10th International Scientific Conference
Technics, Informatics, and Education – TIE 2024
20-22 September 2024

Session: *Computer Sciences and Information Technology*
Professional paper
DOI: 10.46793/TIE24.083C

Application of the YOLO algorithm for Medical Purposes in the Detection of Skin Cancer

Stefan Ćirković[1*] [0009-0004-6775-1543] and Nikola Stanić[1] [0000-0002-8306-0273]
[1] University of Kragujevac/ Faculty of Technical Sciences, Čačak, Serbia
[*] stefan.cirkovic@ftn.kg.ac.rs

Abstract: *Skin cancer is one of the most common forms of cancer worldwide. Exposure to ultraviolet (UV) radiation increases the risk of its development. Early preventive examinations and early detection of suspicious skin changes are key factors for successful treatment. Due to the rapid development of AI technologies, neural networks have found application in various fields, including medicine. Neural networks can be used to create various applications, which would facilitate self-examination for patients and alert them to potential problems. This method would further save time and reduce healthcare costs. The paper presents the application of a neural network using the YOLO (You Only Look Once) algorithm on a dataset of mole images with the aim of identifying and classifying moles, which facilitates early intervention and improves treatment outcomes.*

Keywords: *skin cancer; yolo; neural network*

1. INTRODUCTION

Skin cancer has emerged as a significant health issue in modern society, particularly among young people. The rise in the incidence of this disease in recent decades can be attributed to various factors, including increased exposure to UV radiation due to frequent sun exposure and the use of tanning beds. Although skin cancer is the most common form of cancer, it is often preventable and treatable if detected early.

Young people are especially vulnerable due to the growing trend of spending time outdoors without adequate sun protection. Many are unaware of the risks associated with excessive UV exposure, which can lead to serious skin damage and an increased risk of developing malignant conditions. Education and prevention are crucial components in the fight against skin cancer, as raising awareness about the importance of skin protection can significantly reduce the number of affected individuals.

In 2020, there were an estimated 325,000 new cases of melanoma diagnosed globally, resulting in 57,000 deaths. Incidence rates vary significantly by region. Melanoma is generally more common in men than women in most regions. The highest incidence rates per 100,000 people were recorded in Australia and New Zealand (42 in men and 31 in women), followed by Western Europe (19 in both men and women), Northern America (18 in men and 14 in women), and Northern Europe (17 in men and 18 in women). In contrast, melanoma remains rare in most African and Asian countries, with incidence rates typically below 1 per 100,000. Based on global population trends, scientists predict that by 2040, there could be over 500,000 new melanoma cases and nearly 100,000 deaths annually [1].

Determining whether a mole is dangerous can be challenging, especially for someone without medical training. However, there are certain criteria and characteristics that can help assess the risk. The most commonly used method is the ABCDE rule:

- A (Asymmetry): Asymmetrical moles, where one half doesn't look like the other, can be a warning sign.
- B (Border): Irregular, jagged, or unclear borders of the mole can indicate malignancy.
- C (Color): Different colors or shades within the same mole (brown, black, red, white, blue) can be suspicious.
- D (Diameter): Moles larger than 6 mm (approximately the size of a pencil eraser) often warrant attention.
- E (Evolving): Changes in size, shape, color, or symptoms (such as itching or bleeding) of the mole are important to monitor.

For a definitive diagnosis and assessment, it is necessary to consult a dermatologist. Dermatologists use dermoscopy, a technique that allows for a detailed examination of the skin, to better evaluate moles. A dermatological examination cannot definitively determine if a mole is dangerous with complete certainty. Due to busy lifestyles, many people rarely find the time to visit a dermatologist. Additionally, these examinations can be costly and need to be conducted multiple times if changes in the mole are observed. Regular

monitoring and follow-up appointments are often necessary to keep track of any developments, making it challenging for individuals to maintain consistent check-ups. As a result, early detection and treatment of potentially malignant moles can be delayed, underscoring the need for accessible and affordable dermatological care. In most cases, a biopsy may be needed to definitively determine whether a mole is benign or malignant.

With the accelerated advancement of technology, people now have the opportunity to monitor their health using various applications, which can be useful for initial self-monitoring. These apps can serve as a starting point for later consultations with a doctor. It is important to note that for accurate diagnosis, seeking the opinion of a specialist such as a dermatologist is always necessary. This innovative approach to healthcare not only increases the accessibility of medical services but also empowers individuals to take proactive steps in monitoring their health. By using these applications, individuals can potentially identify concerning symptoms or changes in moles at an early stage, facilitating timely intervention and medical attention when needed.

One example of this is the use of the YOLO algorithm in dermatology for the early detection of skin cancer. The YOLO algorithm in medicine is utilized for the detection and classification of medical objects in images or video recordings. This technology enables rapid and precise identification of pathological changes, such as tumors on X-ray images, anomalies on MRI scans, or skin alterations that may indicate diseases like skin cancer. The application of the YOLO algorithm in medicine promises more efficient diagnostics and early disease detection, potentially enhancing healthcare and reducing the number of missed cases.

The aim of this scientific study is to train a model using the YOLO v8 algorithm, which could accurately assess whether a mole belongs to one of two categories with as high precision as possible.

2. LITERATURE REVIEW

Various studies have been conducted on the topic of skin cancer, employing older versions of the YOLO algorithm or alternative combinations of neural networks. These are showcased below.

Ünver et al. (2019) introduced a robust methodology for segmenting skin lesions in dermoscopic images, which merges the GrabCut algorithm with the YOLO deep convolutional neural network. Their approach underwent testing on the PH2 and ISBI 2017 datasets, both widely utilized public datasets within the domain of Skin Lesion Analysis Towards Melanoma Detection Challenge Dataset. The proposed method achieved an impressive accuracy rate of 93.39% [2].

In a study [3], a YOLO-based deep neural network was proposed for classifying nine types of skin cancer. Both YOLOv3 and YOLOv4 versions were analyzed and compared objectively for improved skin cancer diagnosis. The proposed neural network achieved a mean average precision score of 88.03% for YOLOv3 and 86.52% for YOLOv4. Experimental analysis indicated that both YOLOv3 and YOLOv4 are well-suited for classifying various skin diseases, with YOLOv4 generally outperforming YOLOv3 in most cases. YOLOv4 demonstrated its superiority by achieving the highest scores across all evaluation metrics when compared to conventional methods.

In a study [4], the authors proposed a deep learning CNN model utilizing AlexNet as a pre-trained model. This transfer learning approach was chosen because AlexNet has been extensively trained for object recognition, which is closely related to skin lesion classification. The last three layers of AlexNet—the final fully connected layer, softmax layer, and classification layer—were replaced with layers suitable for binary classification. All images were resized to 227×227 to match the input size of AlexNet. The model was trained using the training set with the SGDM algorithm, an initial learning rate of 0.0001, a mini-batch size of 30, and 40 epochs. The proposed model achieved an area under the receiver operating characteristic (ROC) curve of 0.91. With a confidence score threshold of 0.5, the model obtained a classification accuracy of 84%, sensitivity of 81%, and specificity of 88%. The authors suggested that the proposed approach could be deployed to assist dermatologists in skin cancer detection and could also be integrated into smartphones for the self-diagnosis of malignant skin lesions. They concluded that this could expedite cancer detection, which is critical for effective treatment.

In a study [5], the authors introduced a novel approach integrating YOLO v3 with a deep convolutional neural network (DCNN) for the purpose of skin lesion detection and classification... The authors reported that their YOLO v3-DCNN technique achieved a remarkable accuracy rate of 95% when evaluated on the HAM10000 dataset, demonstrating superior performance compared to earlier methodologies.

3. YOLO (You Only Look Once)

YOLO is an object detection algorithm introduced in 2015 by Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. YOLO's architecture brought a significant revolution in real-time object detection, surpassing its predecessor – the Region-based Convolutional Neural Network (R-CNN). YOLO is a single-pass algorithm, where one neural network predicts bounding boxes and class probabilities using a full image as input [6].

3.1. Architecture of YOLO v8

Figure 1 depicts the architecture of YOLOv8, which encompasses three key components: backbone, neck, and head. The configuration of these components may vary across different versions of YOLO, and improvements in each of them have led to significant enhancements in accuracy and prediction speed. The latest versions of YOLOv8 introduce enhancements in all three components to achieve better performance [7].

- **Backbone**: It plays a crucial role in extracting significant features from the input image. Typically, a convolutional neural network trained on large datasets such as ImageNet is used. The backbone is responsible for feature extraction and generating feature maps from input images. Within YOLO, some commonly used backbone networks include VGG16, ResNet50, CSPDarknet53, and EfficientNet [7].
- **Neck** It represents the connection between the backbone and head in the YOLO architecture. It utilizes advanced techniques of feature pyramid aggregation, such as PANet, to combine features from different layers. The role of the neck is to fuse feature maps from various layers of the backbone network and pass them to the head. Popular options for the neck in YOLO include Spatial Pyramid Pooling (SPP), Feature Pyramid Network (FPN), NAS-FPN, and Rep-PAN [7].
- **Head** It's a part of the architecture consisting of predictive layers that generate final classifications. Each level (P2 to P5) is connected to specific detection blocks that predict bounding boxes and object classes [7].

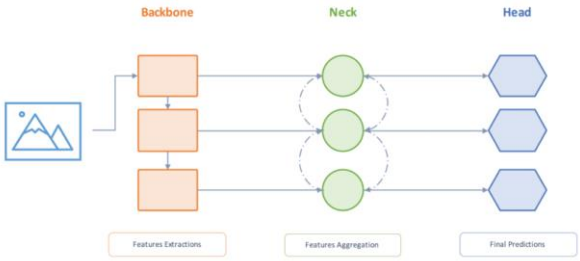

Figure 1. *The architecture of YOLO consists of a backbone, neck, and head [7]*

3.2. Comparison of YOLO Versions

The YOLO algorithm has gone through several versions, with each bringing significant improvements in object detection accuracy and speed. YOLOv2 introduced techniques such as batch normalization and anchor boxes, enhancing precision and processing speed. YOLOv3 added feature extraction at multiple scales and a new backbone network, allowing for a better balance between speed and accuracy. Versions YOLOv4 and YOLOv5 improved backbone networks, data augmentation, and training strategies, increasing accuracy without significant loss in real-time performance. PP-YOLO, developed by Alibaba Group, introduced a new backbone network and spatial attention module, making it faster and more accurate than YOLOv5. YOLOv6 implemented the EfficientNet architecture, resulting in even faster and more precise results. YOLOv7 reduced parameters and computational costs while maintaining high speed and accuracy. YOLOv8 improved feature aggregation and introduced a model without anchor boxes, directly predicting object centers [7].

3. METHODOLOGY

During the training of the model, a private dataset consisting of two classes (benign and malignant) was used, with each class containing 1500 images. The dataset was carefully collected and annotated in accordance with ethical standards. Figure 2 shows the appearance of the dataset.

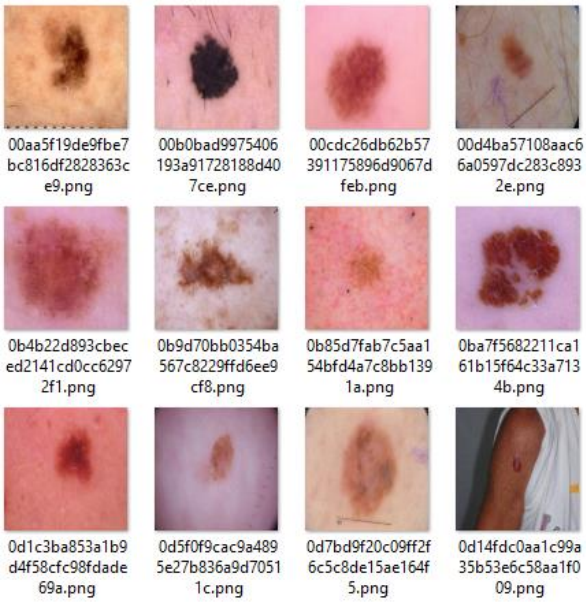

Figure 2. *Dataset*

The images were pre-processed to ensure size and quality. This step involved resizing the images to 224x224 pixels. To enable the application of the YOLO algorithm for classification, it was necessary to annotate the images. Each image was individually labeled using the Labelme tool, which is available as a Python package. After annotation, each image has a corresponding JSON file that matches its name. The JSON file consists of coordinate points that indicate the relevant parts of the image. Figure 3 shows the annotation of an image in the Labelme tool.

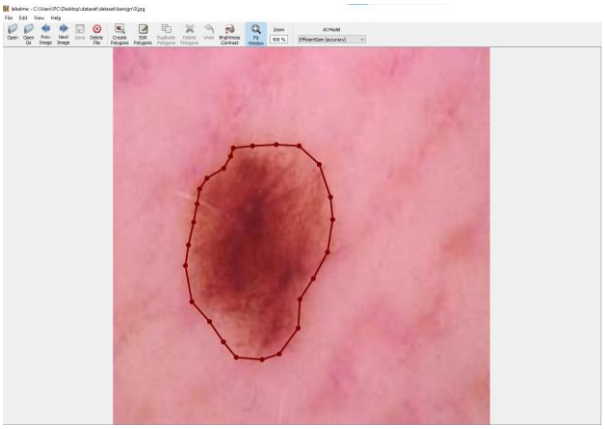

Figure 3. *Annotated image in the Labelme tool*

Images that were not suitable for annotation were excluded from the dataset. Since YOLO does not use JSON files for training, but rather TXT files, the annotations were converted from JSON to TXT format using the labelme2yolo tool. The dataset was then divided into three parts: training, testing, and validation in a ratio of 80:10:10.

After preparing the dataset, the YOLO algorithm was trained using this data.

Model was trained using NVIDIA GeForce 1080. During training, the following parameters were used: 100 epochs and a batch size of 16. The total training duration was 4 hours and 10 minutes.

4. RESULT AND INTERPRETATION

In this chapter, the results and their interpretations obtained after training the YOLO model are presented. Based on the confusion matrix shown in Figure 4 and Table 1, performance metrics were analyzed. The matrix illustrates how the model classifies images into three categories: benign, malignant, and background. Background represents parts of the image that were not annotated. Each cell of the matrix displays the number of instances correctly or incorrectly classified by the model.

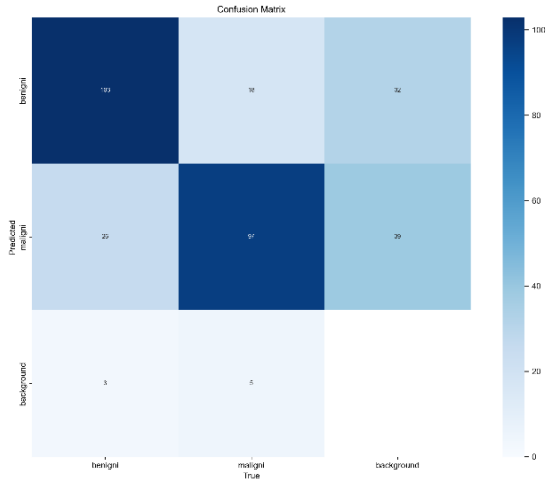

Figure 4. *Confusion Matrix*

In this example, the model correctly classified 103 instances as benign, but made an error by classifying 26 malignant instances as benign and 3 background instances as benign. As for malignant instances, the model correctly identified 97, but misclassified 18 benign and 5 background instances as malignant. The model did not identify any background instances, which is acceptable. Based on the obtained confusion matrix, it can be concluded that the model provides good and acceptable results for preliminary screening, where positive findings will be further reviewed by dermatologists.

Table 1. *Confusion Matrix*

	Benigni (True)	Malignant (True)	Background (True)
Benigni (Pred.)	103	18	32
Malignant (Pred.)	26	97	39
Background (Pred.)	3	5	0

Figure 5 displays a chart for all classes, precision reached a value of 1.00 when the confidence threshold was 0.848. This means that all predicted positive instances were correctly classified when the model had a precision of 84.8%. Precision is defined as the ratio of true positive predictions to the total positive predictions. It is particularly useful in scenarios where the false positive rate is high.

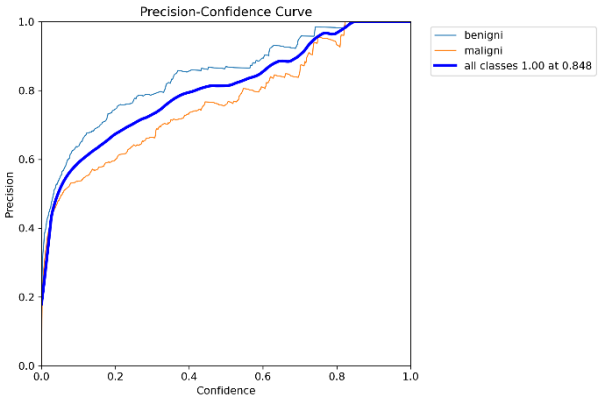

Figure 5. *Precision Confidence Curve*

This metric is expressed by formula (1):

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

where:

- TP (True Positive) is the number of correctly identified positive predictions by the model.
- FP (False Positive) is the number of predictions that are incorrectly classified as positive by the model.

At this confidence threshold, the model achieved 100% precision, meaning there were no false positive predictions. This result demonstrates that the model exhibits a high level of reliability and accuracy at high confidence thresholds, which is crucial for applications where false positive results are unacceptable.

Figure 6 displays the recall, which measures the model's ability to identify all positive instances. This metric becomes particularly significant when the false negative rate is high. In the example, the model achieves a recall of 0.99 (99%) when the confidence threshold is 0.000. This means that the model successfully identifies 99% of true positive instances, practically missing almost none of the actual positive instances.

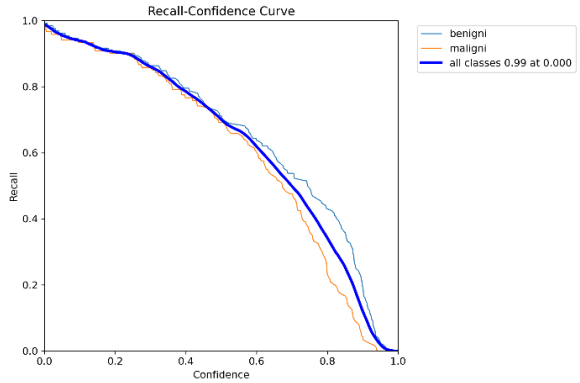

Figure 6. *Recall-Confidence Curve*

It is calculated by the following formula (2):

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where:

- TP (True Positive) is the number of correctly identified positive instances by the model.
- FN (False Negative) is the number of instances that are actually positive but incorrectly classified as negative by the model.

On Figure 7, a graph is shown indicating that the F1-score for all classes reached a value of 0.80 when the confidence threshold was 0.357.

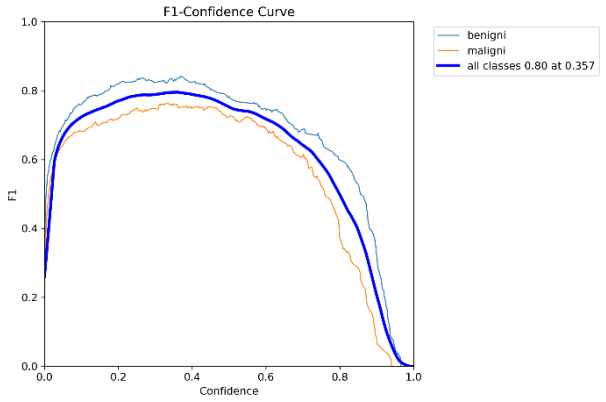

Figure 7. F1 Confidence curve

The F1-score is the harmonic mean between precision and recall and is used as a measure of overall model performance, especially when there is class imbalance. It is calculated by the following formula (3).

$$F1 - score = 2\,x\,\left(\frac{Precision\ x\ Recall}{Precision + Recall}\right) \tag{3}$$

The F1-score value indicates that the model has good overall performance in classifying instances when the confidence threshold is relatively low. This result is useful for evaluating the model in situations where both precision and recall are important.

Validation was performed on the validation dataset consisting of 248 images, as shown in Table 2, achieving an overall precision of 78.2%. This result is considered acceptable for preliminary screening, as dermatologists will further examine positive findings.

Table 2. *Results at the validation set*

Class	Instances	Precision	Recall
All	252	0.782	0.816
Benign	132	0.845	0.826
Malignant	120	0.718	0.807

In the following figure, Figure 8, lesion detection during the validation process is depicted.

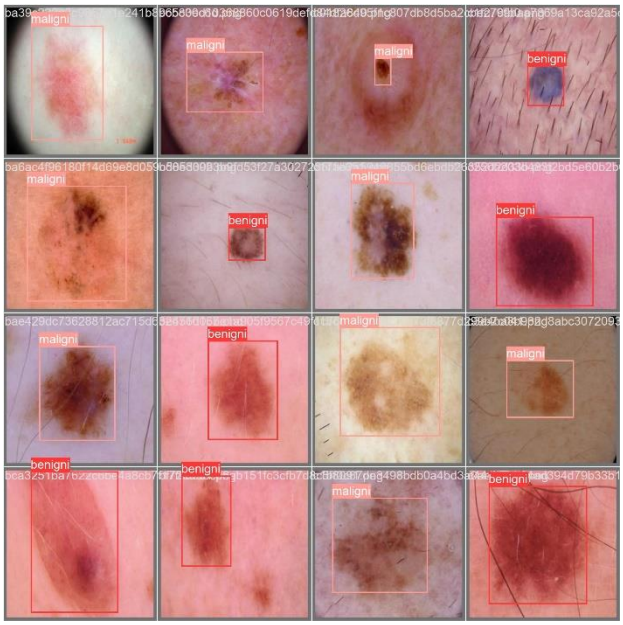

Figure 8. *Lesion Detection during Validation*

5. CONCLUSION

One of the purposes of this study is to raise awareness about the importance of regular mole checks and to find a solution that allows for simpler monitoring accessible to all classes of the population. In this study, the Yolo v8 algorithm was applied for the first time to detect two types of moles (benign and malignant). The theoretical foundation of the YOLO algorithm was comprehensively presented, along with an analysis of its architecture and the evolution of previous versions, emphasizing the advantages of the new versions. The methodology described the dataset and its preparation for the application of the YOLO algorithm. In the results section, the results with metrics Precision, Recall, and F1 score were presented. The achieved precision on the validation set for predicting benign lesions is 84.5%, for malignant lesions 71.8%, while the overall precision is 78.2%. Although the results are not impressive in terms of precision, we emphasize that

the goal of the application is preliminary screening, where dermatologists will further review the findings. This algorithm cannot be entirely relied upon without the supervision of dermatologists and medical assessment, as even the examination by dermatologists alone is not always precise without additional analysis.

Further improvements include tracking new versions of the YOLO algorithm, as well as considering other datasets to achieve higher precision, and the obtained model should be used to create some type of application.

ACKNOWLEDGEMENTS

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-66 / 2024-03 / 200132 with University of Kragujevac - Faculty of Technical Sciences Čačak.

REFERENCES

[1] Arnold, M., Singh, D., Laversanne, M., Vignat, J., Vaccarella, S., Meheus, F., Cust, A. E., de Vries, E., Whiteman, D. C., Bray, F. (2022). Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. *JAMA Dermatology*, 158(5), 495-503. doi: 10.1001/jamadermatol.2022.0160

[2] Ünver, H. M., Ayan, E. (2019). Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. *Diagnostics*, 9(3), Article 72. doi: 10.3390/diagnostics9030072

[3] Aggarwal, A., Das, N., Sreedevi, I. (2019). Attention-guided deep convolutional neural networks for skin cancer classification. *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 1-6. doi: 10.1109/ipta.2019.8936100

[4] Ameri, A. (2020). A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images. *Journal of Biomedical Physics and Engineering*, 10(6), 801-806. doi: 10.31661/jbpe.v0i0.2004-1107

[5] Manoj, S. O., Abirami, K. R., Victor, A., Arya, M. (2023). Automatic Detection and Categorization of Skin Lesions for Early Diagnosis of Skin Cancer Using YOLO-v3 - DCNN Architecture. *Image Analysis & Stereology*, 42(2), 101-117. doi:10.5566/ias.2773

[6] Ultralytics. (n.d.). *Ultralytics Documentation*. [Online]. Available: https://docs.ultralytics.com/. Accessed: 4 June 2024.

[7] Ragab, M. G., Abdulkadir, S. J., Muneer, A., Alqushaibi, A., Sumiea, E. H., Qureshi, R., Al-Selwi, S. M., Alhussian, H. (2024). A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access*. doi: 10.1109/access.2024.3386826