IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 February 2024, accepted 7 April 2024, date of publication 10 April 2024, date of current version 29 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3386826

TOPICAL REVIEW

A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)

MOHAMMED GAMAL RAGAB[1,2], SAID JADID ABDULKADIR[1,2], (Senior Member, IEEE), AMGAD MUNEER[1], ALAWI ALQUSHAIBI[1,2], EBRAHIM HAMID SUMIEA[1,2], RIZWAN QURESHI[3], (Senior Member, IEEE), SAFWAN MAHMOOD AL-SELWI[1,2], AND HITHAM ALHUSSIAN[1,2]

[1]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[2]Centre for Research in Data Science, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[3]Fast School of Computing, National University of Computer and Emerging Sciences, Karachi 75030, Pakistan

Corresponding author: Rizwan Qureshi (engr.rizwanqureshi786@gmail.com)

This work was supported by the Ministry of Higher Education (MOHE), Malaysia for providing financial assistance under Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UTP/02/4) and Universiti Teknologi PETRONAS under the Yayasan Universiti Teknologi PETRONAS (YUTP-FRG/015LC0-308) for providing the required facilities to conduct this research work.

ABSTRACT YOLO (You Only Look Once) is an extensively utilized object detection algorithm that has found applications in various medical object detection tasks. This has been accompanied by the emergence of numerous novel variants in recent years, such as YOLOv7 and YOLOv8. This study encompasses a systematic exploration of the PubMed database to identify peer-reviewed articles published between 2018 and 2023. The search procedure found 124 relevant studies that employed YOLO for diverse tasks including lesion detection, skin lesion classification, retinal abnormality identification, cardiac abnormality detection, brain tumor segmentation, and personal protective equipment detection. The findings demonstrated the effectiveness of YOLO in outperforming alternative existing methods for these tasks. However, the review also unveiled certain limitations, such as well-balanced and annotated datasets, and the high computational demands. To conclude, the review highlights the identified research gaps and proposes future directions for leveraging the potential of YOLO for medical object detection.

INDEX TERMS YOLO, healthcare applications, artificial intelligence, medical object detection, medical imaging, systematic review.

I. INTRODUCTION

Object detection is an essential task in computer vision, with numerous applications in various domains, including medical imaging [1], [2], surgical procedures [3], and personal protective equipment detection [4]. It plays a crucial role in medical imaging by enabling the identification and localization of abnormalities or objects of interest within medical images [5], [6]. Medical diagnosis relies heavily on the accurate detection and localization of abnormalities in medical images. The traditional approach to object detection in medical imaging involves manual annotation and segmentation of the regions of interest, which is a time-consuming and error-prone process [7]. In recent years, artificial intelligence methods have shown great promise for healthcare applications [8], [9], [10], [11], [12]. Deep learning-based object detection algorithms have shown exceptional performance in real time object detection and localization [13], [14], [15]. Accurate and efficient object detection algorithms are essential for assisting healthcare professionals in diagnosing and treating various medical conditions [16], [17], [18]. One such object detection algorithm is You Only Look Once (YOLO) [19], [20], [21], [22], which has gained significant attention.

YOLO is a state-of-the-art, real time, end-to-end object detection algorithm that has gained significant attention in the computer vision community [23]. YOLO is a single-stage detector, which means that it can detect all the objects in an image in a single forward pass through a convolutional neural

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Loconsole.

network (CNN). It uses a single neural network to predict the bounding boxes and class probabilities of the objects present in an image [24]. This makes YOLO very fast, and it can achieve real time speeds on even a modest GPU [25].

The application of YOLO in the medical domain has garnered interest due to its ability to detect and localize anatomical structures [26], [27], lesions [28], [29], tumors [30], [31], [32], and other clinically relevant medical objects [33], [34]. It can detect and localize abnormalities in medical images, which can aid in the early detection and diagnosis of various diseases, including breast cancer, lung cancer, narrowing of blood vessels [35], brain atrophy [36], and abnormal protein deposits [37], cardiovascular diseases [38], and neurological disorders [39]. The adoption of YOLO in medical applications has the potential to improve the accuracy and efficiency of medical diagnosis, which can have a significant impact on patient outcomes.

The real time performance of YOLO makes it particularly appealing for time-sensitive medical procedures and clinical decision-making [40]. By accurately and efficiently identifying objects of interest, YOLO can potentially aid in early disease detection, treatment planning, and monitoring of disease progression. However, as the adoption of YOLO in medical imaging increases, it is essential to evaluate its performance, strengths, limitations, and the specific medical domains in which it has been applied [41]. Therefore, a systematic literature review (SLR) may provide a comprehensive and rigorous approach to analyze the existing literature on YOLO in medical applications. By systematically collecting, evaluating, and synthesizing the available evidence, this review aims to identify the strengths, limitations, and potential of YOLO in medical applications. The findings of this review will assist researchers, healthcare professionals, and developers in understanding the performance of YOLO and its suitability for different medical object detection tasks.

Survey Motivation: There are already existing review articles on YOLO, such as algorithmic developments in YOLO [25], challenges and architectural developments for object detection using YOLO in [23], and a review on object detection techniques in [42]. A survey on object detection for medical images using deep learning techniques was published in [43] and a comprehensive analysis of applying object detection methods for medical image analysis in [44]. In this paper, we focused on the YOLO architecture, its evolution, and applications for three key medical applications; medical images, personal protective equipment detection, and surgical procedures. To the best of our knowledge, it is the first article to discuss three key medical applications using the YOLO series architecture.

The main contributions of this paper are:

- Identify and select relevant peer-reviewed articles published between 2018 and 2023 that focus on the application of YOLO in medical imaging.
- Analyze and summarize the characteristics of the selected studies, including the medical domains, datasets, evaluation metrics, and findings.
- Evaluate the performance of YOLO in medical applications by synthesizing its accuracy, precision, recall, and other relevant metrics as reported in the selected studies.
- Identify common challenges, limitations, and gaps in the existing literature on the use of YOLO in medical imaging.
- Provide insights and recommendations for future research directions, improvements, and potential applications of YOLO in the medical domain.

The rest of this section contains the following sub-sections: You Only Look Once (I-A), YOLO algorithm and architecture (I-B), YOLO in action (I-C), image annotation (I-D), how YOLO operates (I-E), and advantages and drawbacks of YOLO (I-F).

A. YOU ONLY LOOK ONCE (YOLO)

This sub-section provides a brief description of YOLO, its versions, structure, and how it works. YOLO, proposed by Redmon et al. [45], is an object detection algorithm that uses convolutional neural networks (CNN) [46], [47] to detect objects in real time [25]. It is a single-stage method that can achieve real time performance on a standard GPU [48]. It divides the image into a grid of cells, and each cell is responsible for detecting objects within a certain area [49], which allows for faster object detection compared to traditional two-stage methods and is particularly useful for real time applications. It has evolved over multiple versions, each offering improvements in speed, detection accuracy, and capability to detect objects of varying sizes [50].

YOLO exhibits a high level of generalizability, making it less prone to failure when applied to novel domains or unexpected situations [23]. Unlike previous approaches that repurpose classifiers for detection, YOLO is a versatile detector that learns to detect various objects. It acquires generalized representations of objects, enabling it to surpass leading detection methods like Deformable Parts Model (DPM) [51] and Region-based Convolutional Neural Network (R-CNN) [52], [53] by a good margin. However, YOLO has some problems with detecting small objects and will do worse with scenes with many overlapping objects. The main advantage of YOLO lies in its real time object detection capability, which is crucial in time-sensitive applications.

The YOLO architecture has evolved significantly from its inception in v1 to the cutting-edge advancements in v8 as shown in Figure 1. With v1, the initial foundation was laid, introducing a groundbreaking concept of real time object detection through a single network pass in 2015. YOLOv2 (2016): Improves YOLOv5 by using a larger input size, more anchor boxes, and a new loss function. YOLOv3 (2018): Introduces a new network architecture called Darknet-53, which is deeper and more accurate than the previous architectures used in YOLO.

YOLOv4 (2019): Improves upon YOLOv3 by using a new training method called Mosaic data augmentation, which helps to improve the model's robustness to different object scales and orientations. YOLOv5 (2020): Introduces a new

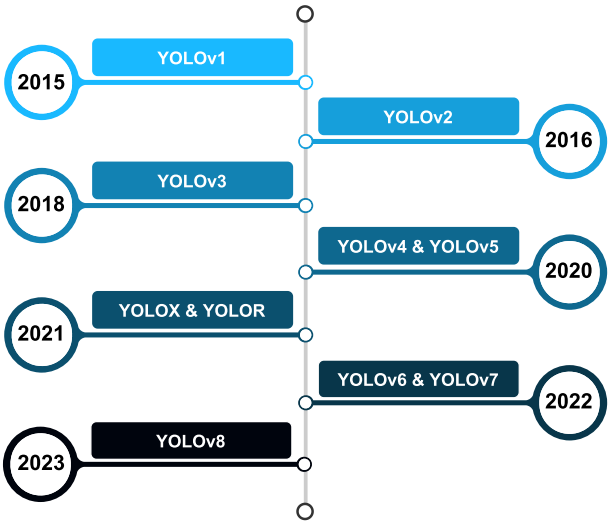


FIGURE 1. Timeline of YOLO versions.

network architecture called CSPDarknet, which is more efficient than Darknet-53. YOLOv6 (2021): Improves upon YOLOv5 by using a new loss function called GIoU, which helps to improve the model's accuracy. YOLOv7 (2022): Introduces a new network architecture called Panoptic YOLO, which can detect both objects and semantic segments in an image.

YOLOv8 (2023): The latest version of YOLO, which introduces a number of new features, including a new network architecture called BiFPN, a new loss function called CIoU, and a new training method called Cross-GPU training.

B. YOLO ALGORITHM AND ARCHITECTURE

The YOLO architecture consists of three main components: (i) the backbone; (ii) the neck; (iii) the head [23], [48], [50], [54]. The architecture of the backbone, neck, and head can vary between different versions of YOLO, and improvements in these components have led to significant improvements in the overall accuracy and speed of the YOLO network [25], [55]. The choice of backbone, neck, and head can affect the speed and accuracy of the YOLO model and depends on the specific application and the desired trade-off [49]. More recent versions of YOLO have introduced improvements in all three components to achieve better performance. Figure 2 represents YOLO Architecture.

- **Backbone:** The primary role of the backbone is to extract valuable characteristics from the initial image. Usually, a convolutional neural network is employed as the backbone, which has been trained on extensive datasets like ImageNet. The backbone functions as the network responsible for extracting features and generating feature maps from the input images. In YOLO, some commonly utilized backbones include VGG16, ResNet50, CSPDarknet53, and EfficientRep.
- **Neck:** The neck serves as the intermediary between the backbone and the head in the architecture. It consists of two main components, namely the spatial pyramid pooling (SPP) module and the path aggregation network (PAN). The neck's function is to combine the feature maps from various layers of the backbone network and forward them to the head. In YOLO, popular neck options include Spatial Pyramid Pooling (SPP), Feature Pyramid Network (FPN), NAS-FPN, and Rep-PAN.
- **Head:** The head component is responsible for handling the combined features and making predictions regarding bounding boxes, objectness scores, and classification scores. In YOLO, a one-stage object detection approach, like YOLOv3, is employed as the detection head. The head's primary role is to generate the ultimate output of the network, which includes predicted bounding boxes and class labels. YOLO utilizes various popular heads, such as Efficient decoupled, Multi-scale, and Anchor-based detection heads.

The YOLOv1 architecture, Figure 3, is inspired by GoogLeNet and it replaces inception modules with 1×1 and 3×3 convolutional layers. This architecture utilizes two fully connected layers on the convoluted feature map, outputting a final prediction grid of size S x S x (5B + K) [48].

This initial YOLO model prioritized speed by utilizing a single CNN [58] to directly predict object locations and classes in real time. However, this approach led to decreased accuracy, particularly for small objects or those with overlapping bounding boxes [59]. The detection architecture of the original YOLO model performed a single pass on the image to predict object locations and class labels [60]. Subsequent versions of YOLO introduced improvements to address these limitations.

YOLOv2, for instance, introduced batch normalization, anchor boxes, and passthrough layers to enhance object localization. Additionally, it incorporated multiscale training and achieved a processing speed of 40-90 frames per Second (FPS) [50]. These refinements aimed to enhance both accuracy and speed in object detection tasks. The performance of YOLOv3 was improved by incorporating a multi-scale feature extraction architecture. New backbone network, feature pyramid network, and more anchor boxes. It allowed a tradeoff between speed and accuracy [25].

YOLOv4 and YOLOv5 introduced new network backbones, improved data augmentation, and optimized training strategies. This enhanced accuracy without severely impacting real time performance [61], [62]. YOLO framework for object detection consistently evolved to balance speed and accuracy in detection tasks. PP-YOLO [63], developed by Alibaba Group, provided further improvements with a new backbone network, spatial attention module, and path aggregation network, making it faster and more accurate than YOLOv5 [64].

YOLOv6 implemented a variant of the EfficientNet architecture named EfficientNet-L2, surpassing the EfficientDet architecture used in YOLOv5 in terms of efficiency. It also introduced a design called EfficientRep Backbone and Rep-PAN Neck, leading to faster and more accurate results. The head of the network was decoupled, adding layers, and

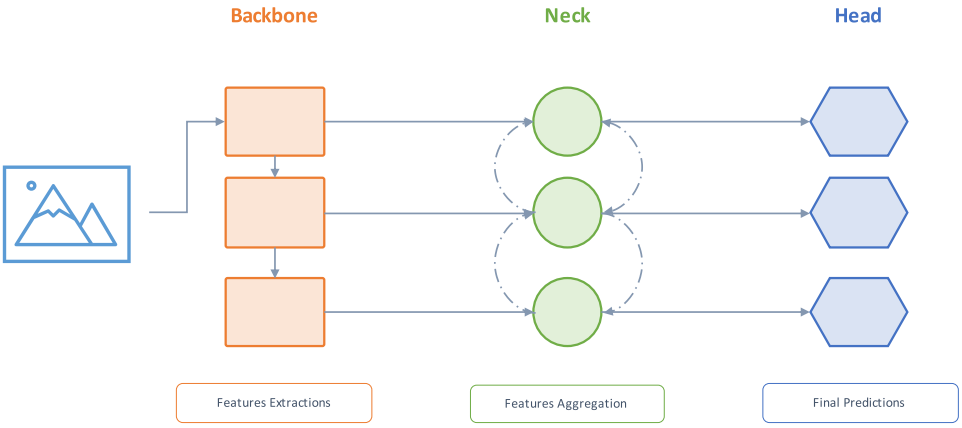


FIGURE 2. The architecture of YOLO consists of a backbone, neck, and head. The backbone, neck, and head vary in different versions of YOLO. For backbone, normally Darknet, VGG16, or Resnet are used; for neck feature pyramid network (FPN) [56], and for neck Densenet [57] or sparsenet are used.

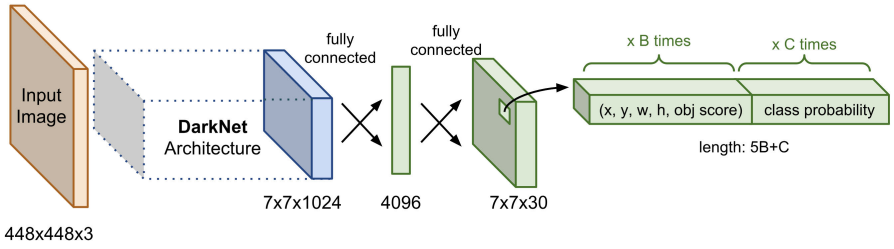


FIGURE 3. YOLOv1 network architecture [48].

separating these features from the final head, which improved performance [65].

YOLOv7 improved accuracy without raising inference costs, reducing parameters and computation by 40% and 50% respectively, compared to other leading real time object detectors [66]. It had a faster, stronger network architecture, more accurate detection performance, a more robust loss function, and enhanced label assignment and model training efficiency. It also required cheaper computing hardware and could be trained faster on small datasets without pre-trained weights [50], [67].

YOLOv8 [68], the most advanced model at the time of writing, had better feature aggregation and a mish activation function that improved detection accuracy and processing speed. It is an anchor-free model, predicting object centers directly without known anchor boxes. YOLO-NAS [69], created by Deci AI, outperformed its predecessors (especially YOLOv6 and YOLOv8) by achieving higher mAP values on the COCO dataset while maintaining lower latency. It also performed best on the Roboflow 100 dataset benchmark, indicating its ease of fine-tuning on custom datasets. Thus, the YOLO family of object detection models has consistently evolved to optimize both speed and accuracy, providing a variety of models to cater to diverse applications and hardware requirements. Table 1 summarizes the key features of each version of YOLO.

C. YOLO IN ACTION

YOLO revolutionized the process of object detection by simultaneously detecting all bounding boxes within an $S \times S$ region using grids. It predicts B bounding boxes for each class, accompanied by confidence scores for C different classes per grid element [48]. Each bounding box prediction comprises of five values: P_c, b_x, b_y, b_h, b_w. Here, P_c represents the confidence score, reflecting the model's confidence in the presence and accuracy of the object within the box. The coordinates b_x and b_y denote the box center relative to the grid cell, while b_h and b_w indicate the box height and width relative to the entire image. The output of YOLO is a tensor of size $S \times S \times (B \times 5 + C)$, which may undergo non-maximum suppression (NMS) to eliminate duplicate detections. These grid cells facilitate operations related to bounding box estimation and class probabilities [50]. Consequently, YOLO estimates the likelihood of the detection element's bounding box center being located within the grid cell, as formulated by Equation 2.

$$C(P) = Prob(p) \times IoU(prediction, target) \tag{1}$$

where: $C(P)$ is the confidence of prediction P, $Prob(p)$ is the probability of presence of object p, and $IoU(prediction, target)$ is the Intersection over Union between the predicted and target bounding boxes.

$$IoU = \left| \frac{B \cap B^{gt}}{B \cup B^{gt}} \right| \tag{2}$$

TABLE 1. Key features of each version of YOLO.

Version	Date	Anchor	Framework	Backbone	AP (%)
YOLO	2015	×	Darknet	Darknet24	63.4
YOLOv2	2016	✓	Darknet	Darknet24	63.4
YOLOv3	2018	✓	Darknet	Darknet53	36.2
YOLOv4	2020	✓	Darknet	CSPDarknet53	43.5
YOLOv5	2020	✓	Pytorch	Modified CSP v7	55.8
PP-YOLO	2020	✓	PaddlePaddle	ResNet50-vd	45.9
Scaled-YOLOv4	2021	✓	Pytorch	CSPDarknet	56.0
PP-YOLOv2	2021	✓	PaddlePaddle	ResNet101-vd	50.3
YOLOR	2021	✓	Pytorch	CSPDarknet	55.4
YOLOX	2021	×	Pytorch	Modified CSP v5	51.2
PP-YOLOE	2022	×	PaddlePaddle	CSPRepResNet	54.7
YOLOv6	2022	×	Pytorch	EfficientRep	52.5
YOLOv7	2022	×	Pytorch	RepConvN	56.8
DAMO-YOLO	2022	×	Pytorch	MAE-NAS	50.0
YOLOv8	2023	×	Pytorch	YOLOv8	53.9
YOLO-NAS	2023	×	Pytorch	YOLO-NAS	52.2

TABLE 2. YOLO included studies categorized in the Oncology domain.

Ref.	Application	YOLO v.	Dataset	Metrics
Cheng *et al.* 2021, [93]	Bone metastasis detection	YOLOv4	D32	Sensitivity 0.72 ± 0.04 Precision 0.90 ± 0.04
Li *et al.* 2021, [94]	Cell viability assay	YOLOv3	D31	Accuracy 0.94 Sensitivity 0.936 Specificity 0.944
Ku *et al.* 2022, [92]	Gastroscopy	YOLOv5	D24	Precision 0.98 Recall 0.89 mAP 0.902

In the context of the YOLO algorithm, the target box is denoted as B^{gt}, while the predicted box is represented as B. The probability (p) signifies whether the object exists within the detected bounding box. The IoU metric, defined by Equation 2, calculates the intersection area between the ground truth and predicted bounding boxes. It determines an acceptable area for each detected object in the input image and makes decisions based on it. To obtain the most suitable bounding box, the confidence value is applied after the estimation. The process of computation of the IoU can be illustrated as shown in Figure 4.

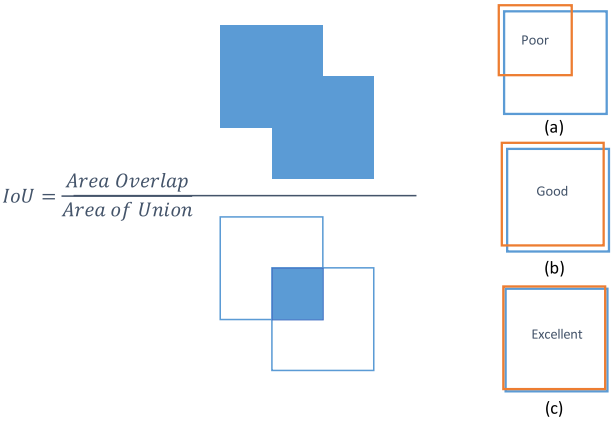


FIGURE 4. Computing the Intersection over Union: (a) poor detection performance, (b) good detection performance, (c) excellent detection performance.

D. IMAGE ANNOTATION

Image annotation [70] is a vital process in computer vision and machine learning. It is the process of labeling or marking specific objects or regions of interest within an image [71]. It involves adding metadata or annotations to images to provide additional information about the objects or features present in the image. The purpose of image annotation is to create a labeled dataset that serves as training data for learning algorithms, particularly for tasks like object detection, object recognition, and image segmentation [72]. By annotating images, human annotators or data scientists manually outline or mark the objects of interest within the image, often by drawing bounding boxes, polygons, or semantic segmentations around those objects. Accurate annotations are crucial for training models to accurately detect, recognize, and segment objects. They provide ground truth data, enable object localization, ensure model accuracy and performance, and facilitate diverse and domain-specific datasets. Annotations also aid in model evaluation and serve as a valuable resource for transfer learning. In summary, image annotation is a fundamental step that underpins

TABLE 3. YOLO included studies categorized in the Pathology domain.

Ref.	Application	YOLO v.	Dataset	Metrics
Larpant *et al.* 2022, [95]	Hematology (Blood transfusion)	YOLOv4-tiny	D1	Accuracy 0.966 Sensitivity 0.919 Specificity 0.9894 AUC 0.99
Han *et al.* 2023, [96]	Hematology (White blood cell detection)	MID-YOLO	D3	mAP@.5: 99.11% mAP@.75: 96.54 mAP@.5:.95: 81.13 Precision: 1.81% higher than original
Rong *et al.* 2023, [31]	Histology-based Nucleus Segmentation and Detection	YOLOv5	D16	Accuracy 0.7110 Precision 0.8308 Recall 0.6743 F1-score 0.7409 mIoU 0.8423
Quan *et al.* 2023, [97]	Human epidermal growth factor receptor 2 (HER2)	YOLOv5	D12	Precision 0.81
Zhu *et al.* 2023, [98]	Sperm detection	YOLOv5s	D36	Sensitivity 0.9362 Precision 0.6435 F1-score 0.7363
Sun *et al.* 2021, [99]	Neuroscience	YOLOv1	D41	Accuracy 0.75 AUC 0.743 Sensitivity 0.80 Specificity 0.72
Afshari *et al.* 2018, [81]	Tumor localization	YOLOv3	D17	Precision 0.75 to 0.98 Recall 0.94 to 0.1 IoU 0.72
Guo *et al.* 2022, [80]	Vitiligo Lesions	YOLOv3	D28	Sensitivity 0.962 CPM value of 0.905
Huang *et al.* 2023, [100]	Yeast cell detection	YOLOv5	D43	Accuracy 0.942

the development of reliable and effective computer vision systems in various industries and applications [41].

There are several popular tools available for annotating images. The choice of tool often depends on personal preference, project requirements, and the specific desired features. Commonly used tools include Visual Object Tagging Tool (VoTT) [73], VGG Image Annotator (VIA) [74], and Roboflow [75] which is a popular platform for managing, preprocessing, and annotating datasets for computer vision tasks. It provides a comprehensive end-to-end solution for dataset management, annotation, and preprocessing, offering a range of features that can help streamline your object detection workflow. When it comes to annotating datasets for YOLO, there are a few commonly used annotation formats that work well with YOLO-based object detection models. The most popular annotation format for YOLO datasets is the Darknet format, which is the native format used by the Darknet framework, the original implementation of YOLO.

$$\begin{bmatrix} c_i & x & y & w & h \\ c_i & x & y & w & h \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_i & x & y & w & h \\ c_i & x & y & w & h \end{bmatrix} \quad (3)$$

In the above matrix, each line represents a single object annotation where c represents the class or label of the object being annotated. It is usually represented by an integer index corresponding to the class label defined in the YOLO configurations. x and y represents the normalized x, y coordinate of the bounding box's center point. The value is relative to the width of the image, ranging from 0 to 1. w and h represent the normalized width and height of the bounding box. The value is relative to the width of the image, ranging from 0 to 1. Each annotation line corresponds to one annotated object in the image. Multiple lines can be present in the annotation file, each representing a different object.

E. HOW YOLO OPERATES

During the process of predicting bounding boxes, YOLO employs "dynamic anchor boxes" utilizing a clustering algorithm. This algorithm groups the ground truth bounding boxes into clusters and utilizes the centroids of these clusters as anchor boxes [76]. By doing so, the anchor boxes become better aligned with the size and shape of the detected objects. However, the primary source of error in YOLO arises from localization. This is due to the fact that the bounding box ratios are entirely learned from the data, causing YOLO to

TABLE 4. YOLO included studies categorized in the Radiology domain.

Ref.	Application	YOLO v.	Dataset	Metrics
Safdar *et al.* 2020, [32]	Brain tumor detection	YOLOv3	D27	Accuracy 0.60 to 0.96
Al-Antari *et al.* 2020, [17]	Breast cancer detection	YOLOv3	D10	Accuracy 0.991 and 0.972 F1-scores 0.992 and 0.980
Aly *et al.* 2021, [89]	Breast cancer detection	YOLOv3	D11	Accuracy 0.98 Precision 0.95 Recall 0.90 F1-score 0.925
Fu *et al.* 2022, [91]	Breast cancer detection	YOLOv3	D15	Accuracy 0.9067 Precision 0.9281
Baccouche *et al.* 2022, [19]	Breast cancer diagnosis	YOLOv1	D13	Accuracy 0.92
Al-Antari *et al.* 2020, [79]	Breast lesions diagnosis	YOLOv3	D22	Accuracy 0.9727 F1-score 0.9802
Su *et al.* 2022, [90]	Breast mass detection and segmentation	YOLOv5	D14	Accuracy 0.957 Precision 0.650 F1-score 0.745 IoU 0.640
Zhuang *et al.* 2020, [38]	Cardiovascular disease	YOLOv3	D8	Accuracy 0.952 Precision 0.964 Recall 0.940 F1-score 0.952
Chen *et al.* 2023, [101]	Cardiovascular disease	YOLOv3	D39	Accuracy 0.9481
Al-Masni *et al.* 2020, [102]	Cerebral microbleeds detection	YOLOv2	D20	Sensitivity 0.93.62 FP avg 52.18 and 155.50
Nambu *et al.* 2022, [103]	Cervical cytology of squamous cell atypia	YOLOv4	D30	Accuracy 0.905 F-measure 0.705
Boonrod *et al.* 2022, [104]	Cervical spine injury detection	YOLOv4	D21	Accuracy 0.75 Sensitivity 0.80 Specificity 0.72
Tang *et al.* 2022, [105]	Colon polyps detection	YOLOv4	D34	Accuracy 0.8428, 0.7517, and 0.8643 mAP 86.8422, 72.1863 and 77.6162
Pacal *et al.* 2022, [40]	Colorectal cancer (CRC)	YOLOv4	D4	Precision 82.33 to 88.62 Recall 71.01 to 77.55 F1-score 76.25 to 82.16 mAP 74.52 to 96.22
Matsui *et al.* 2021, [106]	Colorectal cancer (CRC)	YOLOv3	D33	IoU 0.9374 and 0.9077
Tang *et al.* 2023, [107]	Colorectal cancer (CRC)	YOLOv3	D35	AP 0.5460 and 0.7541 mAP 0.7007 IoU 0.5724
Ozturk *et al.* 2020, [108]	COVID-19 detections	YOLOv3	D2	Accuracy 0.9808 Sensitivity 0.8535 Specificity 0.9218 F1-score 0.8737
Kong *et al.* 2023, [109]	Dentistry	YOLOv4	D42	Accuracy 0.762 FPS 39.4
Panyarak *et al.* 2023, [110]	Dental caries in bitewing radiographs	YOLOv7 vs YOLOv3	D45	Precision (0.557 vs. 0.268) F1-score (0.555 vs. 0.375) mAP (0.562 vs. 0.458)
Tian *et al.* 2022, [111]	Dorsal hand veins detection	YOLO Nano-Vein	D9	AP 0.9323
Pang *et al.* 2019, [112]	Gallstone detection	YOLOv3	D5	Accuracy 0.927 mAP 0.943
Azam *et al.* 2022, [113]	Laryngeal cancer detection	YOLOv5s	D6	Precision 0.92 Recall 0.90 F1-score 0.91

TABLE 4. *(Continued.)* YOLO included studies categorized in the Radiology domain.

Tsai *et al.* 2021, [114]	Lumbar disc herniation detection	YOLOv3	D38	Accuracy 0.736 mAP 0.769
Huang *et al.* 2022, [87]	Lung nodule detection	YOLOv3	D47	Sensitivity 0.962 FPR 8 CPM 0.905
Zhang *et al.* 2023, [115]	Microaneurysm detection	YOLOv8	D44	Recall 0.8823 Precision 0.9798 F1-score 0.9285 AP 0.9462
Rouzrokh *et al.* 2021, [116]	Orthopedics	YOLOv3	D37	Accuracy 0.495 Sensitivity 0.890 Specificity 0.488
Ma *et al.* 2021, [117]	Pulmonary hypertension	YOLOv3	D26	mAP 0.8225 Recall 0.9412 Precision 0.9661
Ahmadyar *et al.* 2023, [86]	Pulmonary nodule detection	YOLOv5s	D46	Accuracy 0.984 AUC 0.989
Lee *et al.* 2023, [118]	Rotator cuff tear screening	YOLOv8	D48	Accuracy 0.96 Sensitivity 0.98 Precision 0.98 Specificity 0.91 F1-score 0.97 AUC 0.94
Fu *et al.* 2022, [119]	Spinal cerebrospinal fluid segmentation	YOLOv3	D29	Specificity 0.917 AUC 0.810
Lv *et al.* 2022, [120]	Traditional Chinese medicine recognition	YOLOv5	D25	Accuracy 0.9433 FPS 75
Till *et al.* 2023, [121]	Wrist fractures in children	YOLOv7	D40	mAP@0.5 +25.51% mAP@[0.5:0.95] +39.78%

TABLE 5. YOLO included studies categorized in the Surgical Procedures domain.

Ref.	Application	YOLO v.	Dataset	Metrics
Wang *et al.* 2022, [83]	Parathyroid surgery	YOLOv3, Faster R-CNN, and Cascade algorithms	D7	Precision 0.887 Recall 0.923 F1-score 0.905
Amiri Tehrani Zade *et al.* 2023, [84]	Ultrasound Interventions (needle placement)	YOLOv3	D23	Accuracy 0.98

struggle with atypical ratio bounding boxes [77].

$$\begin{aligned} b_x &= \sigma\left(t_x\right) + c_x \\ b_y &= \sigma\left(t_y\right) + c_y \\ b_w &= \left(p_w\right) * e^{tw} \\ b_h &= P_h * e^{th} \end{aligned} \tag{4}$$

In the YOLO framework, the bounding box coordinates are denoted as b_x, b_y, b_w, and b_h, while the center coordinates are represented by x, y, and the width and height are given by b_w and p_h respectively. Each bounding box has estimated coordinates t_x, t_y, t_w, and t_h. The values c_x and c_y correspond to the upper-left coordinates of the grid cell. YOLO defines a threshold value for the confidence score, where predictions below this threshold are discarded. Non-maximum suppression is then applied to generate the final positions for the detected bounding boxes. Finally, a loss function is computed for the detected bounding boxes in the last stage. Figure 5 provides a clear understanding of the work of YOLO.

F. ADVANTAGES AND DRAWBACKS OF YOLO FOR MEDICAL OBJECT DETECTION

YOLO is also highly generalized and can recognize a wide range of objects. However, it is important to be aware of the

TABLE 6. YOLO included studies categorized in the Personal Protective Equipment Detection domain.

Ref.	Application	YOLO v.	Dataset	Metrics
Han *et al.* 2022, [21]	Face mask detection	YOLOv4-tiny	D18	mAP 0.6701% Speed 92.81 FPS
Loey *et al.* 2021, [4]	Face mask detection	YOLOv2	D19	mAP 0.81

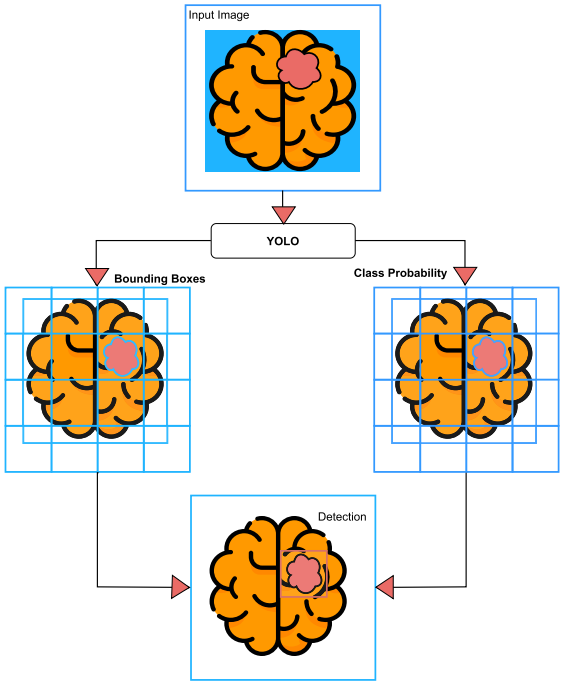


FIGURE 5. YOLO works by dividing the input image into a grid of cells. Each cell predicts a bounding box, a confidence score, and the class probabilities for the objects in that cell. The bounding box is a rectangle that is used to surround the object. A confidence score, between 0 and 1, indicates how confident the model is that the object is present in the cell. The class probabilities are the probabilities that the object in the cell belongs to each of the possible classes.

advantages and disadvantages of YOLO before using it for a specific application.

1) ADVANTAGES OF YOLO FOR MEDICAL OBJECT DETECTION

- **Real-Time Detection**: YOLO's inherent ability to detect objects in a single pass enables real time processing of medical images, which is crucial in time-sensitive medical scenarios such as surgeries or emergency diagnostics.
- **Efficiency**: YOLO's single-pass architecture is computationally efficient, allowing for faster processing speeds and reduced computational requirements, which is especially important for medical applications that demand rapid results.
- **Accuracy**: YOLO's holistic approach to object detection considers contextual information, leading to accurate localization and classification of medical objects within complex medical images, supporting precise diagnoses.
- **Multi-medical Object Detection**: YOLO excels in detecting multiple objects simultaneously, making it valuable for identifying various medical anomalies, tumors, lesions, pathologies factors, or abnormalities within a single image.
- **Adaptability**: The YOLO architecture is versatile and can be fine-tuned for specific medical domains, such as surgical procedures, personal protective detection equipment, and medical images.
- **Minimal False Positives**: YOLO's ability to incorporate global context reduces the likelihood of false positive detections, enhancing the reliability of medical diagnoses.

2) DRAWBACKS OF YOLO FOR MEDICAL OBJECT DETECTION

There are some drawbacks of YOLO such as:

- **Large Dataset Requirement**: One of the primary drawbacks of YOLO is its need for a substantial dataset of images for training. The collection of these images can be both time-consuming and expensive.
- **Sensitivity to Object Scale**: YOLO's performance can be significantly affected by the scale of objects in the input image, leading to potential false positives or negatives.
- **Difficulty with Small Objects**: YOLO often struggles to detect smaller objects as efficiently as it does larger ones. The system divides the image into a grid of cells, and small objects might not be large enough to occupy an entire cell, affecting detection accuracy.
- **Issues with Occluded Objects**: YOLO tends to falter when tasked with detecting objects obscured by others. As it predicts the bounding box of each object, an obscured or partially visible object might not have a well-defined bounding box, limiting detection performance.
- **Limited Object Diversity**: The system often struggles with a diverse set of object classes. As YOLO is trained on a finite dataset of images, it may find it challenging to generalize its detection capabilities to unfamiliar objects.

The rest of this research work is structured as follows. Section II provides the methodology for conducting this study. Section III highlights the results of the literature's applicable studies and responds to the previously stated

research questions using synthesized data from the included research. Finally, section IV concludes this paper.

II. METHODS

This paper seeks to assemble a comprehensive compilation of relevant studies focusing on YOLO in the medical domain, covering the period from 2018 to 2023. The aims are to explore YOLO's potential capabilities in medical applications, diagnosis, and treatment planning. This paper was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [78]. The rest of this section has been thoroughly organized into two subsections. The first subsection (II-A) highlights the evidence acquisition, which explains the aim (II-A1), search strategy (II-A2), and study selection criteria (II-A3). Meanwhile, evidence synthesis of this SLR is presented in the second subsection (II-B).

A. EVIDENCE ACQUISITION

Evidence acquisition in PRISMA involves systematically searching and selecting relevant studies from databases. Medical applications with YOLO capabilities can aid in automating the identification and extraction of relevant evidence from medical images. This integration streamlines the evidence acquisition process, improving efficiency and accuracy in SLRs.

1) AIM

The primary goal of this systematic review is to comprehensively analyze and evaluate the application of YOLO in medical imaging with a focus on assessing its accuracy, efficiency, and effectiveness in medical applications. Additionally, the review aims to identify medical areas where YOLO has been successfully utilized, examining the literature to determine its strengths and potential for significant advancements. Moreover, it provides practical implications for integrating YOLO into medical applications, considering how its utilization can enhance efficiency, accuracy, and automation in medical image and diagnosis. To this end, the following questions were established as our SLR research focus:

RQ1: What are the key medical domains in which YOLO has been applied for medical object detection?
RQ2: What are the performance metrics used to evaluate the effectiveness of YOLO in medical applications, and how does YOLO perform in terms of accuracy, precision, recall, and other relevant metrics?
RQ3: What are the specific tasks and objects of interest within the medical domain where YOLO has demonstrated strong performance?
RQ4: What are the limitations and challenges encountered when using YOLO in medical applications, such as imbalanced datasets, small sample sizes, and computational requirements?
RQ5: How does YOLO compare to other existing object detection algorithms in terms of performance, efficiency, and applicability to medical imaging?

2) SEARCH STRATEGY

A systematic search of the literature was conducted using the National Library of Medicine's PubMed database (https://pubmed.ncbi.nlm.nih.gov, accessed and last searched on 26 January 2024) to identify relevant papers published between 01/01/2018 and 31/12/2023). The search strategy employed a combination of keywords, specifically (YOLO AND ((medical application) OR (medical image))), and adhered to the PRISMA guidelines [78]. The inclusion criteria focused on original research articles, while review papers, abstracts, and reports from meetings were excluded. Each identified article underwent a thorough evaluation to determine its eligibility for inclusion in this SLR. Figure 6 illustrates the PRISMA flowchart conducted in this study.

3) STUDY SELECTION CRITERIA

The selection criteria were carefully designed to ensure the identification of research articles that best align with the objectives of the systematic review while excluding those that are not relevant. The following inclusion criteria were applied to identify the studies to be included in this review:

1) **Inclusion Criteria**:
   - ❒ Peer-reviewed articles published between 2018 and 2023 that focus on the application of YOLO in the medical domain.
   - ❒ Studies that report original research findings.
   - ❒ Studies that primarily focus on medical applications of YOLO.
   - ❒ Studies published in English language.
2) **Exclusion Criteria**:
   - ❒ Studies published before 1 January 2018.
   - ❒ Review papers, abstracts, and reports from meetings or conferences.
   - ❒ Studies that do not involve the YOLO algorithm or its application in the medical domain.
   - ❒ Studies with insufficient information or incomplete methodology.
   - ❒ Studies that primarily focus on non-medical applications of YOLO.
   - ❒ Studies published in languages other than English.

B. EVIDENCE SYNTHESIS

Initially, 124 articles were obtained from the database search, and 48 were included in this study. The quality assessment guidelines employed in this systematic review are designed to reduce bias, enhance transparency, and ensure repeatability. These guidelines focus on the YOLO medical applications utilizing medical datasets. The following criteria were considered during the assessment process: alignment of the selected studies with the primary objective of the systematic review, evaluation of the utilization of performance metrics, scrutiny of the soundness of the conclusions drawn in the

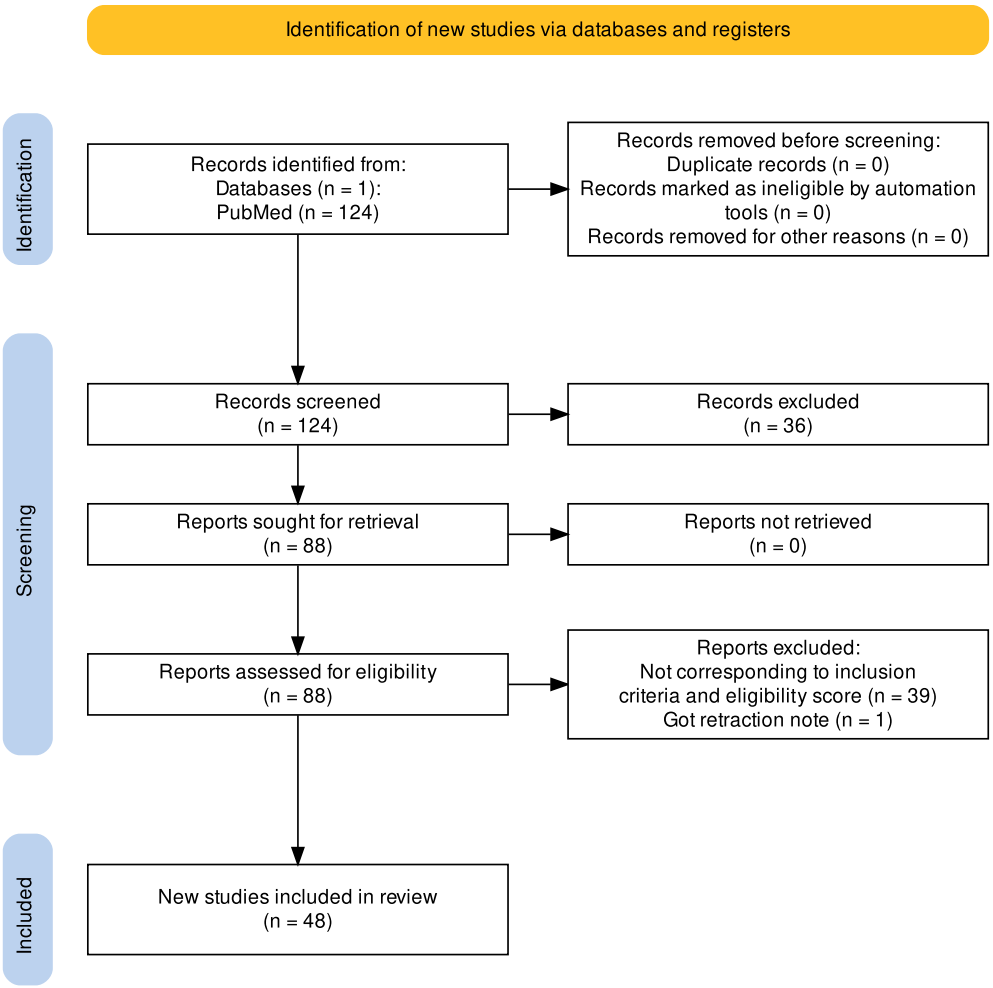


FIGURE 6. PRISMA flowchart.

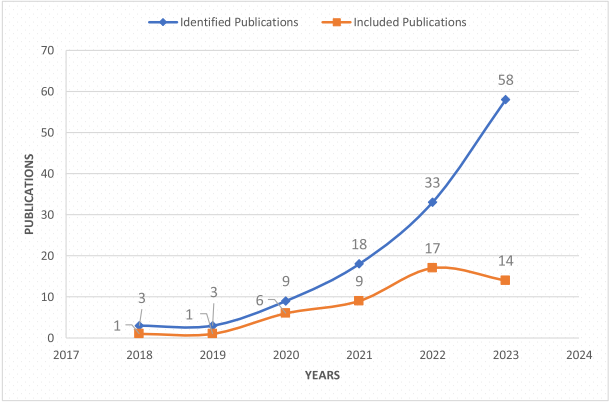


FIGURE 7. Trend of publications on the YOLO applications in medical imaging.

selected studies, and verification of the use of valid and reliable datasets.

Figure 7 displays the years and the corresponding counts of identified and included publications. A clear upward trend is observable in the graph of the identified publications, starting with a modest count of 3 identified publications in 2018. This number remains relatively stable in 2019, suggesting a period of early academic interest. The subsequent years illustrate a notable increase in research activity. The year 2020 sees the number double to 6 identified publications, indicating a growing recognition of YOLO's relevance in medical research. This momentum is maintained and amplified in the following years, with a steady rise to 18 papers in 2021 and a sharper ascent to 33 in 2022. The most striking leap occurs in 2023, where the count of identified publications surges to 58. This substantial growth over five years signifies a robust and accelerating interest in the application of YOLO within medical imaging. It is worth mentioning that, the observed decrease in the number of included studies in 2023 is attributable to the implementation of more strict inclusion and exclusion criteria, ensuring that only the most pertinent and high-quality research is selected for review.

Figure 8 illustrates the distribution of the main versions of YOLO detected from the 48 included studies. The pie chart indicates that YOLOv3 is the most frequently encountered

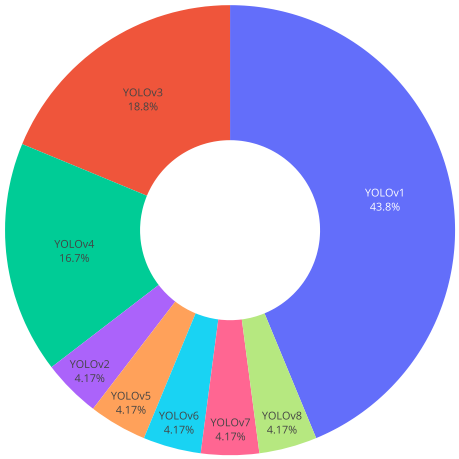


FIGURE 8. Distribution of YOLO versions in medical imaging.

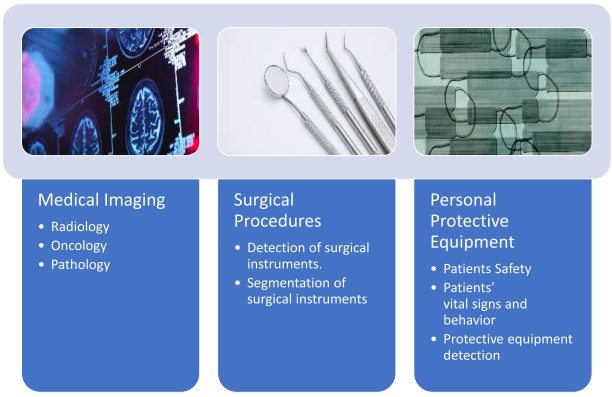


FIGURE 9. Three major domains of YOLO in healthcare; Medical Imaging, Surgical Procedures, and Personal Protective Equipment Detection.

version in the literature, with 21 instances, followed by YOLOv5 in 9 instances, and YOLOv4 in 8 instances. This could suggest that these versions have reached a level of maturity and performance that makes them popular among researchers and practitioners. YOLOv7 and YOLOv8 are represented as well with 2 instances each, indicating their emerging presence in the field and suggesting that more research papers on these versions could be yet to be published.

III. RESULTS

This section highlights the results of the literature's applicable studies and responds to the previously stated research questions (Sub-section II-A1) using synthesized data from the included research.

The key medical domains and applications where YOLO has been applied for medical object detection, along with their performance metrics used to evaluate the YOLO's effectiveness in these applications, with a focus on accuracy, precision, recall, and other relevant metrics, are categorized from the 48 included studies into three main domains as shown in Figure 9: I) Medical Imaging (III-A), II) Surgical Procedures (III-B), and Personal Protective Equipment Detection (III-C).

A. MEDICAL IMAGING

YOLO has exhibited strong performance in various tasks and objects of interest within medical imaging [5], including lesion detection [28], [79], anomaly detection [80], organ segmentation [31], [81], and instrument tracking [82], [83], [84]. These applications highlight YOLO's potential to enhance diagnostic accuracy, improve workflow efficiency, and ultimately advance patient care in the field of medical imaging. One of the primary applications of YOLO in medical imaging is the detection and localization of lesions [85]. Lesions can include tumors, nodules, masses, or any abnormality that may indicate disease [29]. YOLO has demonstrated strong performance in detecting lung nodules in chest X-rays, breast masses in mammograms, and brain tumors in MRI scans [17], [30], [86], [87]. Its ability to identify these objects of interest accurately and efficiently enables earlier detection, leading to timely intervention and improved patient outcomes. In Figure 10, different medical image applications clearly explain the advancement of YOLO in the field.

Moreover, YOLO has shown promise in detecting anomalies in medical images. Anomalies can encompass a wide range of abnormalities, such as fractures, hemorrhages, or foreign objects. YOLO's real time performance makes it well-suited for tasks such as identifying fractures in X-rays, detecting bleeds in brain CT scans, or identifying foreign bodies in radiographic images [87], [88]. By rapidly flagging anomalies, YOLO assists radiologists and clinicians in prioritizing critical cases and expediting appropriate treatment.

In addition, YOLO has been applied successfully in segmenting organs such as the heart, liver, kidney, and brain in different imaging modalities, including MRI and CT scans [32], [81]. Accurate identification and delineation of organs are crucial for surgical planning, radiation therapy, and monitoring disease progression [31]. Its high accuracy and efficiency streamline the segmentation process and contribute to more precise diagnoses and treatment plans. In surgical settings, YOLO has shown promise in detecting and tracking surgical instruments during procedures [83], [84]. By analyzing live video feeds, YOLO can identify instruments, surgical tools, and other objects of interest in real time, providing valuable assistance to surgeons and improving surgical safety. Its ability to handle fast-paced environments and track objects accurately makes it a valuable tool in computer-assisted interventions and robotic surgeries.

Also, YOLO has shown tremendous success in the detection and classification of breast masses in mammograms [19], [79], [89], [90], [91]. Its use in such applications has proven to significantly reduce the time, cost, and potential for human error inherent in traditional methods of mammogram evaluation. The YOLO system preprocesses the mammo-

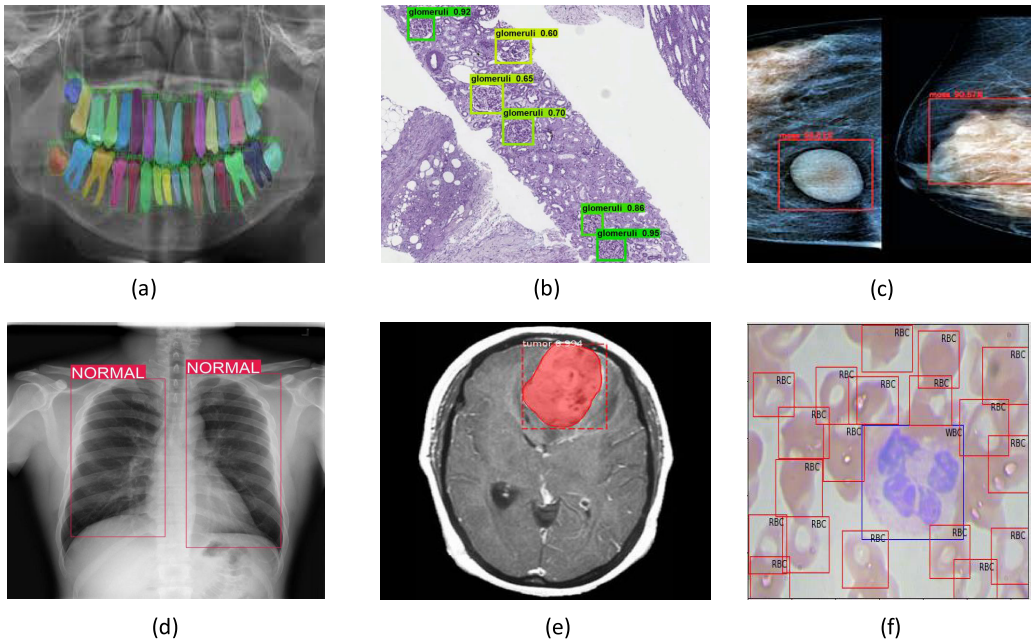


FIGURE 10. YOLO in different medical images applications: (a) periodontitis bone loss diagnosis, (b) glomerular detection, (c) breast cancer detection (d) lung normal and abnormal detection, (e) brain tumor detection, (f) white and red blood cells detections.

grams and then detects masses in them, distinguishing between malignant and benign lesions without any human intervention [80], [92].

The included studies in the medical imagining domain have been further classified into three sub-domains: Oncology (Table 2), Pathology (Table 3), and Radiology (Table 4).

However, while the applications of YOLO in healthcare have been fruitful, there are challenges, including the need for large, diverse, and high-quality datasets for model training. The algorithm's sensitivity to the scale of objects in images is another aspect that needs further improvement. Despite these challenges, with continuous research and refinement, YOLO's application in medical imaging holds significant promise for advancing healthcare diagnostics. One of the most promising areas of application for YOLO is medical imaging. YOLO has shown promising results in various fields such as radiology, oncology, and pathology [19]. For instance, in tumor detection, YOLO can identify and locate abnormal growths in medical images, assisting healthcare professionals in early disease diagnosis and treatment planning [30], [31], [32]. In the context of COVID-19, YOLO has demonstrated its value in detecting and quantifying infection patterns in lung CT scans, contributing to rapid and effective patient management [21].

As healthcare moves towards a more digitized environment, the volume of medical imaging data is growing. This data can be effectively analyzed using AI and machine learning tools like YOLO to extract valuable insights that can aid in diagnosis and treatment planning. For instance, YOLO has been used to detect and classify breast masses in mammograms [19], [79], [89], [90], [91]. The traditional evaluation process of screening mammograms is a laborious task, requiring significant time, cost, and human resources, and is prone to errors due to fatigue and the inherent subjectivity of human evaluation. However, with the introduction of YOLO into this process, an end-to-end computer-aided diagnosis system has been proposed and implemented [122]. The described system performs preprocessing on DICOM-format mammograms to convert them into images while preserving all the data. It is capable of detecting masses in full-field digital mammograms and can differentiate between malignant and benign lesions automatically, without requiring any human intervention, significantly reducing the potential for human error and streamlining the entire process.

B. SURGICAL PROCEDURES

YOLO could also be applied in surgical procedures. It offers great potential for surgical procedures, particularly in the context of computer-assisted and robotic surgery. The algorithm's ability to detect, classify, and locate objects in real time can be of significant value in the surgical environment [82], [83], [84]. For instance, YOLO can be utilized to identify and locate specific surgical instruments within the operating field, helping to streamline instrument tracking and potentially reducing surgical errors. Additionally, it could play a role in enhancing the safety and precision of robotic surgical systems by improving their ability to recognize and interact with various surgical elements in real time. Figure 11 demonstrates the usage of YOLO in surgical procedures.

The study by Wang et al. [83] developed an AI model based on YOLOv3 to identify parathyroid glands during

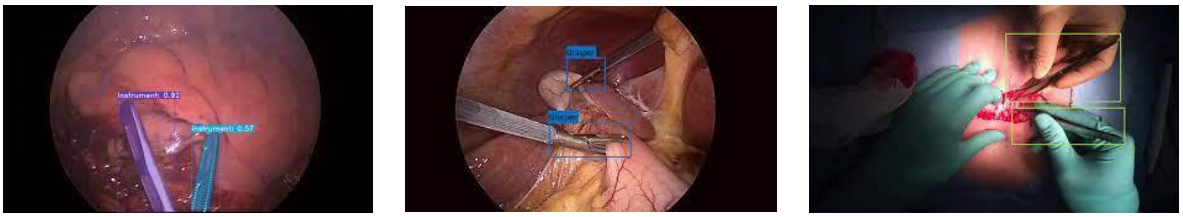


FIGURE 11. YOLO surgical procedures applications of (a)surgical tool detection in open surgery videos, (b) surgical instruments, (c) real time instance segmentation of surgical instruments.

endoscopic thyroid surgery. Using 1,700 images from thyroidectomy videos, the model outperformed junior surgeons and was comparable to senior surgeons in identification rates. Amiri Tehrani Zade et al. [84] developed a CNN-based method to enhance needle tracking in ultrasound for medical procedures. Using advanced motion estimation and the YOLOv3 framework, it accurately locates needles in real time ultrasound, outperforming current methods, and promising better ultrasound-guided interventions. Table 5 shows YOLO included studies categorized in the surgical procedures domain.

Surgeons often rely on various imaging technologies, such as MRI or CT scans, to guide their procedures. However, interpreting these images and applying their insights to the surgical process can be challenging. With the use of YOLO, these images could be analyzed in real time, providing surgeons with immediate feedback and guidance during the procedure [82]. This could potentially lead to more precise surgeries, fewer complications, and better patient outcomes. Furthermore, YOLO could be integrated with surgical navigation systems to improve real time imaging, enabling surgeons to better visualize the surgical field and carry out complex procedures with higher precision. This is particularly relevant for minimally invasive surgeries where real time imaging plays a crucial role. Despite these potential applications, integrating YOLO into surgical procedures also brings challenges. These include ensuring the algorithm's robustness and reliability in a highly variable and complex surgical environment, and addressing concerns related to patient safety and data privacy. Nevertheless, with continued research and technological refinement, YOLO's application in surgical procedures promises to enhance surgical precision and patient outcomes [84].

C. PERSONAL PROTECTIVE EQUIPMENT DETECTION

YOLO has demonstrated significant potential in the field of patient monitoring as shown in Figure 12. Real-time patient monitoring is a critical component of healthcare, providing valuable insights into a patient's condition and enabling timely interventions. Continuous patient monitoring is crucial in many healthcare scenarios, from intensive care units to home-based care. The ability of YOLO to detect and recognize objects in real time can be adapted to monitor various aspects of patient care. Traditionally, this monitoring has been done through a combination of manual observations by healthcare professionals and the use of various medical devices. However, these traditional methods can be time-consuming, costly, and subject to human error.

For instance, YOLO could be deployed to monitor patient activity and movements in an in-patient setting, identifying potential falls or other hazardous events before they occur [21]. This real time alert system could greatly enhance patient safety and improve healthcare outcomes. Similarly, for patients with chronic conditions, YOLO could potentially be utilized to monitor medication intake or adherence to certain therapeutic exercises, promoting better disease management. YOLO could provide valuable support, identifying significant health events from recorded or live video feeds. Despite the promising prospects, integrating YOLO into patient monitoring systems presents challenges, including ensuring patient privacy and dealing with diverse and complex real-world scenarios. However, with ongoing research and development, YOLO's application in patient monitoring can revolutionize care delivery, enhancing patient safety and health outcomes.

Moreover, YOLO has been applied in face mask detection. Han et al [21] developed an enhanced, lightweight YOLOv4-tiny-based detector for real time mask status detection, offering improved precision and speed with fewer parameters, suitable for public health applications. Loey et al. [4] proposed a deep learning model combining ResNet-50 and YOLOv2 to detect medical face masks in images, achieving 81% precision and outperforming related models in accuracy. Table 6 shows YOLO-included studies categorized in the personal protective equipment detection domain.

With the applications of YOLO, it is possible to build automated patient monitoring systems that can continuously monitor patients' vital signs and behavior and alert healthcare professionals to any abnormal patterns or signs of distress. Such systems could potentially lead to faster response times in emergency situations, more effective use of healthcare resources, and better overall patient outcomes.

Finally, Table 7 shows the data sources of all the 48 included studies. From those studies reviewed, no common datasets were identified across the medical applications, except for a few instances:

- In lung nodule detection: the Lung Nodule Analysis 2016 (LUNA16) dataset was used in 2 papers [86] and [87].

FIGURE 12. Medical personal protective equipment categories: (a) suit, (b) face shield;(c) goggles, (d) mask, and (e) glove.

- In breast cancer detection: three public benchmark datasets were utilized.
  - INbreast was used in 4 papers [17], [79], [89], and [90].
  - Digital Database for Screening Mammography (DDSM) was used in 2 papers [17] and [91].
  - An enhanced version of DDSM, the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) was used once by [90].
- In brain tumor detection: the Tumor Cancer Imaging Archive (TCIA) dataset was utilized by [32] and [81].

On the other hand, the majority of the included studies have used different datasets, with many not publicly accessible due to patient privacy concerns. Legal and ethical guidelines require tight data-sharing controls to protect patient confidentiality, highlighting the challenge in medical research to balance scientific progress with data privacy and protection norms.

IV. SUMMARY

This section summarizes this SLR into four sub-sections: the limitations of YOLO in healthcare (IV-A), future directions (IV-B), ethical considerations (IV-C), and the final conclusion (IV-D).

A. LIMITATIONS OF YOLO IN HEALTHCARE

While YOLO has made notable strides in object detection, it has inherent limitations [123]. The system leverages a one-stage algorithm that directly predicts object bounding boxes and class probabilities from images, significantly improving detection speed. YOLO backbones network structure excludes pooling and fully connected layers, instead accomplishing image convolutional transformations by modifying the step size of the convolutional core [124]. While this technique augments the network's depth and ability to extract features, it simultaneously escalates the model's complexity and the computational resources needed.

Though promising, the adoption of YOLO in healthcare brings with it an array of challenges and limitations that must be acknowledged. However, a prominent drawback is its comparative need for more accuracy in detecting small targets, a shortfall with potential ramifications in areas such as pill recognition or the identification of tiny lesions in medical imaging where the identification of minuscule objects is crucial [125]. It utilizes a deep network structure for feature extraction, enhancing accuracy at the cost of considerable computational power. This requirement can be a limiting factor in healthcare settings where resources are constrained [126], [127].

Despite these limitations, YOLO is a robust object detection algorithm used in various applications. As the algorithm develops, its accuracy and performance will likely improve Chaudhary et al. 2023, [128]. The best alternative to YOLO will depend on the specific application. If speed is essential, then YOLO may be the best choice [23]. Faster R-CNN or RetinaNet may be the best choice if accuracy is critical. It is also essential to consider the size of the objects that need to be detected. YOLO is not as good at detecting small objects as more significant objects [50]. If small objects must be detected, Faster R-CNN or RetinaNet may be better choices. Finally, it is essential to consider the diversity of the objects that need to be detected [129].

B. FUTURE DIRECTIONS

Developing YOLO variants specifically designed for healthcare applications is another promising research direction. Such customized systems could cater to the unique demands of healthcare, resulting in more effective and efficient tools. Integrating YOLO with other AI techniques, such as reinforcement learning and transfer learning, could significantly

TABLE 7. Data Sources of the selected studies.

Key	Application	Description	Source
D1 [95]	Blood transfusion	4692 blood samples tested for five Rh antigens (D, C, E, c, e) collected from blood banks in Thailand.	Not open access.
D2 [108]	COVID-19 detections	2,000 chest X-rays, half COVID-19 positive and half negative, sourced globally from multiple medical facilities.	Chest X-rays
D3 [96]	White blood cell detection	Raabin-WBC: 40,000 annotated microscopic white blood cell images, divided into 32,000 for training and 8,000 for testing.	Raabin-WBC
D4 [40]	Colorectal cancer	The SUN polyp database with 56,668 images, and the PICCOLO database with 10,000 images.	SUN, PICCOLO
D5 [112]	Gallstone detection	223,846 CT images of gallstones from 1,369 patients, collected from The Third Hospital in Shandong Province, China.	Not open access.
D6 [113]	Laryngeal cancer detection	624 laryngeal cancer video frames from 219 patients, gathered via video and rigid endoscopies.	Not open access.
D7 [83]	Parathyroid surgery	1,700 images of parathyroid glands extracted from 166 endoscopic thyroidectomy video recordings. These images were obtained from the First Affiliated Hospital of China Medical University.	Not open access.
D8 [38]	Cardiology	Ultrasonic apical long-axis view images of the heart from different patients, obtained from Shantou Institute of Ultrasonic Instruments.	Not open access.
D9 [111]	Dorsal hand veins	160 vein images of left and right dorsal hands from 80 subjects (47 males, 33 females, ages 10-45).	Not open access.
D10 [17]	Breast cancer	The DDSM dataset with over 26,000 mammograms and the INbreast dataset with 410 mammograms.	DDSM, INbreast
D11 [89]	Breast mass detection	INbreast: 410 mammograms from 115 cases: 360 from 90 bilateral breast cases (4 per case) and 50 from 25 mastectomy cases (2 per case), including normal, benign, and malignant cases.	INbreast
D12 [97]	Breast cancer	Pre-operative breast ultrasound videos and clinical parameters from 807 breast cancer patients who visited the Shanghai Cancer Center at Fudan University between February 2019 and July 2020.	Not open access.
D13 [19]	Breast cancer	A digital mammographic database with 413 cases, classified into four categories: Mass, Calcification, Architectural Distortions, and Normal.	Not open access.
D14 [90]	Breast cancer	I) INbreast: 410 mammograms from 115 cases. II) CBIS-DDSM: 10,240 mammograms which is a subset of the DDSM data selected and curated by a trained mammographer.	INbreast, CBIS-DDSM
D15 [91]	Breast cancer detection	I) Breast Ultrasound Image (BUSI): 1,577 ultrasound images of breast lesions. II) DDSM database contains 2,620 breast cases across 43 volumes, with four mammograms per case.	BUSI, DDSM
D16 [31]	Pathology	Different datasets for lung, liver, and breast cancer tissue (National Lung Screening Trial data set, NuCLS, and Cancer Genome Atlas)	Link
D17 [81]	Tumor Localization	PET scans of 156 patients from The Tumor Cancer Imaging Archive (TCIA) which contains PET scans of the whole body.	TCIA
D18 [21]	Face mask detection	I) MAFA: 30,811 Internet images and 35,806 annotated masked faces. II) WIDER FACE Mask: 30,000 images of people wearing face masks in a variety of settings, and collected from public places.	MAFA, WIDER FACE
D19 [4]	Face mask detection	A combined dataset from MMD (682 images) and FMD (853 images) for different people wearing face masks.	MMD, FMD
D20 [102]	Cerebral microbleeds detection	The study used two MR image datasets from Gachon University Gil Medical Center: a high-resolution set with 72 subjects (188 CMBs) and a low-resolution set with 107 subjects (572 CMBs), both annotated with CMB locations.	Link.
D21 [104]	Cervical spine injury detection	Lateral neck or cervical spine X-rays from 625 patients with CT scans, annotated for cervical spine injuries based on CT reports.	Not open access.
D22 [79]	Breast lesions	The INbreast database has 410 mammograms from 115 cases: 360 from 90 bilateral breast cases (4 per case) and 50 from 25 mastectomy cases (2 per case), including normal, benign, and malignant cases.	INbreast
D23 [84]	Ultrasound Interventions	Obtained from an injection needle inserted into two ultrasound phantoms: the Ultrasound Compatible Lumbar Epidural Simulator and the Femoral Vascular Access Ezono phantom, consisting of 14 ultrasound sequences, each with 100 frames.	Not open access.
D24 [92]	Gastroscopy	Microsoft COCO: 31,117 images from 3,747 patients, all digitized using the same endoscope and annotated with lesion type and location.	COCO
D25 [120]	Traditional Chinese Medicine	TCM: manually collected dataset of 100 common TCM images from pharmacies, all digitized using the same camera and annotated with the location and type of TCM.	Not open access.
D26 [117]	Pulmonary hypertension	MRI images of PH patients were digitized with the same scanner and annotated for the right ventricle location. All images are in DICOM format.	Not open access.
D27 [32]	Brain tumor detection	1,961 MRI brain scan images of low-grade glioma patients, all digitized with the same scanner and annotated with tumor locations and obtained from TCIA	TCIA
D28 [80]	Vitiligo lesions	The research utilized data from two sources: a dataset of vitiligo lesions from Chinese patients with Fitzpatrick skin types III or IV, and a test set of 145 images of vitiligo lesions from patients with Fitzpatrick skin types I, II, or V.	Not open access.
D29 [119]	Spontaneous intracranial hypotension	25,603 MRM image slices, acquired from various scanners and protocols, these images varied in quality but consistently depicted spinal CSF. Expert manual labeling identified the spinal CSF in each slice.	Not open access.
D30 [103]	Cervical cytology	Images were obtained through liquid-based cytology (LBC), a less invasive cervical cell collection method than traditional Pap smears. Two expert cytopathologists annotated the images.	Not open access.
D31 [94]	Oncology (Cell viability assay)	The HT29 Cell Dataset features images of various sizes and resolutions, showcasing diverse light scattering patterns from live, dead, and dying cells. It includes images of 575 live HT29 cells and 575 dead HT29 cells.	Not open access.
D32 [93]	Oncology (Bone metastasis detection)	Chest bone scintigraphy images from 205 prostate cancer patients and 371 breast cancer patients, were acquired using a SPECT scanner. Radiologists labeled the images to indicate the presence or absence of bone metastasis.	Not open access.
D33 [106]	Colorectal cancer	Extracted from video records from 20 colonoscopy procedures, with images extracted at five-second intervals focusing on colorectal lesions. Radiologists identified and labeled these lesions.	Not open access.
D34 [105]	Colonoscopy	The images were acquired using a narrow-band imaging system. Radiologists labeled the images as either being a polyp or not a polyp.	Not open access.
D35 [107]	Colorectal cancer	The dataset features 1,000 polyp images in three categories: SSA (sessile serrated adenoma), TA (traditional adenoma), and HP (hyperplastic polyp), captured using diverse colonoscopes and techniques across multiple clinical settings.	Not open access.
D36 [98]	Sperm detection	Microscopic images of sperm from 10 men, at 1280 x 960 resolution, averaging 274.6 cells/image. It is divided into training (480 images), validation (32,768 images), and test (120 images) sets.	Not open access.
D37 [116]	Orthopedics	1490 radiographs of dislocated total hip arthroplasties (THAs) and 91,094 radiographs of non-dislocated THAs.	Not open access.
D38 [114]	Lumbar Disc Herniation Detection	The paper uses a dataset of 550 lumbar vertebrae MRI images for training and evaluation. The images were collected from 20 to 65-year-old patients with at least 6 months of work experience. The images were labeled by radiologists to indicate the presence or absence of LDH	Not open access.
D39 [101]	Cardiovascular disease	Data from 265 patients, including 116 high-risk and 149 low-risk for plaques, were collected from Fujian Medical University's Second Affiliated Hospital.	Not open access.
D40 [121]	Wrist fractures in children	Annotated pediatric trauma wrist radiographs from 6,091 patients, totaling 10,643 studies with 20,327 images from GRAZPEDWRI-DX Medical University of Graz.	Link
D41 [99]	Neuroscience	Standard webcams were used for capturing test videos later resized to 360x640x3 pixels for analysis, and involved six adult C57BL/6 mice.	DeepBhvTracking
D42 [109]	Dentistry	Periodontitis: Large-scale, comprehensive, and high-definition panoramic radiograph dataset for identifying periodontitis.	Periodontitis
D43 [100]	Yeast cell detection	Self-prepared dataset of 1000 raw microscopic images (500×500 pixels) of yeast cells.	Not open access.
D44 [115]	Microaneurysm detection	1240 FFA images (768×768 pixels), from two sources: 1200 images from Nanjing Medical University and 70 images from the Persian Eye Clinic at the Isfahan University of Medical Sciences, all annotated by clinical doctors.	Not open access.
D45 [110]	Dental caries in bitewing radiographs	A set of 2575 bitewing radiographs were labeled using seven categories as per the ICCMS™ radiographic scoring system.	Not open access.

TABLE 7. *(Continued.)* **Data Sources of the selected studies.**

D46 [86]	Pulmonary nodule detection	LUNA16 dataset, comprising 888 CT scans, highlighting 1186 nodules and 400,000 non-nodular regions	LUNA16
D47 [87]	Lung nodule detection	LUNA16 dataset, comprising 888 CT scans, highlighting 1186 nodules and 400,000 non-nodular regions	LUNA16
D48 [118]	Rotator cuff tear screening	A combined 794 shoulder MRI scans were used, involving 374 men and 420 women with an average age of 59 $\pm$ 11 years.	Not open access.

enhance its performance. Reinforcement learning could enable the system to learn from its errors, thereby continually improving, while transfer learning would allow the application of knowledge acquired from one task to related tasks, potentially boosting accuracy and efficiency [130]. Moreover, it will be interesting to see the potential applications of large language models [131] in YOLO or object detection tasks.

Given YOLO's success in medical imaging, future studies are likely to concentrate on enhancing its accuracy for small target detection and extending its application to other healthcare areas, such as patient activity monitoring, real time anomaly detection during surgical procedures, or disease progression prediction based on image data. Moreover, there is a growing interest in integrating YOLO more effectively into clinical workflows. This could involve developing interfaces for seamless interaction between healthcare professionals and the system or devising protocols to ensure appropriate communication and utilization of the system's outputs in clinical decision-making. This rapidly evolving field will continue to reveal novel applications, benefits, and limitations of this technology.

1) DEVELOPING NEW DATASETS FOR MEDICAL OBJECT DETECTION

Developing new datasets for medical object detection using YOLO models can be challenging due to several reasons:

- **Data Privacy and Ethics:** Medical data is sensitive and protected by strict privacy regulations.
- **Annotating Medical Images:** Medical images often require precise and detailed annotations. Expert knowledge is needed to accurately label abnormalities, making annotation time-consuming and labor-intensive. The cost of annotating a large dataset can be significant.
- **Limited Data Availability:** Unlike general object detection, medical datasets are smaller due to the limited availability of medical images, especially for rare conditions. This scarcity can affect the model's performance and generalization.
- **Class Imbalance:** Medical conditions are often rare, leading to a class imbalance where certain classes have very few instances. This can lead to biased models that perform poorly on underrepresented classes.
- **Complexity and Variability:** Medical images can exhibit variations due to factors like lighting, equipment, patient demographics, and disease progression. Capturing this variability in the dataset is crucial for robust model performance.
- **Clinical Relevance:** The dataset needs to be clinically relevant and accurately represent the challenges that medical professionals face in real-world scenarios.

Addressing these challenges requires collaboration between medical professionals, data annotators, and machine learning experts. Rigorous quality control, careful dataset curation, and domain-specific adaptations of YOLO models are essential for successful medical object detection.

2) TRANSFER LEARNING FOR MEDICAL OBJECT DETECTION

Transfer learning is a valuable technique that can significantly benefit the application of YOLO in the medical imaging domain. Here is how transfer learning can be leveraged to improve YOLO's performance:

- **Pre-Trained Models:** Begin by training YOLO on a large dataset from a related domain, such as natural images. This pre-training imparts general object recognition capabilities to YOLO, capturing low-level features that can be valuable for medical object detection.
- **Fine-Tuning:** After pre-training, fine-tune the YOLO model using a smaller but domain-specific medical image dataset. This step adapts the model's learned features to the specific characteristics of medical images, enhancing its ability to detect medical objects.
- **Transfer of Knowledge:** Transfer learning facilitates the transfer of knowledge from the pre-trained model to the medical domain. This approach jumpstarts the training process and reduces the amount of labeled medical data required, a critical advantage in medical imaging where labeled data is often limited.
- **Improved Convergence:** Transfer learning allows the YOLO model to converge faster during fine-tuning, leading to quicker deployment and reducing the risk of overfitting, especially when working with smaller medical datasets.
- **Enhanced Feature Extraction:** The pre-trained features capture valuable information about edges, textures, and basic shapes. These features can be particularly beneficial in medical image analysis, aiding in the detection of various anomalies.

C. ETHICAL CONSIDERATIONS

As YOLO and similar technologies become more prevalent in healthcare, it is important to consider the ethical implications. Implementing YOLO in healthcare elicits various ethical and legal quandaries. Issues such as data privacy, informed consent, and the potential for bias in AI algorithms will need to be addressed. Future research will need to not only focus on improving the technical aspects of these systems but also

ensure that they are used in a way that respects patient rights and upholds the principles of medical ethics.

D. CONCLUSION

To conclude, this SLR offers a comprehensive analysis of the utilization of YOLO in various medical applications, encompassing tumor detection, blood transfusion medicine, COVID-19, colorectal cancer, radiology, laryngeal cancer, parathyroid surgery, and dorsal hand veins recognition, among others. The review incorporated a significant body of literature, aggregating insights from 124 papers published between 2018 and 2023. The findings reveal the pivotal role YOLO plays in enhancing the efficiency and accuracy of medical diagnoses and procedures.

The study also has a few limitations. In this study, we only focused on the Pubmed database, other databases may have relevant articles as well. However, in the medical domain, Pubmed is considered as a gold standard. Another limitation is that we considered object detection tasks only in medical images, personal protective equipment, and surgical procedures. Further medical instruments are not considered as medical objects.

By rapidly identifying and localizing ailments ranging from tumors to various cancers, YOLO has significantly improved patient outcomes while reducing diagnosis and treatment times. However, despite the remarkable successes of YOLO, its deployment is not without challenges. These include its sensitivity to object scale, difficulty in detecting small or occluded objects, and considerable computational resource requirements. To harness the full potential of YOLO, these issues need to be addressed by ongoing and future research.

Also, ethical considerations like data privacy and algorithmic bias need to be considered in the development of YOLO-based systems, particularly in healthcare. In summary, the integration of YOLO into healthcare applications represents a significant stride towards a future where AI not only enhances the accuracy and speed of medical processes but also democratizes access to quality healthcare. Nevertheless, continued research and development are essential for further improvements and for the optimal integration of YOLO into healthcare settings.

ACKNOWLEDGMENT

The authors would like to thank Ministry of Higher Education (MOHE), Malaysia for providing financial assistance under Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UTP/02/4) and Universiti Teknologi PETRONAS under the Yayasan Universiti Teknologi PETRONAS (YUTP-FRG/015LC0-308) for providing the required facilities to conduct this research work.

REFERENCES

[1] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," *NPJ Digit. Med.*, vol. 4, no. 1, p. 5, Jan. 2021.

[2] Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, and Z. Zeng, "CLU-CNNs: Object detection for medical images," *Neurocomputing*, vol. 350, pp. 53–59, Jul. 2019.

[3] J. Peng, Q. Chen, L. Kang, H. Jie, and Y. Han, "Autonomous recognition of multiple surgical instruments tips based on arrow OBB-YOLO network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.

[4] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustain. Cities Soc.*, vol. 65, Feb. 2021, Art. no. 102600.

[5] R. Yang and Y. Yu, "Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis," *Frontiers Oncol.*, vol. 11, Mar. 2021, Art. no. 638182.

[6] M. Tsuneki, "Deep learning models in medical image analysis," *J. Oral Biosci.*, vol. 64, no. 3, pp. 312–320, Sep. 2022.

[7] Y. Zhao, K. Zeng, Y. Zhao, P. Bhatia, M. Ranganath, M. L. Kozhikkavil, C. Li, and G. Hermosillo, "Deep learning solution for medical image localization and orientation detection," *Med. Image Anal.*, vol. 81, Oct. 2022, Art. no. 102529.

[8] R. Qureshi, M. Irfan, H. Ali, A. Khan, A. S. Nittala, S. Ali, A. Shah, T. M. Gondal, F. Sadak, Z. Shah, M. U. Hadi, S. Khan, Q. Al-Tashi, J. Wu, A. Bermak, and T. Alam, "Artificial intelligence and biosensors in healthcare and its clinical relevance: A review," *IEEE Access*, vol. 11, pp. 61600–61620, 2023.

[9] Q. Al-Tashi, M. B. Saad, A. Muneer, R. Qureshi, S. Mirjalili, A. Sheshadri, X. Le, N. I. Vokes, J. Zhang, and J. Wu, "Machine learning models for the identification of prognostic and predictive cancer biomarkers: A systematic review," *Int. J. Mol. Sci.*, vol. 24, no. 9, p. 7781, Apr. 2023.

[10] R. Qureshi, M. Irfan, T. M. Gondal, S. Khan, J. Wu, M. U. Hadi, J. Heymach, X. Le, H. Yan, and T. Alam, "AI in drug discovery and its clinical relevance," *Heliyon*, vol. 9, no. 7, Jul. 2023, Art. no. e17575.

[11] T. Panch, H. Mattie, and L. A. Celi, "The 'inconvenient truth' about AI in healthcare," *NPJ Digit. Med.*, vol. 2, no. 1, p. 77, 2019.

[12] R. Qureshi, B. Zou, T. Alam, J. Wu, V. H. F. Lee, and H. Yan, "Computational methods for the analysis and prediction of EGFR-mutated lung cancer drug resistance: Recent advances in drug design, challenges and future prospects," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 1, pp. 238–255, Jan. 2023.

[13] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[14] M. Nawaz, R. Qureshi, M. A. Teevno, and A. R. Shahid, "Object detection and segmentation by composition of fast fuzzy C-mean clustering based maps," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 6, pp. 7173–7188, Jun. 2023.

[15] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[16] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, and A. Veit, "OpenImages: A public dataset for large-scale multi-label and multi-class image classification," *Dataset*, vol. 2, no. 3, p. 18, 2017. [Online]. Available: https://github.com/openimages

[17] M. A. Al-antari, S.-M. Han, and T.-S. Kim, "Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms," *Comput. Methods Programs Biomed.*, vol. 196, 2020, Art. no. 105584, doi: 10.1016/j.cmpb.2020.105584.

[18] Y. E. Almalki, A. I. Din, M. Ramzan, M. Irfan, K. M. Aamir, A. Almalki, S. Alotaibi, G. Alaglan, H. A. Alshamrani, and S. Rahman, "Deep learning models for classification of dental diseases using orthopantomography X-ray OPG images," *Sensors*, vol. 22, no. 19, p. 7370, Sep. 2022, doi: 10.3390/s22197370.

[19] A. Baccouche, B. Garcia-Zapirain, Y. Zheng, and A. S. Elmaghraby, "Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques," *Comput. Methods Programs Biomed.*, vol. 221, Jun. 2022, Art. no. 106884, doi: 10.1016/j.cmpb.2022.106884.

[20] M. Dobrovolny, J. Benes, J. Langer, O. Krejcar, and A. Selamat, "Study on sperm-cell detection using YOLOv5 architecture with labaled dataset," *Genes*, vol. 14, no. 2, p. 451, Feb. 2023, doi: 10.3390/genes14020451.

[21] Z. Han, H. Huang, Q. Fan, Y. Li, Y. Li, and X. Chen, "SMD-YOLO: An efficient and lightweight detection method for mask wearing status during the COVID-19 pandemic," *Comput. Methods Programs Biomed.*, vol. 221, Jun. 2022, Art. no. 106888, doi: 10.1016/j.cmpb.2022.106888.

[22] Z. Huang, Y. Li, T. Zhao, P. Ying, Y. Fan, and J. Li, "Infusion port level detection for intravenous infusion based on YOLO v3 neural network," *Math. Biosciences Eng.*, vol. 18, no. 4, pp. 3491–3501, 2021, doi: 10.3934/mbe.2021175.

[23] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.

[24] 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.

[25] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of YOLO algorithm developments," *Proc. Comput. Sci.*, vol. 199, pp. 1066–1073, Jan. 2022.

[26] M. J. Mortada, S. Tomassini, H. Anbar, M. Morettini, L. Burattini, and A. Sbrollini, "Segmentation of anatomical structures of the left heart from echocardiographic images using deep learning," *Diagnostics*, vol. 13, no. 10, p. 1683, May 2023.

[27] P. Zeng, S. Liu, S. He, Q. Zheng, J. Wu, Y. Liu, G. Lyu, and P. Liu, "TUSPM-NET: A multi-task model for thyroid ultrasound standard plane recognition and detection of key anatomical structures of the thyroid," *Comput. Biol. Med.*, vol. 163, Sep. 2023, Art. no. 107069.

[28] A. Baccouche, B. Garcia-Zapirain, C. C. Olea, and A. S. Elmaghraby, "Breast lesions detection and classification via YOLO-based fusion models," *Comput., Mater. Continua*, vol. 69, no. 1, pp. 1407–1425, 2021.

[29] C. Santos, M. Aguiar, D. Welfer, and B. Belloni, "A new approach for detecting fundus lesions using image processing and deep neural network architecture based on YOLO model," *Sensors*, vol. 22, no. 17, p. 6441, Aug. 2022.

[30] F. J. P. Montalbo, "A computer-aided diagnosis of brain tumors using a fine-tuned YOLO-based model with transfer learning," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 12, pp. 4816–4834, 2020.

[31] R. Rong, H. Sheng, K. W. Jin, F. Wu, D. Luo, Z. Wen, C. Tang, D. M. Yang, L. Jia, M. Amgad, L. A. D. Cooper, Y. Xie, X. Zhan, S. Wang, and G. Xiao, "A deep learning approach for histology-based nucleus segmentation and tumor microenvironment characterization," *Modern Pathol.*, vol. 36, no. 8, Aug. 2023, Art. no. 100196, doi: 10.1016/j.modpat.2023.100196.

[32] M. Safdar, S. Kobaisi, and F. Zahra, "A comparative analysis of data augmentation approaches for magnetic resonance imaging (MRI) scan images of brain tumor," *Acta Inf. Medica*, vol. 28, no. 1, p. 29, 2020, doi: 10.5455/aim.2020.28.29-36.

[33] J. Zhou, B. Zhang, X. Yuan, C. Lian, L. Ji, Q. Zhang, and J. Yue, "YOLO-CIR: The network based on YOLO and ConvNeXt for infrared object detection," *Infr. Phys. Technol.*, vol. 131, Jun. 2023, Art. no. 104703.

[34] S. Bashir, R. Qureshi, A. Shah, X. Fan, and T. Alam, "YOLOv5-M: A deep neural network for medical object detection in real-time," in *Proc. IEEE Symp. Ind. Electron. Appl. (ISIEA)*, Jul. 2023, pp. 1–6, doi: 10.1109/ISIEA58478.2023.10212322.

[35] H. M. Ali and N. K. El Abbadi, "Optic disc localization in retinal fundus images based on you only look once network (YOLO)," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 2, pp. 1–11, 2023.

[36] J. Liang, Z. Wang, and X. Ye, "Application of deep learning in imaging diagnosis of brain diseases," in *Proc. 3rd Int. Conf. Mach. Learn., Big Data Bus. Intell. (MLBDBI)*, Dec. 2021, pp. 166–175.

[37] H. Honda, S. Mori, A. Watanabe, N. Sasagasako, S. Sadashima, T. Dong, K. Satoh, N. Nishida, and T. Iwaki, "Abnormal prion protein deposits with high seeding activities in the skeletal muscle, femoral nerve, and scalp of an autopsied case of sporadic Creutzfeldt–Jakob disease," *Neuropathology*, vol. 41, no. 2, pp. 152–158, Apr. 2021.

[38] Z. Zhuang, G. Liu, W. Ding, A. N. J. Raj, S. Qiu, J. Guo, and Y. Yuan, "Cardiac VFM visualization and analysis based on YOLO deep learning model and modified 2D continuity equation," *Comput. Med. Imag. Graph.*, vol. 82, 2020, Art. no. 101732, doi: 10.1016/j.compmedimag.2020.101732.

[39] K. K. Wong, M. Ayoub, Z. Cao, C. Chen, W. Chen, D. N. Ghista, and C. W. J. Zhang, "The synergy of cybernetical intelligence with medical image analysis for deep medicine: A methodological perspective," *Comput. Methods Programs Biomed.*, vol. 240, Oct. 2023, Art. no. 107677.

[40] I. Pacal, A. Karaman, D. Karaboga, B. Akay, A. Basturk, U. Nalbantoglu, and S. Coskun, "An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets," *Comput. Biol. Med.*, vol. 141, Feb. 2022, Art. no. 105031, doi: 10.1016/j.compbiomed.2021.105031.

[41] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, pp. 1–11, Dec. 2021.

[42] K. Li and L. Cao, "A review of object detection techniques," in *Proc. 5th Int. Conf. Electromechanical Control Technol. Transp. (ICECTT)*, May 2020, pp. 385–390.

[43] A. Kaur, Y. Singh, N. Neeru, L. Kaur, and A. Singh, "A survey on deep learning approaches to medical images and a systematic look up into real-time object detection," *Arch. Comput. Methods Eng.*, vol. 29, no. 4, pp. 2071–2111, Jun. 2022.

[44] N. Ganatra, "A comprehensive study of applying object detection methods for medical image analysis," in *Proc. 8th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2021, pp. 821–826.

[45] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[46] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.

[47] M. G. Ragab, S. J. Abdulkadir, and N. Aziz, "Random search one dimensional CNN for human activity recognition," *Int. Conf. Comput. Intell. (ICCI)*, pp. 86–91, Oct. 2020.

[48] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, p. 677, Jun. 2023.

[49] C. Chen, Z. Zheng, T. Xu, S. Guo, S. Feng, W. Yao, and Y. Lan, "YOLO-based UAV technology: A review of the research and its applications," *Drones*, vol. 7, no. 3, p. 190, Mar. 2023.

[50] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," 2023, *arXiv:2304.00501*.

[51] L. Aziz, M. S. B. Haji Salam, U. U. Sheikh, and S. Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review," *IEEE Access*, vol. 8, pp. 170461–170495, 2020.

[52] P. Bharati and A. Pramanik, "Deep learning techniques—R-CNN to mask R-CNN: A survey," in *Computational Intelligence in Pattern Recognition*. Singapore: Springer, 2020, pp. 657–668.

[53] C. Yu, Z. Hu, R. Li, X. Xia, Y. Zhao, X. Fan, and Y. Bai, "Segmentation and density statistics of mariculture cages from remote sensing images using mask R-CNN," *Inf. Process. Agricult.*, vol. 9, no. 3, pp. 417–430, Sep. 2022.

[54] Y. Cao, D. Pang, Y. Yan, Y. Jiang, and C. Tian, "A photovoltaic surface defect detection method for building based on deep learning," *J. Building Eng.*, vol. 70, Jul. 2023, Art. no. 106375.

[55] R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," *Digit. Signal Process.*, vol. 132, Jan. 2023, Art. no. 103812.

[56] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[57] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869*.

[58] M. G. Ragab, S. J. Abdulkadir, N. Aziz, Q. Al-Tashi, Y. Alyousifi, H. Alhussian, and A. Alqushaibi, "A novel one-dimensional CNN with exponential adaptive gradients for air pollution index prediction," *Sustainability*, vol. 12, no. 23, p. 10090, Dec. 2020.

[59] A. B. Amjoud and M. Amrouch, "Object detection using deep learning, CNNs and vision transformers: A review," *IEEE Access*, vol. 11, pp. 35479–35516, 2023.

[60] W. Chen, Y. Li, Z. Tian, and F. Zhang, "2D and 3D object detection algorithms from images: A survey," *Array*, vol. 19, Sep. 2023, Art. no. 100305.

[61] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digit. Signal Process.*, vol. 126, Jun. 2022, Art. no. 103514.

[62] X. Cong, S. Li, F. Chen, C. Liu, and Y. Meng, "A review of YOLO object detection algorithms based on deep learning," *Frontiers Comput. Intell. Syst.*, vol. 4, no. 2, pp. 17–20, Jun. 2023.
[63] A.-A. Tulbure, A.-A. Tulbure, and E.-H. Dulf, "A review on modern defect detection models using DCNNs—Deep convolutional neural networks," *J. Adv. Res.*, vol. 35, pp. 33–48, Jan. 2022.
[64] N.-N. Dao, T.-H. Do, S. Cho, and S. Dustdar, "Information revealed by vision: A review on the next-generation OCC standard for AIoV," *IT Prof.*, vol. 24, no. 4, pp. 58–65, Jul. 2022.
[65] S. Gupta and S. Nair, "A review of the emerging role of UAVs in construction site safety monitoring," *Mater. Today, Proc.*, 2023, doi: 10.1016/j.matpr.2023.03.135.
[66] Y. Wang, H. Wang, and Z. Xin, "Efficient detection model of steel strip surface defects based on YOLO-V7," *IEEE Access*, vol. 10, pp. 133936–133944, 2022.
[67] L. Cao, X. Zheng, and L. Fang, "The semantic segmentation of standing tree images based on the YOLO v7 deep learning algorithm," *Electronics*, vol. 12, no. 4, p. 929, Feb. 2023.
[68] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLOv8*. [Online]. Available: https://github.com/ultralytics/ultralytics
[69] S. Aharon, Louis-Dupont, O. Masad, K. Yurkova, L. Fridman, E. Khvedchenya, R. Rubin, N. Bagrov, B. Tymchenko, T. Keren, and A. Zhilko, "Super-gradients," Tech. Rep., 2021, doi: 10.5281/ZENODO.7789328.
[70] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognit.*, vol. 45, no. 1, pp. 346–362, Jan. 2012.
[71] C. Sager, C. Janiesch, and P. Zschech, "A survey of image labelling for computer vision applications," *J. Bus. Analytics*, vol. 4, no. 2, pp. 91–110, Jul. 2021.
[72] A. Kumar, A. Kalia, K. Verma, A. Sharma, and M. Kaushal, "Scaling up face masks detection with YOLO on a novel dataset," *Optik*, vol. 239, Aug. 2021, Art. no. 166744.
[73] S. Annadatha, M. Fridberg, S. Kold, O. Rahbek, and M. Shen, "A tool for thermal image annotation and automatic temperature extraction around orthopedic pin sites," in *Proc. IEEE 5th Int. Conf. Image Process. Appl. Syst. (IPAS)*, vol. Five, Dec. 2022, pp. 1–5.
[74] A. Dutta, A. Gupta, and A. Zissermann. (2316). *Vgg Image Annotator (Via)*. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/software/via
[75] F. Ciaglia, F. Saverio Zuppichini, P. Guerrie, M. McQuade, and J. Solawetz, "Roboflow 100: A rich, multi-domain object detection benchmark," 2022, *arXiv:2211.13523*.
[76] Y. Hu, X. Wu, G. Zheng, and X. Liu, "Object detection of UAV for anti-UAV based on improved YOLO v3," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 8386–8390.
[77] Y. Jamtsho, P. Riyamongkol, and R. Waranusast, "Real-time bhutanese license plate localization using YOLO," *ICT Exp.*, vol. 6, no. 2, pp. 121–124, Jun. 2020.
[78] M. J. Page, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Systematic Rev.*, vol. 10, no. 1, p. 89, Dec. 2021, doi: 10.1186/s13643-021-01626-4.
[79] M. A. Al-Antari, M. A. Al-Masni, and T. S. Kim, "Deep learning computer-aided diagnosis for breast lesion in digital mammogram," in *Deep Learning in Medical Image Analysis*, vol. 1213, 2020, pp. 59–72. 10.1007/978-3-030-33128-3_4.
[80] L. Guo, Y. Yang, H. Ding, H. Zheng, H. Yang, J. Xie, Y. Li, T. Lin, and Y. Ge, "A deep learning-based hybrid artificial intelligence model for the detection and severity assessment of vitiligo lesions," *Ann. Transl. Med.*, vol. 10, no. 10, p. 590, 2022, doi: 10.21037/atm-22-1738.
[81] S. Afshari, A. BenTaieb, and G. Hamarneh, "Automatic localization of normal active organs in 3D PET scans," *Computerized Med. Imag. Graph.*, vol. 70, pp. 111–118, Dec. 2018, doi: 10.1016/j.compmedimag.2018.09.008.
[82] Y. Huang, J. Li, X. Zhang, K. Xie, J. Li, Y. Liu, C. S. H. Ng, P. W. Y. Chiu, and Z. Li, "A surgeon preference-guided autonomous instrument tracking method with a robotic flexible endoscope based on dVRK platform," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2250–2257, Apr. 2022.
[83] B. Wang, J. Zheng, J. Yu, S. Lin, S. Yan, L. Zhang, S. Wang, S. Cai, A. H. A. Ahmed, L. Lin, F. Chen, G. W. Randolph, and W. Zhao, "Development of artificial intelligence for parathyroid recognition during endoscopic thyroid surgery," *Laryngoscope*, vol. 132, no. 12, pp. 2516–2523, Dec. 2022, doi: 10.1002/lary.30173.
[84] A. A. T. Zade, M. J. Aziz, H. Majedi, A. Mirbagheri, and A. Ahmadian, "Spatiotemporal analysis of speckle dynamics to track invisible needle in ultrasound sequences using convolutional neural networks: A phantom study," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 18, no. 8, pp. 1373–1382, Feb. 2023, doi: 10.1007/s11548-022-02812-y.
[85] M. Mushtaq, M. U. Akram, N. S. Alghamdi, J. Fatima, and R. F. Masood, "Localization and edge-based segmentation of lumbar spine vertebrae to identify the deformities using deep learning models," *Sensors*, vol. 22, no. 4, p. 1547, Feb. 2022, doi:10.3390/s22041547.
[86] Y. Ahmadyar, A. Kamali-Asl, H. Arabi, R. Samimi, and H. Zaidi, "Hierarchical approach for pulmonary-nodule identification from CT images using YOLO model and a 3D neural network classifier," *Radiological Phys. Technol.*, vol. 17, no. 1, pp. 124–134, Mar. 2024, doi: 10.1007/s12194-023-00756-9.
[87] Y.-S. Huang, P.-R. Chou, H.-M. Chen, Y.-C. Chang, and R.-F. Chang, "One-stage pulmonary nodule detection using 3-D DCNN with feature fusion and attention mechanism in CT image," *Comput. Methods Programs Biomed.*, vol. 220, Jun. 2022, Art. no. 106786, doi: 10.1016/j.cmpb.2022.106786.
[88] C. Liu, S.-C. Hu, C. Wang, K. Lafata, and F.-F. Yin, "Automatic detection of pulmonary nodules on CT images with YOLOv3: Development and evaluation using simulated and patient data," *Quant. Imag. Med. Surg.*, vol. 10, no. 10, pp. 1917–1929, Oct. 2020, doi: 10.21037/qims-19-883.
[89] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO based breast masses detection and classification in full-field digital mammograms," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105823, doi: 10.1016/j.cmpb.2020.105823.
[90] Y. Su, Q. Liu, W. Xie, and P. Hu, "YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms," *Comput. Methods Programs Biomed.*, vol. 221, Jun. 2022, Art. no. 106903, doi: 10.1016/j.cmpb.2022.106903.
[91] Q. Fu and H. Dong, "Spiking neural network based on multi-scale saliency fusion for breast cancer detection," *Entropy*, vol. 24, no. 11, p. 1543, Oct. 2022, doi: 10.3390/e24111543.
[92] Y. Ku, H. Ding, and G. Wang, "Efficient synchronous real-time CADe for multicategory lesions in gastroscopy by using multiclass detection model," *BioMed Res. Int.*, vol. 2022, pp. 1–9, Aug. 2022, doi: 10.1155/2022/8504149.
[93] D.-C. Cheng, T.-C. Hsieh, K.-Y. Yen, and C.-H. Kao, "Lesion-based bone metastasis detection in chest bone scintigraphy images of prostate cancer patients using pre-train, negative mining, and deep learning," *Diagnostics*, vol. 11, no. 3, p. 518, Mar. 2021, doi: 10.3390/diagnostics11030518.
[94] S. Li, Y. Li, J. Yao, B. Chen, J. Song, Q. Xue, and X. Yang, "Label-free classification of dead and live colonic adenocarcinoma cells based on 2D light scattering and deep learning analysis," *Cytometry A*, vol. 99, no. 11, pp. 1134–1142, Nov. 2021, doi: 10.1002/cyto.a.24475.
[95] N. Larpant, W. Niamsi, J. Noiphung, W. Chanakiat, T. Sakuldamrongpanich, V. Kittichai, T. Tongloy, S. Chuwongin, S. Boonsang, and W. Laiwattanapaisal, "Simultaneous phenotyping of five RH red blood cell antigens on a paper-based analytical device combined with deep learning for rapid and accurate interpretation," *Analytica Chim. Acta*, vol. 1207, May 2022, Art. no. 339807, doi: 10.1016/j.aca.2022.339807.
[96] Z. Han, H. Huang, D. Lu, Q. Fan, C. Ma, X. Chen, Q. Gu, and Q. Chen, "One-stage and lightweight CNN detection approach with attention: Application to WBC detection of microscopic images," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106606, doi: 10.1016/j.compbiomed.2023.106606.
[97] M.-Y. Quan, Y.-X. Huang, C.-Y. Wang, Q. Zhang, C. Chang, and S.-C. Zhou, "Deep learning radiomics model based on breast ultrasound video to predict HER2 expression status," *Frontiers Endocrinology*, vol. 14, Apr. 2023, Art. no. 1144812, doi: 10.3389/fendo.2023.1144812.
[98] R. Zhu, Y. Cui, J. Huang, E. Hou, J. Zhao, Z. Zhou, and H. Li, "YOLOv5s-SA: Light-weighted and improved YOLOv5s for sperm detection," *Diagnostics*, vol. 13, no. 6, p. 1100, Mar. 2023, doi: 10.3390/diagnostics13061100.
[99] G. Sun, C. Lyu, R. Cai, C. Yu, H. Sun, K. E. Schriver, L. Gao, and X. Li, "DeepBhvTracking: A novel behavior tracking method for laboratory animals based on deep learning," *Frontiers Behav. Neurosci.*, vol. 15, Oct. 2021, Art. no. 750894, doi: 10.3389/fnbeh.2021.750894.
[100] Z.-J. Huang, B. Patel, W.-H. Lu, T.-Y. Yang, W.-C. Tung, V. Bučinskas, M. Greitans, Y.-W. Wu, and P. T. Lin, "Yeast cell detection using fuzzy automatic contrast enhancement (FACE) and you only look once (YOLO)," *Sci. Rep.*, vol. 13, no. 1, p. 16222, Sep. 2023, doi: 10.1038/s41598-023-43452-9.

[101] Y.-F. Chen, Z.-J. Chen, Y.-Y. Lin, Z.-Q. Lin, C.-N. Chen, M.-L. Yang, J.-Y. Zhang, Y.-Z. Li, Y. Wang, and Y.-H. Huang, "Stroke risk study based on deep learning-based magnetic resonance imaging carotid plaque automatic segmentation algorithm," *Frontiers Cardiovascular Med.*, vol. 10, Feb. 2023, Art. no. 1101765, doi: 10.3389/fcvm.2023.1101765.

[102] M. A. Al-masni, W.-R. Kim, E. Y. Kim, Y. Noh, and D.-H. Kim, "Automated detection of cerebral microbleeds in MR images: A two-stage deep learning approach," *NeuroImage, Clin.*, vol. 28, Apr. 2020, Art. no. 10246410.1016/j.nicl.2020.102464.

[103] Y. Nambu, T. Mariya, S. Shinkai, M. Umemoto, H. Asanuma, I. Sato, Y. Hirohashi, T. Torigoe, Y. Fujino, and T. Saito, "A screening assistance system for cervical cytology of squamous cell atypia based on a two-step combined CNN algorithm with label smoothing," *Cancer Med.*, vol. 11, no. 2, pp. 520–529, Jan. 2022, doi: 10.1002/cam4.4460.

[104] A. Boonrod, A. Boonrod, A. Meethawolgul, and P. Twinprai, "Diagnostic accuracy of deep learning for evaluation of C-spine injury from lateral neck radiographs," *Heliyon*, vol. 8, no. 8, Aug. 2022, Art. no. e10372, doi: 10.1016/j.heliyon.2022.e10372.

[105] C.-P. Tang, C.-H. Hsieh, and T.-L. Lin, "Computer-aided image enhanced endoscopy automated system to boost polyp and adenoma detection accuracy," *Diagnostics*, vol. 12, no. 4, p. 968, Apr. 2022, doi: 10.3390/diagnostics12040968.

[106] H. Matsui, S. Kamba, H. Horiuchi, S. Takahashi, M. Nishikawa, A. Fukuda, A. Tonouchi, N. Kutsuna, Y. Shimahara, N. Tamai, and K. Sumiyama, "Detection accuracy and latency of colorectal lesions with computer-aided detection system based on low-bias evaluation," *Diagnostics*, vol. 11, no. 10, p. 1922, Oct. 2021, doi: 10.3390/diagnostics11101922.

[107] C.-P. Tang, H.-Y. Chang, W.-C. Wang, and W.-X. Hu, "A novel computer-aided detection/diagnosis system for detection and classification of polyps in colonoscopy," *Diagnostics*, vol. 13, no. 2, p. 170, Jan. 2023, doi: 10.3390/diagnostics13020170.

[108] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792, doi: 10.1016/j.compbiomed.2020.103792.

[109] Z. Kong, H. Ouyang, Y. Cao, T. Huang, E. Ahn, M. Zhang, and H. Liu, "Automated periodontitis bone loss diagnosis in panoramic radiographs using a bespoke two-stage detector," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106374, doi: 10.1016/j.compbiomed.2022.106374.

[110] W. Panyarak, K. Wantanajittikul, A. Charuakkra, S. Prapayasatok, and W. Suttapak, "Enhancing caries detection in bitewing radiographs using YOLOv7," *J. Digit. Imag.*, vol. 36, no. 6, pp. 2635–2647, Dec. 2023, doi: 10.1007/s10278-023-00871-4.

[111] Y. Tian, D. Zhao, and T. Wang, "An improved YOLO nano model for dorsal hand vein detection system," *Med. Biol. Eng. Comput.*, vol. 60, no. 5, pp. 1225–1237, May 2022, doi: 10.1007/s11517-022-02551-x.

[112] S. Pang, T. Ding, S. Qiao, F. Meng, S. Wang, P. Li, and X. Wang, "A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0217647, doi: 10.1371/journal.pone.0217647.

[113] M. A. Azam, C. Sampieri, A. Ioppi, S. Africano, A. Vallin, D. Mocellin, M. Fragale, L. Guastini, S. Moccia, C. Piazza, L. S. Mattos, and G. Peretti, "Deep learning applied to white light and narrow band imaging videolaryngoscopy: Toward real-time laryngeal cancer detection," *Laryngoscope*, vol. 132, no. 9, pp. 1798–1806, Sep. 2022, doi: 10.1002/lary.29960.

[114] J.-Y. Tsai, I. Y.-J. Hung, Y. L. Guo, Y.-K. Jan, C.-Y. Lin, T. T.-F. Shih, B.-B. Chen, and C.-W. Lung, "Lumbar disc herniation automatic detection in magnetic resonance imaging based on deep learning," *Frontiers Bioeng. Biotechnol.*, vol. 9, Aug. 2021, Art. no. 708137, doi: 10.3389/fbioe.2021.708137.

[115] B. Zhang, J. Li, Y. Bai, Q. Jiang, B. Yan, and Z. Wang, "An improved microaneurysm detection model based on SwinIR and YOLOv8," *Bioengineering*, vol. 10, no. 12, p. 1405, Dec. 2023, doi: 10.3390/bioengineering10121405.

[116] P. Rouzrokh, T. Ramazanian, C. C. Wyles, K. A. Philbrick, J. C. Cai, M. J. Taunton, H. M. Kremers, D. G. Lewallen, and B. J. Erickson, "Deep learning artificial intelligence model for assessment of hip dislocation risk following primary total hip arthroplasty from postoperative radiographs," *J. Arthroplasty*, vol. 36, no. 6, pp. 2197–2203, Jun. 2021, doi: 10.1016/j.arth.2021.02.028.

[117] T. Ma, Z. Ma, X. Zhang, and F. Zhou, "Evaluation of effect of curcumin on psychological state of patients with pulmonary hypertension by magnetic resonance image under deep learning," *Contrast Media Mol. Imag.*, vol. 2021, pp. 1–10, Jul. 2021, doi: 10.1155/2021/9935754.

[118] K.-C. Lee, Y. Cho, K.-S. Ahn, H.-J. Park, Y.-S. Kang, S. Lee, D. Kim, and C. H. Kang, "Deep-learning-based automated rotator cuff tear screening in three planes of shoulder MRI," *Diagnostics*, vol. 13, no. 20, p. 3254, Oct. 2023, doi: 10.3390/diagnostics13203254.

[119] J. Fu, J.-W. Chai, P.-L. Chen, Y.-W. Ding, and H.-C. Chen, "Quantitative measurement of spinal cerebrospinal fluid by cascade artificial intelligence models in patients with spontaneous intracranial hypotension," *Biomedicines*, vol. 10, no. 8, p. 2049, Aug. 2022, doi: 10.3390/biomedicines10082049.

[120] B. Lv, L. Wu, T. Huangfu, J. He, W. Chen, and L. Tan, "Traditional Chinese medicine recognition based on target detection," *Evidence-Based Complementary Alternative Med.*, vol. 2022, pp. 1–9, Jul. 2022, doi: 10.1155/2022/9220443.

[121] T. Till, S. Tschauner, G. Singer, K. Lichtenegger, and H. Till, "Development and optimization of AI algorithms for wrist fracture detection in children using a freely available dataset," *Frontiers Pediatrics*, vol. 11, Dec. 2023, Art. no. 1291804, doi: 10.3389/fped.2023.1291804.

[122] F. Varçın, H. Erbay, E. Çetin, I. Çetin, and T. Kültü, "End-to-end computerized diagnosis of spondylolisthesis using only lumbar X-rays," *J. Digit. Imag.*, vol. 34, no. 1, pp. 85–95, Feb. 2021, doi: 10.1007/s10278-020-00402-5.

[123] G. Oreski, "YOLO*C—Adding context improves YOLO performance," *Neurocomputing*, vol. 555, Oct. 2023, Art. no. 126655.

[124] M. Baghbanbashi, M. Raji, and B. Ghavami, "Quantizing YOLOv7: A comprehensive study," in *Proc. 28th Int. Comput. Conf., Comput. Soc. Iran (CSICC)*, Iran, Jan. 2023, pp. 1–5.

[125] M. Hu, Z. Li, J. Yu, X. Wan, H. Tan, and Z. Lin, "Efficient-lightweight YOLO: Improving small object detection in YOLO for aerial images," *Sensors*, vol. 23, no. 14, p. 6423, Jul. 2023.

[126] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "YOLO-based deep learning model for pressure ulcer detection and classification," *Healthcare*, vol. 11, no. 9, p. 1222, Apr. 2023.

[127] H. Ghabri, W. Fathallah, M. Hamroun, S. B. Othman, H. Bellali, H. Sakli, and M. N. Abdelkrim, "AI-enhanced thyroid detection using YOLO to empower healthcare professionals," in *Proc. IEEE Int. Workshop Mech. Syst. Supervision (IW_MSS)*, Nov. 2023, pp. 1–6.

[128] D. Chaudhary, A. Mathur, A. Chauhan, and A. Gupta, "Assistive object recognition and obstacle detection system for the visually impaired using YOLO," in *Proc. 13th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence)*, Jan. 2023, pp. 353–358.

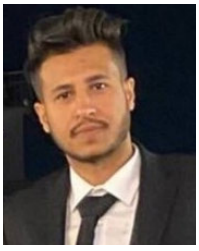
[129] B. Wu, C. Pang, X. Zeng, and X. Hu, "ME-YOLO: Improved YOLOv5 for detecting medical personal protective equipment," *Appl. Sci.*, vol. 12, no. 23, p. 11978, Nov. 2022.

[130] N. Andrade, T. Ribeiro, J. Coelho, G. Lopes, and A. Ribeiro, "Combining YOLO and deep reinforcement learning for autonomous driving in public roadworks scenarios," in *Proc. 14th Int. Conf. Agents Artif. Intell.*, 2022, pp. 793–800.

[131] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, and S. Mirjalili, "Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects," Tech. Rep., 2023, doi: 10.36227/techrxiv.23589741.v4.

MOHAMMED GAMAL RAGAB received the Bachelor of Science degree in software engineering from Universiti Teknologi PETRONAS in 2019, following the completion of his undergraduate degree, he continued his studies at Universiti Teknologi PETRONAS, pursuing the master's degree by research in machine learning. Currently, he is continuing his academic pursuits by pursuing the Ph.D. degree in information technology with Universiti Teknologi PETRONAS. His ongoing research builds on his previous work, focusing on the development of new and innovative techniques for optimizing the performance of deep learning models.

SAID JADID ABDULKADIR (Senior Member, IEEE) received the B.Sc. degree in computer science from Moi University, the M.Sc. degree in computer science from Universiti Teknologi Malaysia (UTM), and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS (UTP). He is currently an Associate Professor and a member of the Centre for Research in Data Science (CeRDaS), UTP. He is involved in flagship consultancy projects for PETRONAS under pipeline integrity, materials corrosion, and inspection. His research interests include machine learning, deep learning architectures, optimizations, and applications in predictive analytics. He is serving as the Treasurer for the IEEE Computational Intelligence Society Malaysia Chapter and the Editor-in-Chief for *Platform* journal.



AMGAD MUNEER received the B.Eng. degree (Hons.) in mechatronic engineering from Asia Pacific University of Technology and Innovation (APU), in 2018, and the master's degree in information technology from Universiti Teknologi PETRONAS, Malaysia, in 2022. He is currently with the Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, as a Research Assistant. He has authored several high-impact articles in well-reputed journals and conferences. His research interests include computer vision, AI applications for cancer data sciences, manufacturing data analytics, medical imaging, and bioinformatics. He is a reviewer of many international impact-factor journals.

ALAWI ALQUSHAIBI received the B.Sc. degree in computer networks and security from Universiti Teknologi Malaysia, in 2012, and the master's (by Research) degree from Universiti Teknologi PETRONAS (UTP), in 2021. He is currently an Academic Researcher. During his academic journey, he has acquired knowledge and skills in conducting independent research, producing academic writing, and teaching computer science courses. His research interests include machine learning, data science, optimization, feature selection, classification, data analytics, and image processing, specifically in generative adversarial networks (GANs).

EBRAHIM HAMID SUMIEA received the B.S. degree in software engineering from Asia Pacific University of Technology and Innovation (APU), in 2014, and the master's degree in management. He is currently pursuing the Ph.D. degree with Universiti Teknologi PETRONAS, delving deeper into the realm of artificial intelligence. He honed his skills in various programming languages and software development methodologies at APU. He is eager to explore the fusion of technology and business. His work aims to leverage DDPG's potential to address complex problems, demonstrating the powerful capacity of AI to transform various sectors. His research interests include reinforcement learning, specifically in the area of deep deterministic policy gradient (DDPG).

RIZWAN QURESHI (Senior Member, IEEE) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2021. His Ph.D. thesis focused on lung cancer drug resistance analysis using molecular dynamics simulation and machine learning. He joined the Fast School of Computing, National University of Computer and Emerging Sciences, Karachi, Pakistan, as an Assistant Professor. He is currently with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar. He has published his findings and methods in IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, *Pattern Recognition*, and IEEE BIBM Conference. His research interests include AI applications in life sciences, cancer data sciences, computer vision, and machine learning.

SAFWAN MAHMOOD AL-SELWI received the bachelor's degree in software engineering from Taiz University, Yemen, in 2012, and the master's degree in computer applications from Bangalore University, India, in 2018. He is currently a Research Assistant with the Department of Computer and Information Sciences, Universiti Teknologi PETRONAS (UTP), Malaysia. He has a total experience of eight years both in academic institutions and in the industry. His industry working experience is related to Android applications and website development. His research interests include artificial intelligence, machine learning, predictive and time-series analysis, metaheuristic algorithms, and optimization.

HITHAM ALHUSSIAN received the B.Sc. and M.Sc. degrees in computer science from the School of Mathematical Sciences, University of Khartoum, Sudan, and the Ph.D. degree from Universiti Teknologi PETRONAS (UTP), Malaysia. He was a Postdoctoral Researcher with the High-Performance Cloud Computing Center, UTP. He is currently a Senior Lecturer with the Department of Computer and Information Sciences, UTP. His current research interests include real-time parallel and distributed systems, big data and cloud computing, data mining, and real-time analytics.