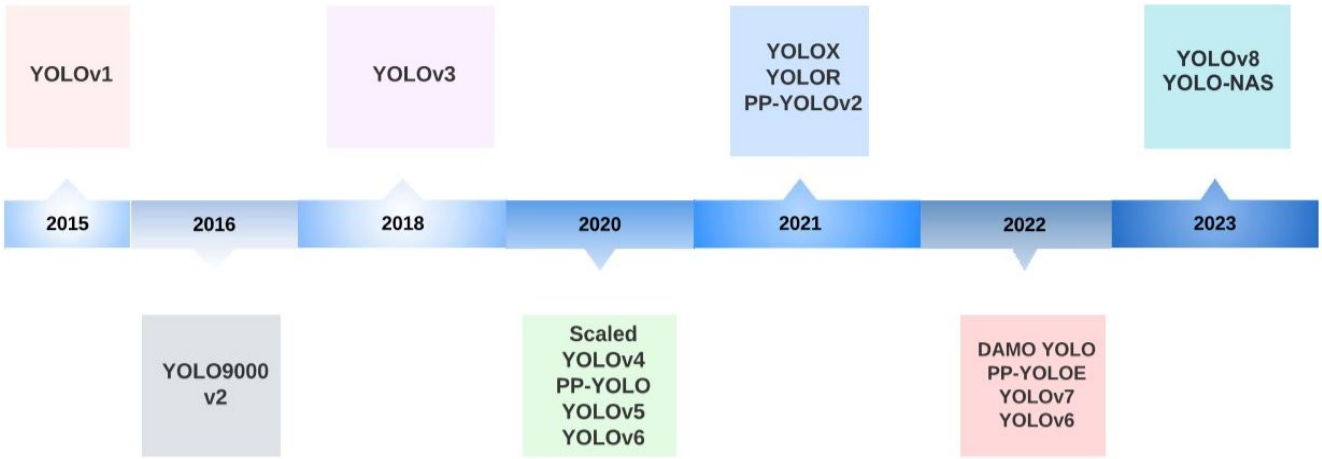

Fig. 1. YOLO evolution [26]

TABLE I. SUMMARY OF YOLO

S.No	YOLO Versions	Enhancements	References
1.	YOLOv1	The issues with boundary box creation and class identification are combined and resolved by a single-shot detector	[13]
2.	YOLOv2	higher-quality detection, Batch normalization, and the use of anchor boxes have all seen iterative improvements.	[18]
3.	YOLOv3	Bounding box prediction with the addition of objectness score, connections to the backbone network layers	[18], [34]
4.	YOLOv4	Enhanced feature aggregations, and mish activation	[35], [36]
5.	YOLOv5	Cross-Stage Partial Network (CSPNet) for architecture and the PANet for its neck, auto-anchoring, and residual structure	[23], [37], [38], [39]
6.	YOLOv6	Backbone and neck designs are Rep-PAN Neck and EfficientRep Backbone.	[20]
7.	YOLOv7	E-ELAN, a trainable bag of freebies, utilizes 35%	[40]
8.	YOLOv8	Use task alignment learning and an anchor-free approach to align classification and regression tasks	[41], [42], [43], [44]

C. *YOLOv8 Model*

- The latest version is the YOLOv8 algorithm [33], which was open-sourced by Ultralytics in January 2023.
- Further Speed and Accuracy Gains: Builds upon YOLOv7's efficiency while pushing the boundaries of accuracy and speed:
- Anchor-free architecture: Similar to YOLOv7, eliminates anchors for dynamic box prediction, but with a novel Path Aggregation Network (PAN) for improved feature propagation and context understanding [45].
- Multi-scale prediction: Enhances YOLOv7's approach by employing multiple prediction heads at different scales, leading to better multi-scale object detection [15].
- Improved backbone network: Introduces a custom YOLOv8 backbone with densely connected convolutions for richer feature extraction and higher accuracy [46].

Lightweight YOLOv8m chosen for this paper. The YOLOv8 method gave rise to the lightweight parameter structure known as YOLOv8m. It consists of a backbone network, neck network, and prediction output head. Convolutional operations used for backbone network to extract properties on various scales from RGB (Red, Green, and Blue) color images. The neck network's job is to combine the features that the backbone network has extracted. To combine low-level characteristics into higher-level representations, feature pyramid structures, or FPNs, are usually used. Three sets of detection detectors of varied sizes are used to pick and detect the image contents [41], [42], [47]. The head layer is in charge of predicting the target category.

Unlike traditional object detection methods that use multiple stages, YOLO performs detection in a single pass. It divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell.

1. Grid-Based Approach:

YOLO divides the input image into a grid of cells. Each cell is responsible for predicting bounding boxes and class probabilities. The grid allows the algorithm to efficiently capture objects at different locations and scales within the image. The bounding boxes and class predictions are made directly at the grid cell level, making YOLO fast and effective. Fig. 2 represents the YOLO Architecture and how much conventional layer has been used.

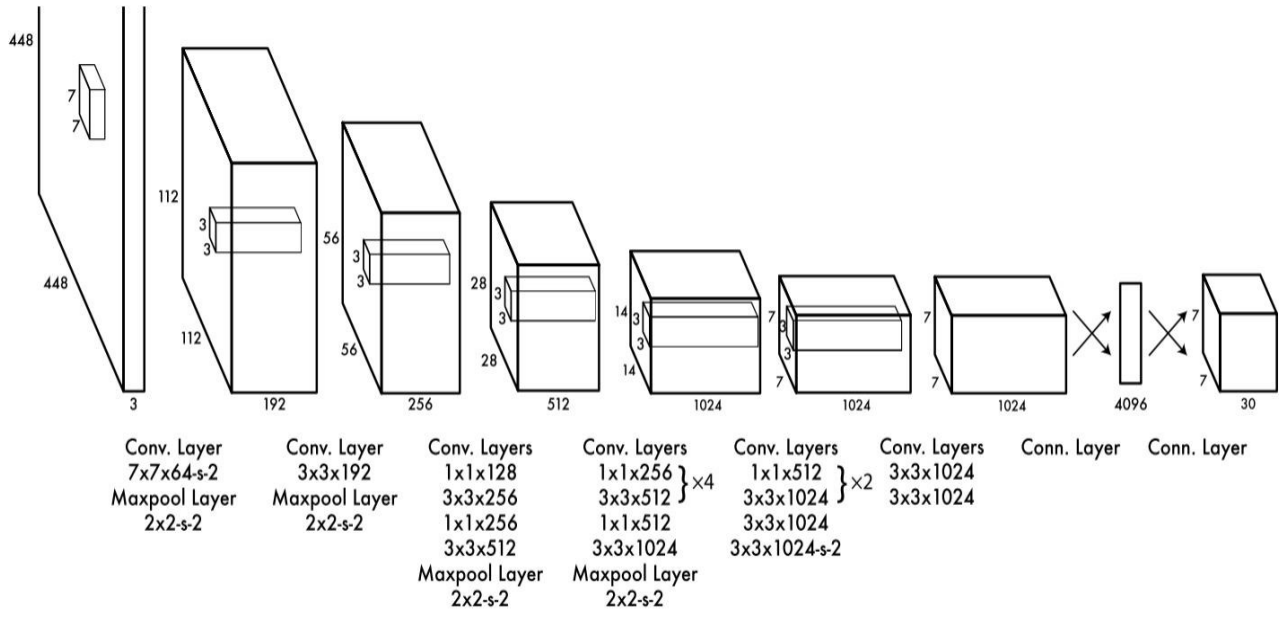

Fig. 2. YOLO architecture [13]

1) Bounding Boxes:

Bounding boxes are rectangular regions that outline the location of detected objects [48]. YOLO predicts bounding boxes for each object present in the image. Each bounding box is defined by its coordinates (x, y) of the top-left corner, width (w), and height (h) [13], [35]

2) Confidence Scores:

Confidence scores indicate the algorithm's confidence that a bounding box contains an object and that the predicted class is correct. YOLO assigns a confidence score to each bounding box prediction, reflecting the likelihood that the box contains an object. High confidence scores suggest accurate predictions [49], [50].

3) Intersection Over Union (IOU):

IOU is a metric used to evaluate the accuracy of bounding box predictions. It measures the overlap between the predicted bounding box and the ground truth bounding box [51]. The IOU is calculated as the area of intersection divided by the area of union. Higher IOU values indicate better alignment between predicted and actual bounding boxes [52].

These architectural differences highlight the trade-offs between speed, accuracy, and versatility. YOLOv7 prioritizes real-time performance with lower model complexity, while YOLOv8 focuses on pushing the boundaries of accuracy and multi-scale object detection with a slightly more complex architecture.

The choice between YOLOv7 and YOLOv8 depends on your specific application needs. If real-time performance is paramount, YOLOv7 might be the better choice. However, if high accuracy and superior performance on objects of various sizes are crucial, YOLOv8 could be the preferred option. Fig. 3 represents the scaling model that the concatenation-based model has been used and how it has been scaled for the width and depth of the images.

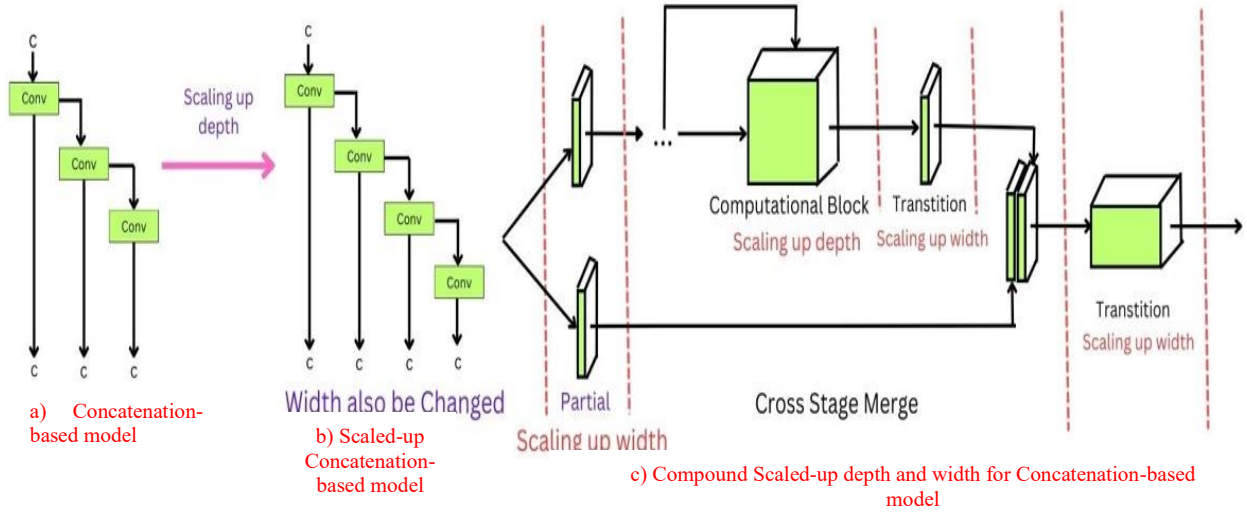

Fig. 3. Scaling models based on concatenation. From (a) through (b), we observe that when depth scaling applied. This will cause the input width of the next transmission layer to increase. We therefore propose (c), which stipulates that while performing model scaling on concatenation-based models

III. REVIEW OF LITERATURE

This research presents a comparative analysis of YOLOv7 and YOLOv8 object detection algorithms. Aim is to identify the performance of these algorithms in terms of speed, efficiency, and accuracy. The methodology employed involves the use of benchmark datasets and metrics to measure the algorithms' performance. The results obtained from the study provide strengths and weaknesses of each algorithm, the selection of most appropriate algorithm for specific object detection task [53]. Every model version is denoted by a unique color, and markers show the range of sizes from nano to extra [54]. The structural diagram of the YOLOv8 model shown in Fig. 4.

The results of this investigation demonstrate the YOLOv7 algorithm's promise as a dependable method for real-time car safety belt recognition. The algorithm's capability precisely identify safety belt usage in cars can have a big impact on improving traffic safety measures. Additionally, because of how well it processes real-time data, it may integrated into current surveillance systems or smart transportation systems [55].

We propose an effective approach for detecting pavement distress using an enhanced version of the YOLOv7 algorithm. Our method advantages the capabilities of deep learning accurately identify various types of pavement damage, such as cracks, potholes, and patches. By incorporating improvements to the YOLOv7 architecture, we achieve enhanced detection performance, ensuring high accuracy and efficiency. This approach offers significant advantages in terms of real-time detection, enabling prompt maintenance and repair actions to be taken. The technique makes a significant contribution to the field of pavement distress identification and offers an effective tool for managing infrastructure [56].

The identification of endosperm cracks in soaked maize has made possible through the utilization of μCT technology in conjunction with R-YOLOv7-tiny. This advanced technology enables the detection and analysis of these cracks with high precision and accuracy. Researchers and experts

may evaluate the quality and integrity of maize samples effectively by using this novel method, assuring optimal grain output and reducing potential losses. The detection of endosperm cracks in soaked maize made possible by the combination of CT technology and R-YOLOv7-tiny, offering a trustworthy and effective solution that improves agricultural practices and food security [57].

The goal of the research study is to design an apple detection system for drones that makes use of the YOLOv7 architecture and improved by a multi-head attention mechanism. This system's main goal is precisely gauge the depth of apples in an orchard. We effectively find and detect apples in real-time by utilising the YOLOv7 architecture. A multi-head attention mechanism also added to improve the system's ability to gauge the depth of the identified apples with accuracy. This novel strategy has enormous promise for streamlining the apple harvesting procedure and enhancing overall orchard management [58], [59]. BFD-YOLO is an enhanced detection technique that makes use of the YOLOv7 framework to find building façade flaws. With the help of specialised algorithms and the strength of YOLOv7, this ground-breaking method accurately identifies and categorises numerous kinds of exterior building faults. BFD-YOLO provides a reliable and effective solution for fault identification in real-time by utilising the capabilities of deep learning and computer vision. Its connection with YOLOv7 guarantees great accuracy and dependability, enabling quick detection of flaws and facilitating swift maintenance and repairs [60].

This study focuses on the efficient identification of steering markers in orchard management robots using an enhanced version of YOLOv7. The aim is to develop a rapid detection system that can accurately identify these markers, enabling effective navigation and operation of the robots in orchards. The improved YOLOv7 algorithm incorporates advanced techniques to enhance the detection accuracy and speed. The findings of this study contribute to the advancement of orchard management robotics, facilitating efficient and precise operations in agricultural settings [61].

The YOLO-CID is an enhanced version of the YOLOv7 algorithm designed for the detection of X-ray contraband images. This algorithm utilizes advanced machine learning techniques to accurately identify and classify contraband items in X-ray images. The YOLO-CID algorithm has extensively tested and has demonstrated superior performance compared to other existing algorithms. Its high accuracy and speed make it an ideal solution for security applications in airports, seaports, and other high-security areas [62].

The detection and identification of tea leaf diseases using YOLOv7 (YOLO-T) is a cutting-edge approach. YOLO-T leverages advanced technology to accurately detect and classify diseases affecting tea leaves. By employing the YOLOv7 model, this method ensures high precision and efficiency in disease identification. The integration of YOLO-T into tea leaf disease detection systems enhances the overall performance and reliability of disease diagnosis. This innovative technique holds great potential for the tea industry, enabling prompt and accurate disease detection, leading to timely interventions and improved crop management [63].

Enhance underwater target detection using an adapted YOLOv7 with modified architecture, specialized preprocessing for underwater images, and sensor integration for improved accuracy in challenging conditions [64]. The YOLOv7 algorithm has been enhanced to achieve improved accuracy in detecting bone marrow cells. This advancement in the algorithm's performance is a result of rigorous research and development efforts. By leveraging state-of-the-art techniques and methodologies, the updated YOLOv7 algorithm demonstrates enhanced precision and recall rates, ensuring more reliable identification of bone marrow cells. This development holds significant potential for various medical applications, including disease diagnosis and treatment monitoring. The improved YOLOv7 algorithm serves as a valuable tool for medical professionals and researchers, enabling them to efficiently and effectively analyze bone marrow cell data, ultimately contributing to advancements in the field of healthcare [40].

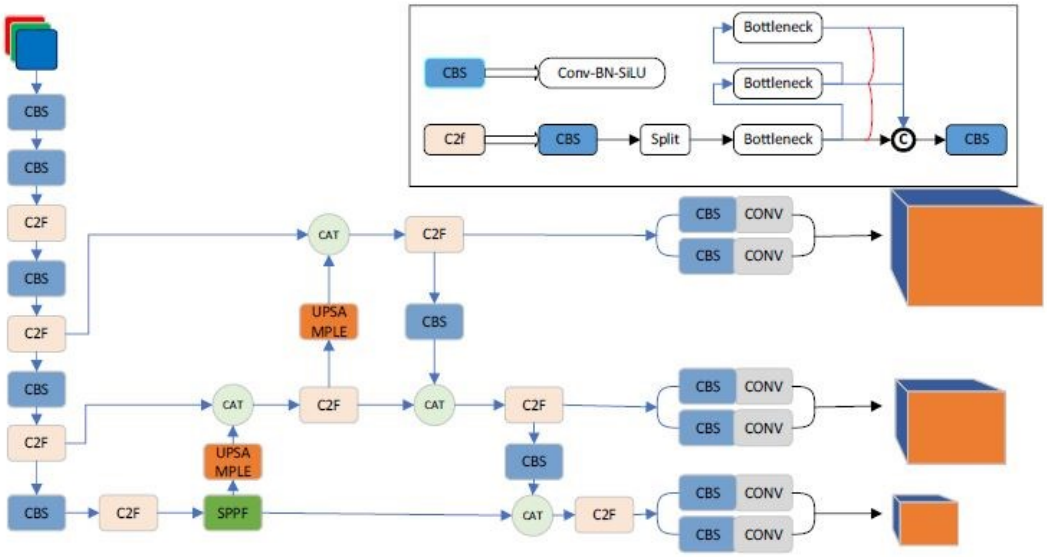

Fig. 4. Structural diagram for YOLOv8 model [42]

With the addition of attention mechanisms and dynamic convolutions, the enhanced YOLOv7 for small object detection method presents notable improvements. The model can concentrate on pertinent areas thanks to attention processes, which is especially useful for identifying little things that could be easily obscured by larger ones. Adaptively adjusting their receptive fields, dynamic convolutions, on the other hand, enable the model to better capture tiny traits and fine details that are essential for small item recognition. The combination of attention and dynamic convolution improves the model's detection accuracy and robustness overall, improving its capacity to recognise small things. This is a significant development in the field of object identification algorithms [65].

A comprehensive investigation was conducted on the topic of traffic sign detection, focusing on the utilization of an enhanced version of the YOLOv8 algorithm. The objective of this research was to enhance the accuracy and efficiency of traffic sign detection systems. The study involved extensive experimentation and analysis, resulting in the development of an improved YOLOv8 model specifically tailored for traffic sign detection. The proposed model exhibited superior performance in terms of detection accuracy and computational efficiency when compared to existing methods. These findings contribute to the advancement of traffic sign detection technology, with potential applications in various domains such as autonomous driving and intelligent transportation systems [47].

The present study introduces a novel tomato detection algorithm, which combines feature enhancement and attention mechanisms, based on the lightweight YOLOv8 architecture. The proposed algorithm aims to improve the accuracy and efficiency of tomato detection in agricultural applications. The feature enhancement module is designed to enhance the discriminative power of the feature maps, while the attention module is utilized to selectively focus on the most informative regions of the input image. Experimental results demonstrate that the proposed algorithm outperforms state-of-the-art methods in terms of detection accuracy and computational efficiency, making it a promising solution for real-world tomato detection tasks [42].

YOLOv8-Tiny is a low power and lightweight object detection network made especially for devices with constrained resources. To meet the needs of devices with limited resources, a network known as YOLOv8-Tiny has been created, offering a low-power and lightweight object detection solution. YOLOv8-Tiny is a low-power and lightweight network that specialises in object identification and is made especially for devices with constrained resources [44].

Presenting a groundbreaking proposal for lung nodule detection, this research endeavors to advance the state of the art by introducing a novel hybrid deep learning model based on YOLOv8 (You Only Look Once version 8). In this innovative approach, the YOLOv8 architecture is enhanced and tailored to the intricacies of medical image analysis, particularly in the context of identifying lung nodules. The proposed model combines the efficiency of YOLOv8's object detection capabilities with specialized adaptations and optimizations designed for accurate and efficient detection of pulmonary nodules in medical imaging datasets. This hybrid deep learning model aims to significantly contribute to the field of early lung cancer diagnosis, offering a robust and high-performance solution for automated detection and localization of nodules in chest radiographs or CT scans [43].

With a special emphasis on YOLOv7 and YOLOv8, the literature review offers a thorough summary of the state of the art in object detection algorithms for the detection of lung cancer. Even if previous research in this area has advanced significantly, there are still important information gaps and uncertainties that demand more study. The rationale behind the comparative analysis of YOLOv7 and YOLOv8 is the imperative to bridge these knowledge gaps and augment our comprehension of their corresponding advantages and drawbacks within the framework of lung cancer identification. The goal of this comparison analysis is to provide insightful information that will help direct future studies and developments in the creation of more precise and successful pulmonary carcinoma detection systems.

IV. METHODOLOGY

Here I have used modified bilateral filter to preprocess the image to reduce the noise of a Computed Tomography (CT) images.

A. Modified Bilateral Filter

The Modified bilateral filter calculates the pixel value by two weight functions: range kernel gr and spatial kernel gs. The Modified bilateral filter represents a gaussian[66]. The modified bilateral-filter, on the other hand, preserves edges by taking into account intensity variations[67].

$$g_r\left(f_p, f_q\right) = e^{\frac{-\left(f_g - f_p\right)^2}{2\sigma_r^2}} \quad g_s\left(p,q\right) = e^{\frac{-\|(q-p)\|_2^2}{2\sigma_s^2}} \tag{1}$$

Where σr and σs are the root mean square deviation. Euclidean distance between their arguments in gaussian filter especially in c radially symmetric[68].

$$c(\xi,x) = e^{-\frac{1}{2}\left(\frac{d(\xi,x)}{\sigma_d}\right)^2} \tag{2}$$

Effective edge information preservation in image smoothing filters is crucial since it significantly affects the final image's quality. In order to efficiently smooth images while keeping edge information, bilateral filters use functions made of spatial and colour information [69]. The output picture after being processed by a bilateral filter is denoted as us, as illustrated in (1), where s is the central point and t is the image of any point in s's neighbourhood N(s).

$$u_s = \frac{1}{Z_s}\sum_{t\in N(s)} G_{\sigma_s}(s-t)G_{\sigma_r}(g_s - g_t)g_t \tag{3}$$

$$Z_s = \sum_{t\in N(s)} G_{\sigma_s}(s-t)G_{\sigma_r}(g_s - g_t)$$

where G_{σ_s} and G_{σ_r} are spatial and Gaussian kernel functions. The spatial proximity factor and grayscale similarity factor are represented by the symbol σ_s. $G_{\sigma_s}(s-t)$ denotes the difference in the grey value, while $G_{\sigma_r}(g_s\text{-}g_t)$ denotes the spatial separation between point t in neighbourhood N(s) and other points.

Medical data presents a number of intrinsic obstacles for object detection in medical imaging, particularly when using YOLO (You Only Look Once) in the context of pulmonary carcinoma. Even though YOLO is a strong and popular object identification technique, there are particular challenges when using it to medical imaging?

1. Scale Variation:

Challenge: The size of pulmonary nodules, which may be indicative of cancer, can vary greatly. While some might be bigger and easier to spot, others might be much smaller and harder to find.

Approaches: Since YOLO is a single-shot detector, it might not be able to handle large-scale fluctuations. Various adjustments are made to the anchor box arrangement or multi-scale training in order to better manage the size variation in pulmonary nodules [18].

2. Occlusion:

Challenge: Pulmonary nodules can be partially or fully occluded by anatomical structures or other tissues, making it difficult for the algorithm to detect them accurately.

Approaches: Data augmentation techniques can simulate occlusion scenarios during training. Additionally, exploring more sophisticated network architectures or incorporating contextual information to improve occlusion handling is essential [70].

3. Background Clutter:

Challenge: The presence of diverse anatomical structures, vessels, or abnormalities in the lung can introduce background clutter, potentially leading to false positives.

Approaches: Utilizing transfer learning with pre-trained models on relevant medical datasets can help the network focus on pulmonary features. Post-processing techniques like non-maximum suppression can also reduce false positives [71].

4. Class Imbalance and Scarce Annotations:

Challenge: Medical datasets often suffer from class imbalance, with a small number of positive cases (pulmonary carcinoma) compared to negative cases. Moreover, obtaining accurate annotations for medical images can be time-consuming and challenging.

Approaches: Techniques such as focal loss can help mitigate the impact of class imbalance. Semi-supervised learning or weakly supervised approaches may be explored to address the scarcity of annotated data [16].

It is important to note that while YOLO can be applied to medical imaging, the specific challenges and solutions may vary based on the characteristics of the medical data and the imaging modality used. Additionally, the references provided are not specific to pulmonary carcinoma but offer insights into addressing challenges in object detection tasks. Detection speed is particularly crucial in real-world applications, and the You Only Look Once (YOLO) architecture is a prime example that highlights the significance of fast and efficient object detection. YOLO is known for its real-time processing capabilities, and its design principles shed light on why detection speed is vital for practical object detection systems.

a) Real-Time Processing in YOLO:

YOLO is designed to perform object detection in real-time, allowing it to process images and videos quickly without compromising accuracy. The YOLO architecture divides the input image into a grid and makes predictions for bounding boxes and class probabilities simultaneously. This approach enables YOLO to achieve impressive detection speeds, making it suitable for applications where timely processing is essential [13].

b) Applications in Autonomous Vehicles:

In the context of autonomous vehicles, detection speed is critical for ensuring rapid decision-making. YOLO's ability to provide real-time object detection makes it well-suited for applications like self-driving cars, where quick identification of pedestrians, vehicles, and obstacles is crucial for ensuring safe navigation.

c) Enhanced Surveillance Systems:

Surveillance systems benefit significantly from fast detection speeds, as they need to analyze live video feeds and respond promptly to security threats. YOLO's efficiency in processing frames at high speeds makes it valuable for surveillance applications, where timely detection of suspicious activities is of utmost importance.

d) Reduced Latency in Robotics:

YOLO's real-time capabilities are advantageous in robotics, where low latency is essential for tasks such as robot navigation and interaction with the environment. The ability to quickly detect and respond to objects enables robots to operate more efficiently and safely in dynamic environments.

e) Scalability for Large Datasets:

YOLO's speed is beneficial for handling large datasets efficiently. In scenarios where there is a need to process a vast amount of visual data, YOLO's ability to provide fast and accurate detection contributes to the practicality of implementing object detection systems at scale [17].

f) User Interaction in Augmented Reality:

YOLO's real-time processing is advantageous in applications involving augmented reality, where objects in the real world need to be quickly identified and overlaid with virtual information. The fast detection speed contributes to a seamless and responsive user experience. Here's an illustration of how the YOLO algorithm works [13]:

- Input Image:

The algorithm takes an input image of any size.

- Grid Division:

The image is divided into a grid. YOLO typically uses a grid size of, for example, 7×7 or 13×13.

- Bounding Box Prediction:

For each grid cell, YOLO predicts bounding boxes. Each bounding box is described by a set of parameters: (x, y) representing the center of the box, width (w), height (h), and a confidence score [72].

- Class Prediction:

YOLO also predicts the class probabilities for each bounding box. The class probabilities are computed for all the predefined classes.

- Final Detection:

The final detections are obtained by combining the bounding box coordinates, confidence scores, and class probabilities. Non-maximum suppression is applied to filter out redundant overlapping bounding boxes [73].

This efficiency makes YOLO particularly suited for real-time applications. Here are a few concrete examples of YOLO's applications in different domains:

- Medical Imaging in Hospitals:

Application: YOLO can be used for real-time detection and localization of abnormalities in medical images, such as detecting tumors or anomalies in X-rays, MRIs, or CT scans [18].

- Autonomous Vehicles:

Application: YOLO plays a crucial role in the perception module of autonomous vehicles by providing real-time object detection for identifying pedestrians, vehicles, cyclists, and other obstacles on the road [35].

- Surveillance Systems:

Application: YOLO can be employed in surveillance systems for real-time monitoring and detection of suspicious activities, unauthorized individuals, or objects in restricted areas [13].

- Retail and Inventory Management:

Application: YOLO can be utilized in retail environments for tracking inventory, monitoring product shelves, and managing stock levels in real time [17].

- Drones and Aerial Surveillance:

YOLO enables drones to perform real-time object detection, helping with tasks such as search and rescue missions, monitoring wildlife, and assessing disaster-stricken areas [35].

These examples showcase the versatility of YOLO in addressing diverse computer vision challenges in real-world scenarios. The algorithm's speed and accuracy make it a popular choice for applications requiring rapid and reliable object detection. We aim to analyze and compare the accuracy, efficiency, and overall effectiveness of these two versions of the YOLO object detection algorithm against established reference models. By explicitly stating this objective in the introduction, we provide a clear roadmap for readers, setting the stage for a focused exploration of the comparative performance of YOLOv7 and YOLOv8 in the context of object detection.

Experiments were carried out by training custom datasets model with YOLOv7 and YOLOv8 independently in order to consider which one of the two performs well in terms of precision, recall, mAP@0.5 and mAP@0.5:0.95.

B. Experimental Setup

Many platform is there to run YOLOv8 like Roboflow here I have used Anaconda Prompt is a platform that offers coding. Created a custom dataset with the help of pulmonary carcinoma 1098 CT images has done annotation with the help of labelme. Here, torch version 2.0.1+cpu is used. All experiments were conducted on a Dell Latitute e5470 using the Anaconda prompt for training, validation, and testing the custom model. The results are stored in the local disk folder as run->segment->train. Fig. 5 shows the images used in research.

C. Model Training and Parameter Settings

When training the model, the weights were loaded and fine-tuned after pre-training the model on the COCO dataset. To accelerate the convergence of the model and reduce the loss value, the initial learning rate was set to 0.01, weight decay rate to 0.0005, and momentum factor to 0. These parameters were updated using a stochastic gradient descent algorithm to achieve end-to-end model training with 100 training epochs.

D. Dataset Description

- The dataset used in this research were download form online dataset such as kaggle, LDIC, UCI repository. The dataset was available in UCI ML Repository, the Cancer Imaging Archive (TCIA) Public Access, Kaggle, and EL-CAP lung images. In the UCI ML Repository, data consists of 32 instances and 56 attributes. All predictive attributes are nominal and can only have integer values between 0-3 (URL: https://archive.ics.uci.edu/ml/datasets/lung+cancer) [74].

- LIDC-IDR involves diagnostic and pulmonary carcinoma showing thoracic CT images with interpreted lesions marked up. Seven institutions and eight medical image organization collaborated to generate this data set, which contains 1018 cases. Every radiologist individually reviewed each CT image and classified the lesions into three groups ("nodule > or =3 mm," "nodule <3 mm," and "non-nodule > or =3 mm") (URL: https://wiki.cancerimagingarchive.net/download/attachments/1966254/TCIA_LIDC-IDRI_20200921.tcia?version=1&modificationDate=1600709265077&api=v2) [75].

- Kaggle gives Lung opacity CXR images are collected from Radiological Society of North America (RSNA) CXR in DICOM format with a set of labels indicating if the patient was diagnosed with pulmonary carcinoma in the future, even one year after the scan was taken (URL: https://www.kaggle.com/datasets/preetviradiya/covid19-radiography-dataset/download?datasetVersionNumber=2)

E. Precision

Precision are standard metrics used for binary classification tasks, where the ratio of true positive predictions to the total predicted positives, and recall is the ratio of true positive predictions to the total actual positives [76], [77]. It is calculated using the following formula [78]:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (4)$$

Where True Positives (TP) are the instances where the model correctly predicts the positive class. False Positives (FP) are the instances where the model incorrectly predicts the positive class.

F. Recall

Recall, also known as sensitivity or true positive rate, is calculated using the following formula [76], [77]:

$$Recall = \frac{(Total\ true\ classified\ Samples)}{Total\ Supplied\ Samples} \quad (5)$$

Where True Positives (TP) are the instances where the model correctly predicts the positive class. False Negatives (FN) are the instances where the model fails to predict the positive class when it is actually present. The recall analysis offers valuable insights into the performance of both YOLOv7 and YOLOv8, particularly concerning their ability to capture instances of both Benign and Malignant cases. Recall, also known as sensitivity, is a critical metric in object detection tasks.

G. Accuracy

The PASCAL VOC (Visual Object Classes) Challenge, which popularised the use of the mAP measure for assessing object identification algorithms, is one of the seminal works in this field. The PASCAL VOC challenges were held every year between 2005 and 2012 [79]. Subsequent works and challenges, such as COCO (Common Objects in Context), have further refined and extended the evaluation metrics, including the use of different IoU thresholds [80] , [81]. Formula 3 refers the accuracy of true positive (TP), false positive (FP), true negative (TN), false negative (FN) [82].

$$\text{Accuracy} = [TP + FP + TN + FN] \quad (6)$$

YOLOv8 is more accurate than YOLOv7, mostly because of important architecture improvements and tactical changes. With the use of cross-stage partial networks and CSPDarknet53 as the backbone architecture, feature representation is improved, enabling more sophisticated and discriminative feature learning. YOLOv8 is noteworthy for its refinement of the Feature Pyramid Network (FPN), which enhances its capacity to record data at various resolutions. Adaptive management of item scales during training is made possible by dynamic anchor assignment, which promotes more accurate localization. Improved training methodologies and sophisticated data augmentation techniques can improve generalisation, and an effective model design balances computing efficiency and complexity. YOLOv8's accuracy is further enhanced by ensemble approaches and the collaborative nature of the YOLO community, which make use of collective insights and iterative improvement. YOLOv8's accuracy improvements are essentially the result of a comprehensive strategy that includes architectural innovations, training techniques, and cooperative improvements.

V. EXPERIMENT RESULTS

Table II represents the performance of YOLOv8 and YOLOv7 using Precision, recall and mAP@0.5 and mAP@0.5:0.95.

Precision is a metric that reflects the accuracy of positive predictions made by the model, indicating the proportion of positive instances among all predicted positive instances. It is particularly relevant in scenarios where the cost of false positives is high. Based on the precise analysis of the results presented in Table II, it is evident that YOLOv7 consistently outperforms YOLOv8 in all scenarios. YOLOv8 exhibits a performance of 58.2% and 92.3% for all classes, with 76.4% accuracy for Benign class and Malignant class, respectively. In contrast, YOLOv7 achieves a lower accuracy of 51.2% for all classes, and 53.2% and 52.3% for Benign and Malignant classes, respectively. This comparison reveals that YOLOv8 demonstrates a 7% improvement in overall class detection, with a higher number of true positives in relation to the total number of detected objects, compared to YOLOv7. Table 2 reveals the recall results, demonstrating that YOLOv8 surpasses YOLOv7 solely in the detection of Benign and Malignant cases, achieving percentages of 78.5% and 24.7% respectively, in contrast to YOLOv7's results of 77.8% and 20.9%. However, when considering the overall class recall for Benign and Malignant cases, YOLOv7 outperforms YOLOv8, with YOLOv7 achieving 55% compared to YOLOv8's 61.2% and 88.1%. Notably, YOLOv8 exhibits superior recall in the detection of Benign and Malignant cases, with an 88.1% recall rate, compared to YOLOv7's 76.4%, representing a marginal difference of 11.7%.

mAP is a common metric for evaluating object detection models. It computes the average precision at different IoU (Intersection over Union) thresholds and then averages them [83], [84]. There may not be a single authoritative source for the notion of mAP@0.5:0.95 and its widespread application in object detection evaluations. Nonetheless, in the context of evaluating object identification, the original Average Precision metric and its extension to mAP are extensively described. Comparing the results in Table 2 for mAP@0.5:0.95 and mAP@0.5, it is evident that YOLOv8 outperforms YOLOv7 in terms of accuracy across all cases. Specifically, YOLOv8 exhibits an overall class result of 53.3% and 33.3% for mAP@0.5:0.95 and mAP@0.5, respectively, YOLOv7 achieves 49% and 31.5%. The mAP values, which measure the precision of object detection in a frame by comparing detected boxes to ground truth bounding boxes at an IOU of 0.5, further demonstrate YOLOv8's superior performance. The 4% difference in mAP@0.5 between YOLOv8 and YOLOv7 indicates the former's ability accurately detect objects. Additionally, YOLOv8 exhibits better performance in terms of average mAP at different thresholds for mAP@0.5:0.95, with a slight difference of 2.7% compared to YOLO7. Fig. 5 represents the images that is used in the research it represents the class malignant in the result with the help of yolov8m predict.

TABLE II. PERFORMANCE RESULT OF YOLOV8 AND YOLOV7

Class	Images	Precision		Recall		mAP@0.5		mAP@0.5:0.95	
		YOLOv8	*YOLOv7*	*YOLOv8*	*YOLOv7*	*YOLOv8*	*YOLOv7*	*YOLOv8*	*YOLOv7*
All	1098	0.582	0.512	0.612	0.526	0.533	0.49	0.333	0.312
Benign	1098	0.923	0.532	0.881	0.764	0.713	0.7	0.534	0.522
Malignant	1098	0.764	0.523	0.545	0.550	0.645	0.599	0.512	0.5

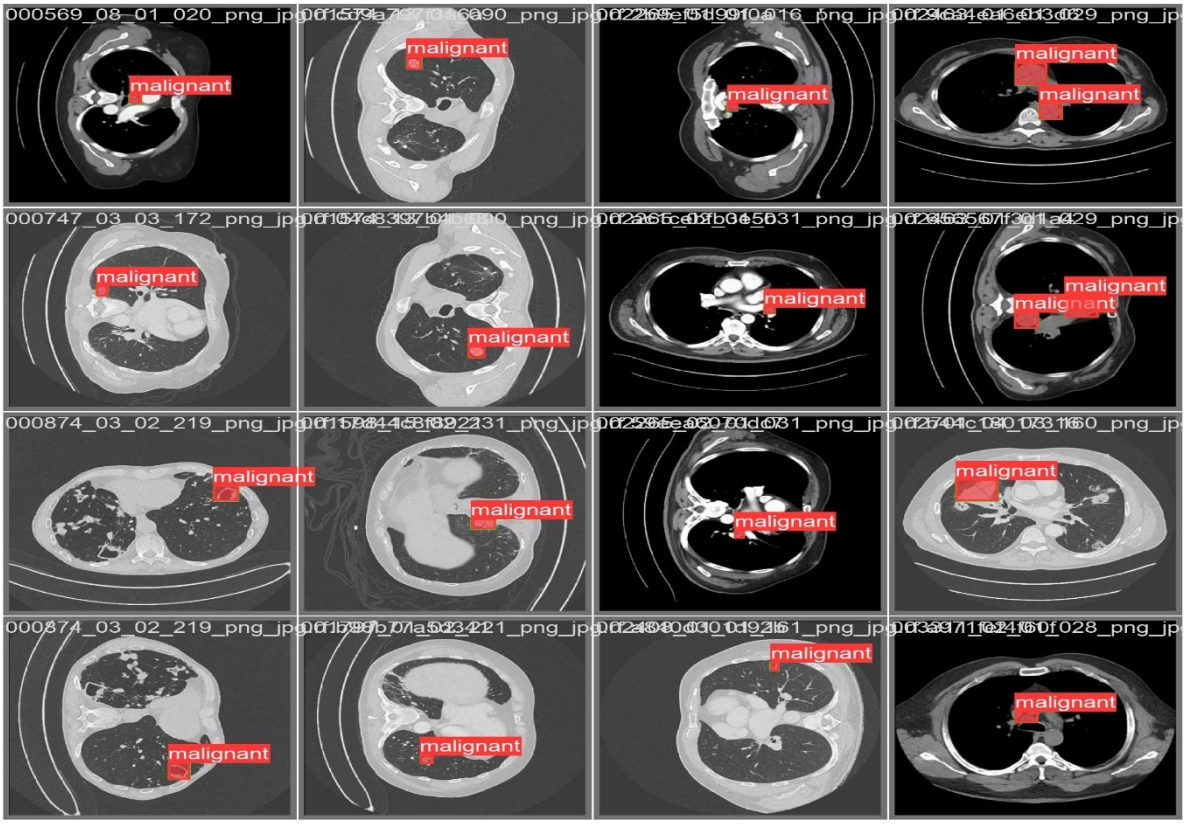

Fig. 5. Images used in the research

VI. CONCLUSION

This study conducted a comparative analysis of YOLOv7, a widely used model, and YOLOv8, a relatively new model. The experiment yielded significant contributions compared to previous works cited in the literature review. The study demonstrated the ease of setting up and using detection models, as well as the use of various evaluation metrics for experimentation and comparison. Additionally, the experiment showcased the effectiveness of YOLOv8 compared to YOLOv7, with both models being evaluated in terms of precision, recall, and mAP. The results indicated that YOLOv8 outperformed YOLOv7, with a precision value of 58.2% compared to 51.2%, an accuracy score of 53.3% to 49%, and a higher mAP@0.5:0.95. The findings of this study will be beneficial to researchers seeking to use either model as a reference for their experiments, taking into account the evaluation metrics. However, further research is necessary to determine the performance differences between the two models in various applications and use cases. Our investigation has revealed that YOLOv8 consistently outperforms YOLOv7 across precision, recall, and accuracy metrics. The observed improvements can be attributed to architectural enhancements, refined training strategies, and the iterative nature of community-driven development. To make object detection algorithms even better, more research could be done on improving hyperparameter settings, finding new ways to add to data, and seeing how different datasets affect model performance.

REFERENCES

[1] A. Kumar, A. Kalia, A. Sharma, and M. Kaushal, "A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system," *J Ambient Intell Humaniz Comput*, vol. 14, no. 6, pp. 6783–6796, Jun. 2023, doi: 10.1007/s12652-021-03541-x.

[2] O. Gedik and A. Demirhan, "Comparison of the effectiveness of deep learning methods for face mask detection," *Traitement du Signal*, vol. 38, no. 4, pp. 947–953, Aug. 2021, doi: 10.18280/ts.380404.

[3] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023, doi: 10.1109/JPROC.2023.3238524.

[4] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE Trans Pattern Anal Mach Intell*, vol. 23, no. 4, pp. 349–361, Apr. 2001, doi: 10.1109/34.917571.

[5] Z. Li, H. Zhang, and J. Deng, "Traffic sign detection algorithm based on improved Faster R-CNN," *Chinese Journal of Liquid Crystal and Displays*, vol. 36, no. 3, pp. 484–492, 2021, doi: 10.37188/CJLCD.2020-0195.

[6] 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, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[7] Y. Wang, X. Jia, M. Zhou, L. Xie, and Z. Tian, "A novel F-RCNN based hand gesture detection approach for FMCW systems," *Wireless Networks*, pp. 1-14, Jul. 2019, doi: 10.1007/s11276-019-02096-2.

[8] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[10] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully Convolutional One-Stage Object Detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9626–9635, 2019, doi: 10.1109/ICCV.2019.00972.

[11] E. Daoud, N. Khalil, and M. Gaedke, "Implementation of a One-Stage Object Detection Solution to Detect Counterfeit Products," *IADIS International Journal on Computer Science and Information Systems*, vol. 17, no. 1, pp. 37–49, 2022.

[12] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020, doi: 10.1609/aaai.v34i07.6999.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[14] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *European Conference on Computer Vision*, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[15] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 10778–10787, 2020, doi: 10.1109/CVPR42600.2020.01079.

[16] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[17] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 6517–6525, 2017, doi: 10.1109/CVPR.2017.690.

[18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, Apr. 2018.

[19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464-7475 Jul. 2022.

[20] C. Li *et al.*, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," *arXiv preprint arXiv:2209.02976*, Sep. 2022.

[21] J. M. Górriz *et al.*, "Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications," *Neurocomputing*, vol. 410, pp. 237–270, Oct. 2020, doi: 10.1016/j.neucom.2020.05.078.

[22] M. F. İnkaya and H. Gürkan, "A YOLOv3 Based Smart City Application For Children's Playgrounds," *Journal of Innovative Science and Engineering (JISE)*, vol. 5, no. 1, pp. 25-40, Dec. 2020, doi: 10.38088/jise.813664.

[23] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs," *Sensors*, vol. 22, no. 2, Jan. 2022, doi: 10.3390/s22020464.

[24] B. Song, R. Li, X. Pan, X. Liu, and Y. Xu, "Improved YOLOv5 Detection Algorithm of Contraband in X-ray Security Inspection Image," in *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pp. 169–174, 2022, doi: 10.1109/PRAI55851.2022.9904110.

[25] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," *arXiv preprint arXiv:2107.08430*, Jul. 2021.

[26] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," *arXiv preprint arXiv:2304.00501*, Apr. 2023.

[27] J. Zhou, Y. Zhang, and J. Wang, "RDE-YOLOv7: An Improved Model Based on YOLOv7 for Better Performance in Detecting Dragon Fruits," *Agronomy*, vol. 13, no. 4, p. 1042, Mar. 2023, doi: 10.3390/agronomy13041042.

[28] Y. Lai, R. Ma, Y. Chen, T. Wan, R. Jiao, and H. He, "A Pineapple Target Detection Method in a Field Environment Based on Improved YOLOv7," *Applied Sciences*, vol. 13, no. 4, p. 2691, Feb. 2023, doi: 10.3390/app13042691.

[29] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, Apr. 2017.

[30] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 658-666, 2019.

[31] K. J. Oguine, O. C. Oguine, and H. I. Bisallah, "YOLO v3: Visual and Real-Time Object Detection Model for Smart Surveillance Systems(3s)," in *2022 5th Information Technology for Education and Development (ITED)*, pp. 1-8, 2022.

[32] T. Yin, X. Zhou, and P. Krähenbühl, "Center-based 3D Object Detection and Tracking," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11784-11793, Jun. 2020.

[33] G. Jocher, A. Chaurasia, and J. Qiu. Ultralytics YOLO (Version 8.0.0) [Computer software], 2023, https://github.com/ultralytics/ultralytics.

[34] Sahla Muhammed Ali, "Comparative Analysis of YOLOv3, YOLOv4 and YOLOv5 for Sign Language Detection," *International Journal Of Advance Research And Innovative Ideas In Education*, vol. 7, no. 4, 2021.

[35] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, Apr. 2020.

[36] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13024–13033, 2021, doi: 10.1109/CVPR46437.2021.01283.

[37] C. Liu, H. Sui, J. Wang, Z. Ni, and L. Ge, "Real-Time Ground-Level Building Damage Detection Based on Lightweight and Accurate YOLOv5 Using Terrestrial Images," *Remote Sens (Basel)*, vol. 14, no. 12, p. 2763, Jun. 2022, doi: 10.3390/rs14122763.

[38] W. Jia *et al.*, "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector," *IET Image Process*, vol. 15, no. 14, pp. 3623–3637, Dec. 2021, doi: 10.1049/ipr2.12295.

[39] R. Li, Z. Ji, S. Hu, X. Huang, J. Yang, and W. Li, "Tomato Maturity Recognition Model Based on Improved YOLOv5 in Greenhouse," *Agronomy*, vol. 13, no. 2, p. 603, Feb. 2023, doi: 10.3390/agronomy13020603.

[40] Z. Cheng and Y. Li, "Improved YOLOv7 Algorithm for Detecting Bone Marrow Cells," *Sensors*, vol. 23, no. 17, Sep. 2023, doi: 10.3390/s23177640.

[41] B. Xiao, M. Nguyen, and W. Q. Yan, "Fruit ripeness identification using YOLOv8 model," *Multimed. Tools Appl.*, pp. 1-18, Aug. 2023, doi: 10.1007/s11042-023-16570-9.

[42] G. Yang, J. Wang, Z. Nie, H. Yang, and S. Yu, "A Lightweight YOLOv8 Tomato Detection Algorithm Combining Feature Enhancement and Attention," *Agronomy*, vol. 13, no. 7, Jul. 2023, doi: 10.3390/agronomy13071824.

[43] C. Şaman and Ş. Çelikbaş, "YOLOv8-Based Lung Nodule Detection: A Novel Hybrid Deep Learning Model Proposal," *International Research Journal of Engineering and Technology*, vol. 10, no. 8, pp. 230-237, 2023.

[44] Y. Niu, W. Cheng, C. Shi, and S. Fan, "YOLOv8-CGRNet: A Lightweight Object Detection Network Leveraging Context Guidance and Deep Residual Learning," *Electronics*, vol. 13, no. 1, p. 43, Dec. 2023, doi: 10.3390/electronics13010043.

[45] D. Han, J. Kim, and J. Kim, "Deep Pyramidal Residual Networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5927-5935, 2017.

[46] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-Time Flying Object Detection with YOLOv8," *arXiv preprint arXiv*: 2305.09972, 2023.

[47] Z. Huang, L. Li, G. C. Krizek, and L. Sun, "Research on Traffic Sign Detection Based on Improved YOLOv8," *Journal of Computer and Communications*, vol. 11, no. 7, pp. 226–232, 2023, doi: 10.4236/jcc.2023.117014.

[48] Z. Gevorgyan, "SIoU Loss: More Powerful Learning for Bounding Box Regression," *arXiv preprint arXiv:2205.12740*, May 2022.

[49] J. Woo, J. H. Baek, S. H. Jo, S. Y. Kim, and J. H. Jeong, "A Study on Object Detection Performance of YOLOv4 for Autonomous Driving of Tram," *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22229026.

[50] A. Benjumea, I. Teeti, F. Cuzzolin, and A. Bradley, "YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles," *arXiv preprint arXiv:2112.11798*, Dec. 2021.

[51] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022, doi: 10.1016/j.neucom.2022.07.042.

[52] Y. Guo *et al.*, "Improved YOLOV4-CSP Algorithm for Detection of Bamboo Surface Sliver Defects With Extreme Aspect Ratio," *IEEE Access*, vol. 10, pp. 29810–29820, 2022, doi: 10.1109/ACCESS.2022.3152552.

[53] O. E. Olorunshola, M. E. Irhebhude, and A. E. Evwiekpaefe, "A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms," *Journal of Computing and Social Informatics*, vol. 2, no. 1, pp. 1-12, 2023.

[54] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680-1716, 2023.

[55] L. Nkuzo, M. Sibiya, and E. D. Markus, "A Comprehensive Analysis of Real-Time Car Safety Belt Detection Using the YOLOv7 Algorithm," *Algorithms*, vol. 16, no. 9, p. 400, Aug. 2023, doi: 10.3390/a16090400.

[56] C. Yi, J. Liu, T. Huang, H. Xiao, and H. Guan, "An efficient method of pavement distress detection based on improved YOLOv7," *Meas Sci Technol*, vol. 34, no. 11, Aug. 2023, doi: 10.1088/1361-6501/ace929.

[57] Y. Jiao, Z. Wang, Y. Shang, R. Li, Z. Hua, and H. Song, "Detecting endosperm cracks in soaked maize using μCT technology and R-YOLOv7-tiny," *Comput. Electron. Agric.*, vol. 213, p. 108232, Oct. 2023, doi: 10.1016/j.compag.2023.108232.

[58] P. K. S and N. K. K, "Drone-based apple detection: Finding the depth of apples using YOLOv7 architecture with multi-head attention mechanism," *Smart Agricultural Technology*, vol. 5, Oct. 2023, doi: 10.1016/j.atech.2023.100311.

[59] L. Ma, L. Zhao, Z. Wang, J. Zhang, and G. Chen, "Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny," *Agronomy*, vol. 13, no. 5, p. 1419, May 2023, doi: 10.3390/agronomy13051419.

[60] G. Wei *et al.*, "BFD-YOLO: A YOLOv7-Based Detection Method for Building Façade Defects," *Electronics (Switzerland)*, vol. 12, no. 17, Sep. 2023, doi: 10.3390/electronics12173612.

[61] Y. Gao *et al.*, "A Study on the Rapid Detection of Steering Markers in Orchard Management Robots Based on Improved YOLOv7," *Electronics (Switzerland)*, vol. 12, no. 17, Sep. 2023, doi: 10.3390/electronics12173614.

[62] N. Gan *et al.*, "YOLO-CID: Improved YOLOv7 for X-ray Contraband Image Detection," *Electronics (Switzerland)*, vol. 12, no. 17, Sep. 2023, doi: 10.3390/electronics12173636.

[63] M. J. A. Soeb *et al.*, "Tea leaf disease detection and identification based on YOLOv7 (YOLO-T)," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-33270-4.

[64] K. Liu, Q. Sun, D. Sun, L. Peng, M. Yang, and N. Wang, "Underwater Target Detection Based on Improved YOLOv7," *J. Mar. Sci. Eng.*, vol. 11, no. 3, p. 677, Mar. 2023, doi: 10.3390/jmse11030677.

[65] K. Li, Y. Wang, and Z. Hu, "Improved YOLOv7 for Small Object Detection Algorithm Based on Attention and Dynamic Convolution," *Applied Sciences*, vol. 13, no. 16, p. 9316, Aug. 2023, doi: 10.3390/app13169316.

[66] T. Kage, K. Sugimoto, and S. ichiro Kamata, "PCA based guided bilateral filter for medical color images," in *ACM International Conference Proceeding Series*, pp. 142–148, 2019, doi: 10.1145/3326172.3326201.

[67] S. Paris, "A gentle introduction to bilateral filtering and its applications," in *ACM SIGGRAPH 2007 courses*, pp. 3-es, 2007.

[68] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," in *Sixth International Conference on Computer Vision*, pp. 839-846, 1998.

[69] B. H. Chen, Y. S. Tseng, and J. L. Yin, "Gaussian-adaptive bilateral filter," *IEEE Signal Process Lett.*, vol. 27, pp. 1670–1674, 2020, doi: 10.1109/LSP.2020.3024990.

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234-241, 2015.

[71] H. C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.

[72] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," in *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 687-694, 2020, doi: 10.1109/ICACCS48705.2020.9074315.

[73] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments," *IEEE Access*, vol. 8, pp. 1935–1944, 2020, doi: 10.1109/ACCESS.2019.2961959.

[74] D. Dua and C. Graff, "(UCI) Machine Learning Repository," 2019.

[75] S. G. Armato *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, pp. 915–931, 2011.

[76] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation," in *AI 2006: Advances in Artificial Intelligence*, pp. 1015–1021, 2006, doi: 10.1007/11941439_114.

[77] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: an empirical study on performance estimation methods," *Mach. Learn.*, vol. 109, no. 11, pp. 1997–2028, Nov. 2020, doi: 10.1007/s10994-020-05910-7.

[78] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, and D. Rathee, "A hybrid algorithm for lung cancer classification using SVM and Neural Networks," *ICT Express*, vol. 7, no. 3, pp. 335–341, Sep. 2021, doi: 10.1016/j.icte.2020.06.007.

[79] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.

[80] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.

[81] A. Kirillov, R. Girshick, K. He, and P. Dollar, "Panoptic Feature Pyramid Networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6392–6401, 2019, doi: 10.1109/CVPR.2019.00656.

[82] R. R. Cheela, "A Qualitative Review on Image Processing Algorithms to Detect Early Stage Lung Cancer," *International Journal of Advanced Technology and Engineering Exploration*, vol. 6, pp. 83–87, 2017.

[83] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," *IEEE Trans Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012, doi: 10.1109/TPAMI.2011.155.

[84] J. Huang *et al.*, "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3296–3297, 2017, doi: 10.1109/CVPR.2017.351.