Detection and Classification of Oral Cancer using YOLO Object Detection Algorithm

Kavyashree C[1], H S Vimala[1] and Shreyas J[2]

[1]*Department of Computer Science , UVCE, Bangalore University, Bangalore, India- 560001*
[2]*Department of Information Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India - 576104*

Abstract: The early detection of oral cancer plays a pivotal role in enhancing the survivable rate of the patients. Recent advancement in artificial intelligence have made the diagnosis rapid and precise. The advent of deep learning has transformed medical image analysis, facilitating more precise, efficient, and automated evaluations of medical images. It serves the purpose of identifying and locating particular objects within medical imaging. The aim of this research is to develop a deep learning-powered system for diagnosing oral cancer, capable of distinguishing between cancerous and non-cancerous areas in a provided image. The Yolo is a cutting edge deep learning model employed for object detection, segmentation and classification. The system was retrained for the oral cancer dataset. The images are annotated with the help of the experts. A balanced dataset is created by data augmentation by rotating and flipping the images. The blurring is used to pre-process the images. The Yolov8 architecture has been enhanced through the integration of EfficientNet-B0 for the generation of feature maps, along with the implementation of a Feature Pyramid Network (FPN), which facilitates the detection of objects across various scales. Following that, the model is trained with the images and then validated using YOLOv8 model. The normal and abnormal part of an images are identified with a precision of 0.901. The mean Average Precision(mAP) obtained for the model is 0.913. The is YOLOv8 model is compared with other objection detection model such as YOLOv7, Mask R-CNN (Region based Convolutional Network) and Faster R-CNN. YOLOv8 is found to be the fastest object detection and classification framework compared to the other three models. These results greatly helps the medical practitioner to perform the initial investigation to and helps in the early detection of an oral cancer.

Keywords: Deep Learning methods, YOLO models, Object Detection, Classification, Oral Cancer

1. INTRODUCTION

Oral cancer is a form of cancer that develops in the tissues of the mouth or throat and is categorized as one of the head and neck cancers. It is essential to comprehend the causes, symptoms, diagnosis, treatment options, and preventive measures in order to effectively manage the condition and improve outcomes. Squamous Cell Carcinoma (SCC) is the predominant form of oral cancer, typically manifesting on the lining of the oral cavity. Tobacco usage is one of the largest risk factors for mouth cancer that includes chewing tobacco, smoking cigarettes [1]. Also, genes can occasionally cause cancer due to DNA mutations. According to the data, men have a higher likelihood of developing oral cancer than women [2]. It further indicates that this type of cancer is among the most preventable forms if detected in its early stages. The oral cancer can be diagnosed through biopsy which is considered as the highest standard in the cancer detection. Also imaging methods like MRI, CT and PET scans can be used for the cancer detection. A photographic images also can be used for the cancer detection. This study mainly focuses on the oral cancer detection using photographic images.

The development in artificial intelligence has assisted in automating the cancer detection by delivering precise and affordable outputs. Deep learning entails instructing artificial neural networks to identify patterns in data and make predictions. Processing input data through the network, calculating the output, and then evaluating it against the anticipated output helps in automating the process [3]. Object detection plays a vital role in computer vision by identifying and locating objects within medical images. YOLO, an acronym for You Only Look Once, is a widely utilized and effective deep learning model employed for real-time object detection. It utilizes a classifier across different areas of an image, approaching the task as a unified regression problem [4]. In 2016, Ross Girshick, Ali Farhadi, Santosh Divvala, and Joseph Redmon [5] introduced Yolo to detect and categorise objects in images as well as in videos. YOLO receives a single image or video frame as input, it instantly creates bounding boxes around any objects that are discovered. The YOLO algorithm partitions an image into a grid, making predictions for bounding boxes and class probabilities in each grid cell. Combining these predictions yields the whole set of bounding boxes and the class with the probabilities. Deep convolutional neural networks are used to train the algorithm, which enables it to learn features that are characteristic of many object categories. The system is trained using massive datasets of labelled photos. There have been numerous iterations of

E-mail address: kavyashree.csd@gmail.com

YOLO, each with unique additions and improvements. An overview of some of the more popular variations is provided below: YOLOv1 was the original iteration of YOLO, released in 2016. Despite its speed and accuracy, there were some limitations, including challenges in identifying small objects. YOLOv2 is the second version of YOLO was released in 2017 and, in addition to a new architecture and the addition of features like batch normalisation and anchor boxes, it addressed some of the shortcomings of the original version. It surpassed its predecessor regarding both accuracy and speed. YOLOv3 was released in 2018 with a few modifications to the model's performance, including the usage of residual blocks and feature pyramid networks and faster and more accurate. It also unveiled a novel feature extractor known as Darknet-53. YOLOv4 was released in 2020 with the implementation of Scaled-YOLOv4 and the usage of the CSPDarknet53 backbone, multi-input weighted residual connections, and other features. On a number of object detection benchmarks, it attained cutting-edge performance. YOLOv5 is a completely redesigned and re-implemented version of YOLO that was released in 2020. The use of anchor-free object detection, improved data augmentation methods, and more effective training are just a few of the enhancements it makes.

YOLOv6 was funded by Meituan in 2022 and used in robots. YOLOv7 implements pose estimation on the COCO keypoints dataset [6]. The most recent version of YOLO is called YOLOv8 by Ultralytics. YOLOv8 expands on the success of earlier editions state-of-the-art model by enhancing the features and performance.YOLOv8 supports a wide range of visual AI tasks such as detection, estimation, segmentation, tracking, and classification. Because of its adaptability, YOLOv8 can be employed in numerous applications[7] and we are using it for the oral cancer detection.

This study is focused on YOLO based technique to detect and classify the oral cancer images that identifies the region of cancer affected area.

The contributions made by this study are as follows:

1) Annotation of Oral cancer images using experts.
2) Augmentation techniques to improve the quality of the dataset.
3) This study employs YOLOv8 architecture In integration with EfficientNet B0 and FPN for object detection in images.
4) The performance analysis is performed comparing with other object detection models.

2. Literature Review

The objective of processing the medical images is to create computer aided algorithms to forecast the future developments in the health care systems. Thus, there are many algorithms developed using several artificial intelligence techniques for detecting several different cancers, including prostate, skin, and breast cancer. The similar idea is adapted for the oral cancer detection and grading of images.

Gao et al. [8] integrated the Yolov7 version with a coordinate attention mechanism to enhance the feature extraction process, ensuring that significant features are not overlooked. The streamlined Feature Pyramid Network (FPN) and the anchor-free model decrease the overall complexity. The sophisticated loss function contributes to enhancing both the accuracy and the resilience of the model. This comprehensive model is capable of identifying irregularities within dental images. Hsu et al. [9] used Yolov7 model for oral mucous lesion detection and classification. The model is trained using 50,000 macroscopic images, each representing various grading. It categorizes images not solely as benign or malignant; it is also capable of identifying potentially malignant lesions. The YOLOv7-E6 variant demonstrated strong performance in both precision and recall, while the YOLOv7-D6 variant excelled in achieving a high F1-score. In addition to these metrics, the proposed model demonstrated enhanced accuracy in the detection of lesions. Mammeri et al. [10] used Yolov7 for lung nodule detection. This technology has the potential to create bounding boxes around the nodules, thereby assisting radiologists in tracking and identifying them within the entirety of the slide images. The model attained a mAP of 81.28%, despite the absence of image preprocessing. It additionally conducted classification across multiple classes utilizing a pretrained VGG16 model. It also emphasized the importance of determining the degree of malignancy, which has the potential to enhance the overall diagnostic process. Prinzi et al. [11] employed the Yolov5 model to identify suspicious objects within mammograms, thereby assisting in the detection of breast cancer. The implementation of transformers in place of convolution layers also facilitated the detection of smaller objects. Transformers assist in reducing the dimension of a vector to achieve a more compact output. Eigen – CAM notably decreases the occurrence of false negatives, although it results in a rise in false positives. The model could able to obtain 0.621 mAP, which can subsequently assist in the evaluation process. Chou et al. [12] proposed a Yolo based framework foe esophageal cancer detection using with version v5 and v8. The application of enhancers for white light images has significantly enhanced the detection of carcinoma in comparison to RGB images. Yolov5 has demonstrated superior performance with the dataset in comparison to Yolov8, particularly excelling in feature learning capabilities, while Yolov8 is noted for its high precision. The model underwent rigorous training for 500 iterations, achieving a precision of 0.85 with Yolov5 and 0.81 with Yolov8. Salman et al. [13] created a model for the prostate cancer diagnosis and grading of biopsy images using Yolov3 by fine tuning its activation function fro ReLu to Sigmoid and Tanh activation function because of the features being non linearly distributed. Pacal et al. [14] suggested a yolo based identification of polyps by improving the performance using Cross Stage Partial Network (CSPNet) that enables instantaneous detection. The model's

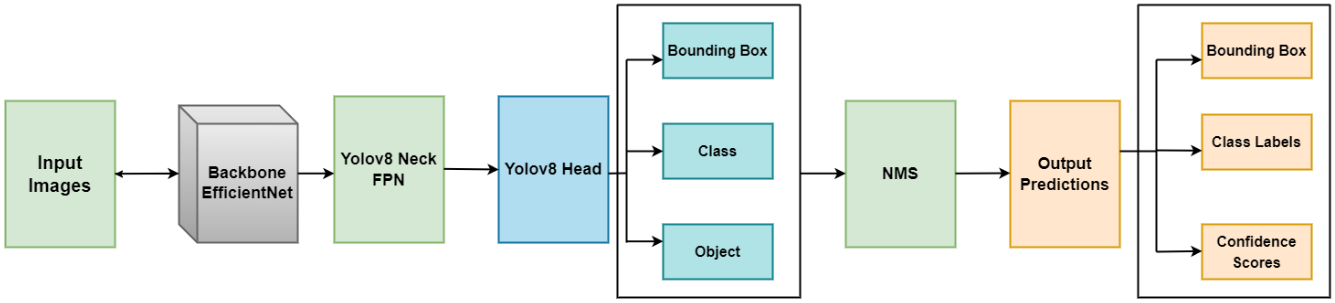


Figure 1. Proposed Yolov8 architecture

performance and clinical applicability are improved by the use of a large dataset during training. Baccouche et al. [15] used Yolo model to identify and categorise worrisome breast lesions from the whole set of mammograms. The model is validated with public as well as the privately collected database. Also suggested the fusion model approach to enhance the performance. [16] also proposed Image-to-image translation methods for mammography pairs data reconstruction. The model performs single class and well as multi class prediction for detecting the mass lesion. Karaman et al. [17] designed a system based on real-time identification of polyps with Yolov5 that uses artificial bee colony to optimize the activation functions. The study employs the different versions of Yolo model saves the best model in order to enhance the activation processes and hyper parameters. Nersisson et al. [18] used Yolov2 model to extract the features from lesion for skin cancer detection. conventional characteristics, such as texture and color features are combined with the featured obtained from Convolutional neural networks (CNN). This fusion model has improved the overall performance in the skin cancer detection. Hamed et al. [19] compared the machine learning models with the YOLO and RetinaNet model and found Yolo it provides the highest accuracy compared to other models. Aly et a. [20] used Yolo to detect mammograms and it is found to be the better object detection model compared to CNN models. The reason for the enhanced object detection is the use of anchor boxes in yolov3 that employs the k-means clustering algorithm. It provides the average precision of 94.2% and the classification accuracy of 84.6%. The study used a small annotated dataset and the major limitation is it could not detect the small masses. Nie et al. [21] employed the Yolov1, yolov2 and yolov3 models based deep CNN for skin cancer identification and achieved a mean precision average of 0.82 with the limited training images of 200. This is designed for light weight applications like mobile applications. Victor et al. [22] used yolov3 for the identify and categorize the skin cancer. Yolov3 is used to generate the feature map and is combined with the color features extracted from Quad histogram. The integration of these features is input into the deep convolution neural network for cancer detection, resulting in a commendable level of accuracy. Ji et al. [23] used Yolo to find lung cancer in computed tomography (CT) images. A one stage model has been developed that improve the feature layer's overall multi-scale representation capability. They compared the model to other cutting-edge models and found that it performed better in terms of particular criteria like recall. Patel et al. [24] used Yolov8 for breast cancer detection along with ResNet50 trained on BreakHis data. Data augmentation is performed to have a balanced dataset. This model performed better with accuracy of 97.8% and false positive rate of 1.2% suitable of real time analysis. Pham et al. [5] used Yolov8 to detect and classify tumors in ovary in the ultrasound images. The research indicates that Yolov8 demonstrated a 19% increase in precision compared to Yolov7, although it also exhibited a decrease in speed. The study showcases a comparative analysis of multiple iterations of Yolo models, highlighting the respective values observed. The research further determined that additional studies are necessary for the real-time analysis and detection of ovarian tumors and ca not afford to miss even a single object.

3. Methodology

A. *YOLOv8*

YOLOv8 was designed by Glenn Jocher at Ultralytics and was released on January 10th, 2023. This most recent iteration of the well-known model is used for real-time object identification and image segmentation. Due to its rapid, accurate, and user-friendly design, YOLOv8 offers an excellent solution for various tasks such as object identification, classification of images and instance segmentation [25]. YOLOv8 estimates a target's centre directly instead of predicting its offset from a predefined anchor box. A complex post-processing technique called Non-Maximum Suppression (NMS), which sorts through candidate detection after inference, is sped up by with the absence of anchors since it reduces the number of boxes being obtained. 3*3 convolution layers are added instead of 6*6 convolution layers.

A deep neural network called YOLOv8 employs a succession of convolution layers for the extraction of data from the obtained image in turn producing the bounding boxes and classes in the output. The architectural structure is composed of multiple essential elements, including the SPPF (Spatial Pyramid Pooling - Fast) layer, C2f module, Detection module, and the backbone. The Backbone is responsible for extracting high-level features from the input image. The SPPF layer, along with subsequent convolutional layers, handles features at various scales. The C2f module integrates advanced functionalities with contextual data to enhance detection precision, while the Detection module employs convolutional and linear layers to produce the bounding boxes and class probabilities. Along with these main components, there are several other layers used in the YOLOv8 architecture, include Up sample and Concat layers, which enhance the resolution of feature maps and merge feature maps from various layers, respectively. The COCO dataset is used for pretraining the Detect, Segment, and Pose models, while the ImageNet dataset is used for pretraining the Classify models. YOLOv8 offers different versions such as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x that runs on different speed and used different parameters.Yolov8 provides the high accuracy rate according to COCO and Roboflow 100 measurements and also easy for the developers to use. In other words, instead of forecasting the deviation of an object from a recognized anchor box, it directly forecasts the object's center. This provides a solution to the major difficult component of previous Yolo models.

B. *Proposed Yolov8 based object detection for oral cancer detection*

Figure.1 shows the methodology followed in our study that incorporates the improvement to the Yolov8 Model. Yolov8 can be integrated with various backbone like CSPDarknet53, EfficientNet, ResNet, Inception and other pretrained models for feature extraction. CSPDarknet53 proves to be highly effective and perfectly compatible with yolov8, providing the ideal combination of balance and top-notch performance. Due to its computational complexity and superior accuracy, EfficientNet has been chosen as the backbone network. EfficientNet-B0 was specifically selected for its lightweight design, while still delivering impressive accuracy at a lower computational expense. EfficientNet utilizes depth-wise separable convolutions as part of its design to maintain simplicity. This aids in lowering the computational expenses while maintaining the same level of accuracy. Accurate feature extraction is crucial for health care needs, and our research with limited resources benefited greatly from the use of EfficientNet-B0. The EfficientNet-B0 has been integrated with Yolov8, enabling it to accept input and process it through the Yolov8 head.

The neck merges the characteristics derived from the backbone and transmits them to the head for the ultimate detection and classification. The proposed methodology uses Feature pyramid network (FPN), that aggregates the feature from different levels. This boosts the model's capability to recognize objects of diverse sizes. FPN enhances higher-level features through up sampling, integrates them with lower-level features, and subsequently forwards the combined information to the next stage. FPN is utilized in conjunction with convolutional layers to enhance the model's performance. The neck must possess sufficient efficiency to support the functions of the spine without impeding its performance.

The head of the network is responsible for producing the final output. It processes the features extracted and aggregated by the backbone and neck components to generate predictions such as bounding boxes, class scores, and objectness scores. The head plays a vital role in transforming the feature maps into significant detection outcomes. Yolov8 employs a regression-based method to anticipate the coordinates of the bounding box surrounding the identified objects. The Objectness Score is a measure of the probability that a bounding box encloses an object, allowing for the distinction between background and object. The classification layers determine the likelihood of each class for the identified object.

4. Experimental Setup

A. *Dataset Preparation*

Dataset preparation is crucial in the proposed algorithm as it is founded on the deep neural network. Oral cancer photographic image dataset is taken from [26] that contains 323 images. The annotation of these images are crucial that greatly helps in the detection. The annotation of the images is done with the help of experts from Suraksha Speciality Dental Care, Hoskote, India. The labels used for annotation are Abnormal and Normal regions. The precise annotation leads to better detection of objects and improves the overall performance. Figure 2 shows the annotation of oral cancer images as Normal and Abnormal objects. Multiple objects are present in each image.

B. *Data Augmentation and Preprocessing*

Utilizing balanced datasets can lead to more accurate outcomes and mitigate the risk of overfitting in the model. The dimension of the dataset is increased with augmentation through horizontal flipping, rotating the images and blurring upto 10px. The data set is significantly increased to 689 images. The images are preprocessed by resizing it to 640*640 as required for YOLOv8 architecture. The augmented and preprocessed dataset is divided with 80% training set, 12% testing set and 8% for validation.

Figure 3 shows the proposed methodology in detecting the lesion in input images and classification with the confidence score . The Yolov8 is used to train the model with the training data. It uses the sigmoid linear units (SiLU) [27] as an activation function that replaces the leaky Relu only in the convolution and batch normalization layer of CNN. The outputs are taken from the sigmoid function.

$$SiLU(x) = x\frac{1}{1+e^{-x}} \tag{1}$$

The intersection over union (IoU) quantifies the overlap between the predicted bounding box and the ground truth bounding box. Figure. 4 shows the intersection of the bounding boxes and extracting the bounding box from the

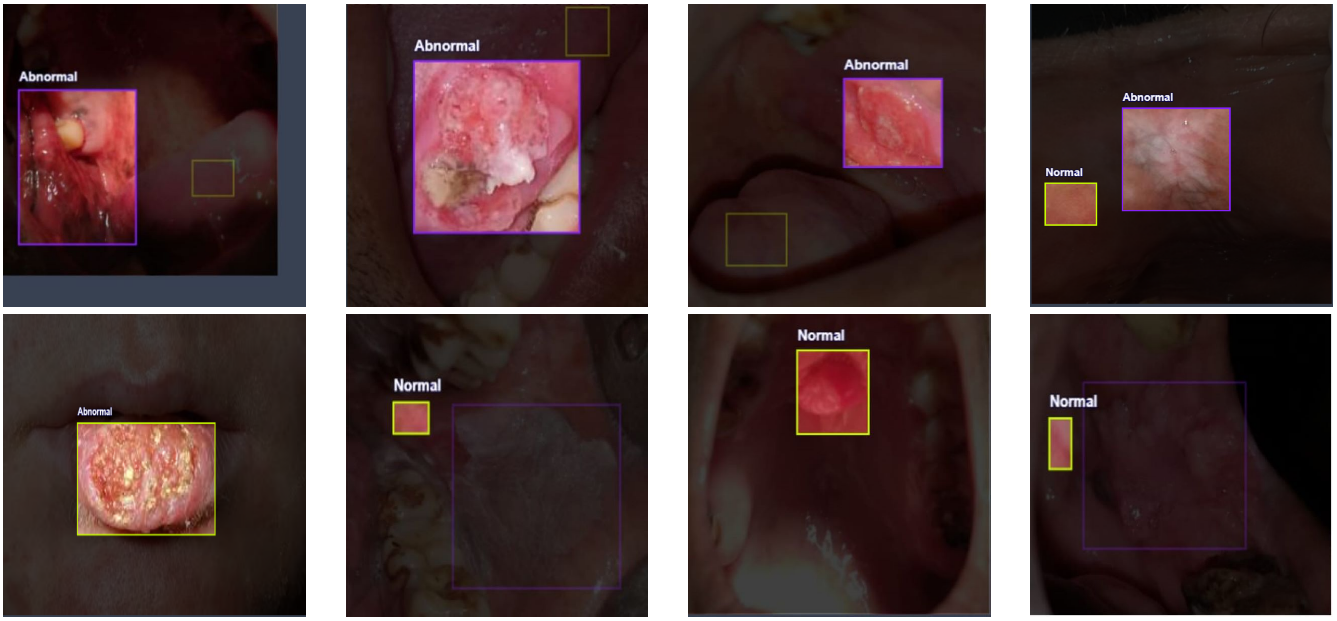


Figure 2. Images annotated with Normal and Abnormal labels

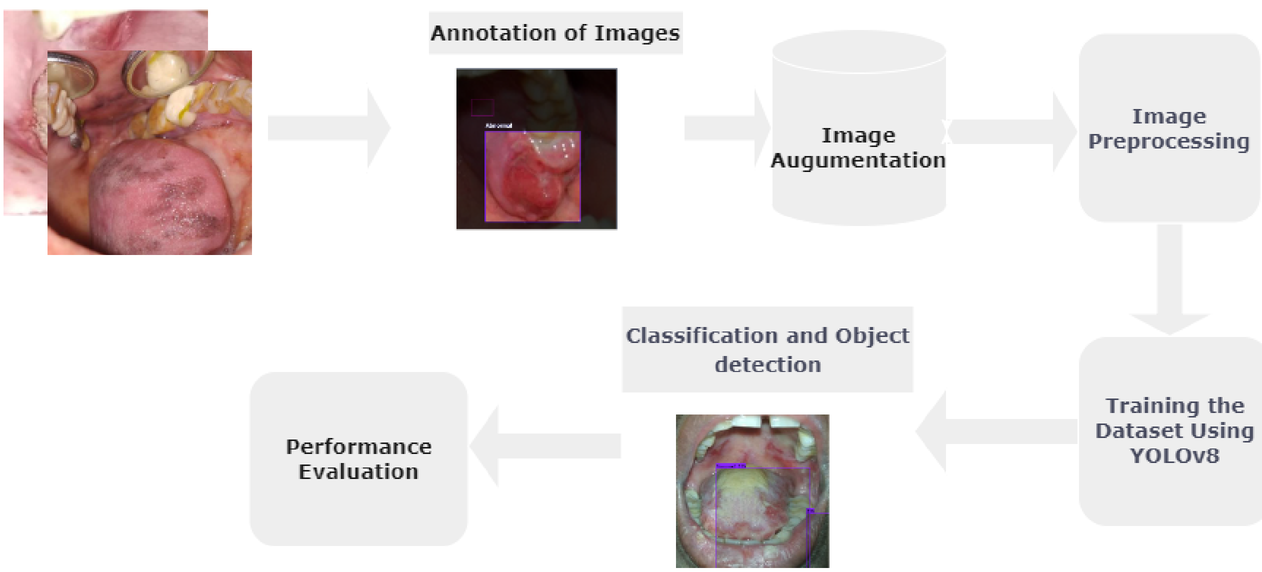


Figure 3. Proposed Design Methodology

background.

$$IoU\ Score = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{2}$$

A loss function helps in adjusting the weights of the neural network to reduce the cost. YOLOv8 employs $C_I oU$ (Complete Intersection over Union) and the DFL (Distributed Focal Loss) for BBox (bounding Box) loss and Binary Cross Entropy(BCE) for Classification (cls) loss. BBox loss measures how closely do the cls loss evaluations and the expected ground truth bounding boxes match how accurately each anticipated bounding box was classified. DFL loss takes care of the class imbalance issue in training the neural network and optimizes the bbox boundary distribution. The $C_I oU$ loss considers three elements: the overlap area, the distance between the center points of the boxes, and the aspect ratio. The loss function for $C_I oU$ is

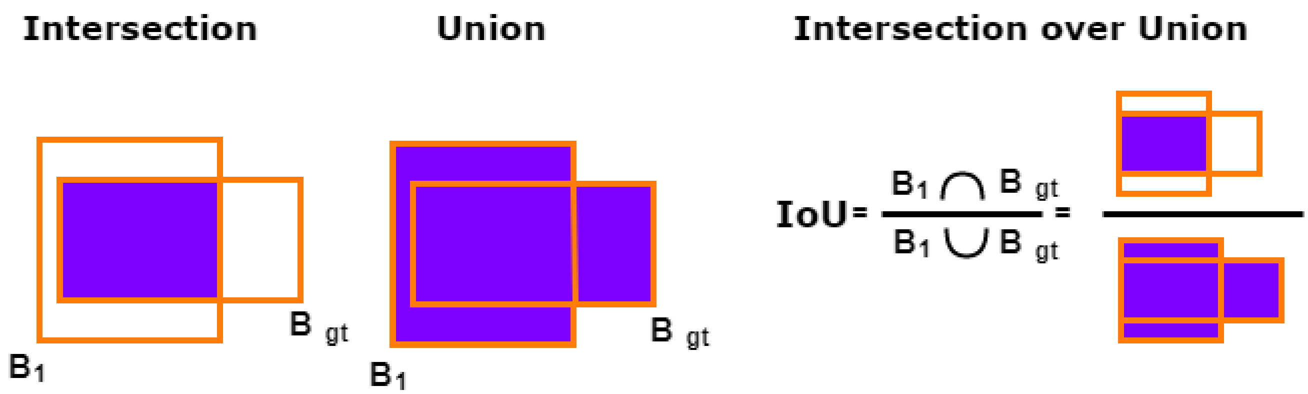


Figure 4. Intersection over Union

given by

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (3)$$

where b , B^{gt}represents the intersection of background with ground truth, ρ is the obtained euclidean distance. The smallest box that encloses the two boxes is represented by c, which is the diagonal measurement. α represents the positive trade-off criterion and v gauges aspect ratio's constancy. α can be given as

$$\alpha = \frac{a}{(1 - IoU) + a} \quad (4)$$

The specified a is

$$a = \frac{4}{\pi^2}(arctan\frac{wt^{gt}}{ht^{gt}} - arctan\frac{wt}{ht})^2 \quad (5)$$

a defines the applying inverse tan function to the the bounding box's height and width.

BCE loss is utilized to measure the difference between the predicted category and the actual labels of the dataset. BCE loss is determined as

$$\mathcal{L}_{BCE} = \frac{1}{A}\sum_{i=1}^{A} -(m_i * log(p_i)) + ((1 - m_i) * log(1 - p_i)) \quad (6)$$

p_i represents probability of Normal class and $(1 - p_i)$ is the probability of class Abnormal. YOLO has originally trained using COCO dataset to identify 80 object classes like cars, books, handbags, phones and so on. This pretrained model has to be retrained recognize normal or abnormal part in an oral cancer images by fine tuning the hyper parameters. The roboflow framework is used to custom train the model for oral cancer images. The model is custom trained using the YOLOv8 model using the train dataset.The classification of an image as Normal and Abnormal is carried out using the test set. The bounded box is used to identify the cancerous part in an image. Yolov8 is crafted to deliver enhanced speed and accuracy, but the performance of the model can be greatly enhanced by fine-tuning the hyperparameters. Table I shows hyper parameters considered for fine tuning the model:

Table I. Hyper parameters for Training the model

Hyper parameters	Value
Image Size	640*640
Batch size	32
Learning rate	1.00E-03
Epochs	25
Object threshold	0.5
NMS Threshold	0.4
Weight decay	0.0001

5. Performance Analysis

In this investigation, the performance metrics considered are mean Average Precision (mAP), precision and recall. The term precision is used to assess how accurate and dependable the experiment's measurement is. Recall measures the model's ability to recognise positive samples. The Precision measure averaged over all classes in a model is referred as mean Average Precision (mAP). This is considered as one of the important performance metric in object detection. Upon obtain the True Positive (TP), False Positive(FP), True Negative (TN) and False Negative (FN) values from the confusion matrix. As determined by the following formula:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (9)$$

TABLE II shows confusion matrix table for the test data and shows the values of precision, recall and mAP for

Table II. Table of the confusion matrix for the test data

Class	Images	Labels	Precision	Recall	mAp@.5	mAP@.5:.95
All	56	76	0.873	0.875	0.913	0.734
Normal	56	29	0.901	0.889	0.972	0.723
Abnormal	56	46	0.812	0.873	0.896	0.768

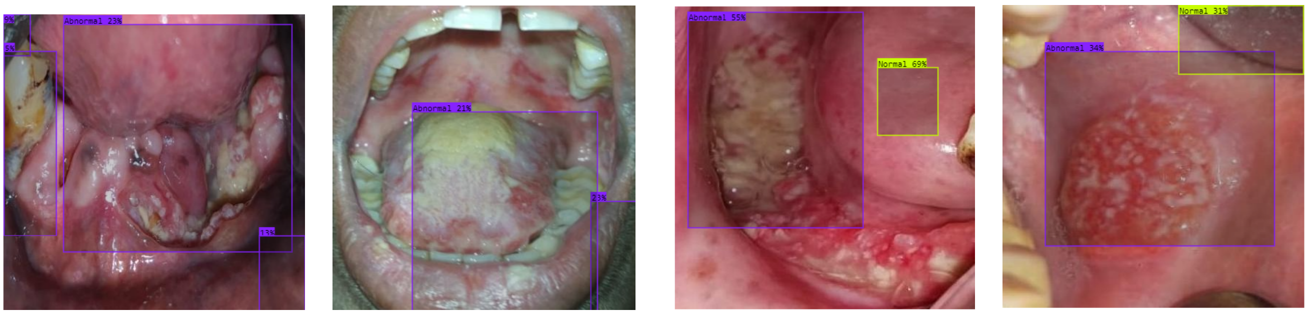


Figure 5. Identification of Normal and Abnormal part in the test images

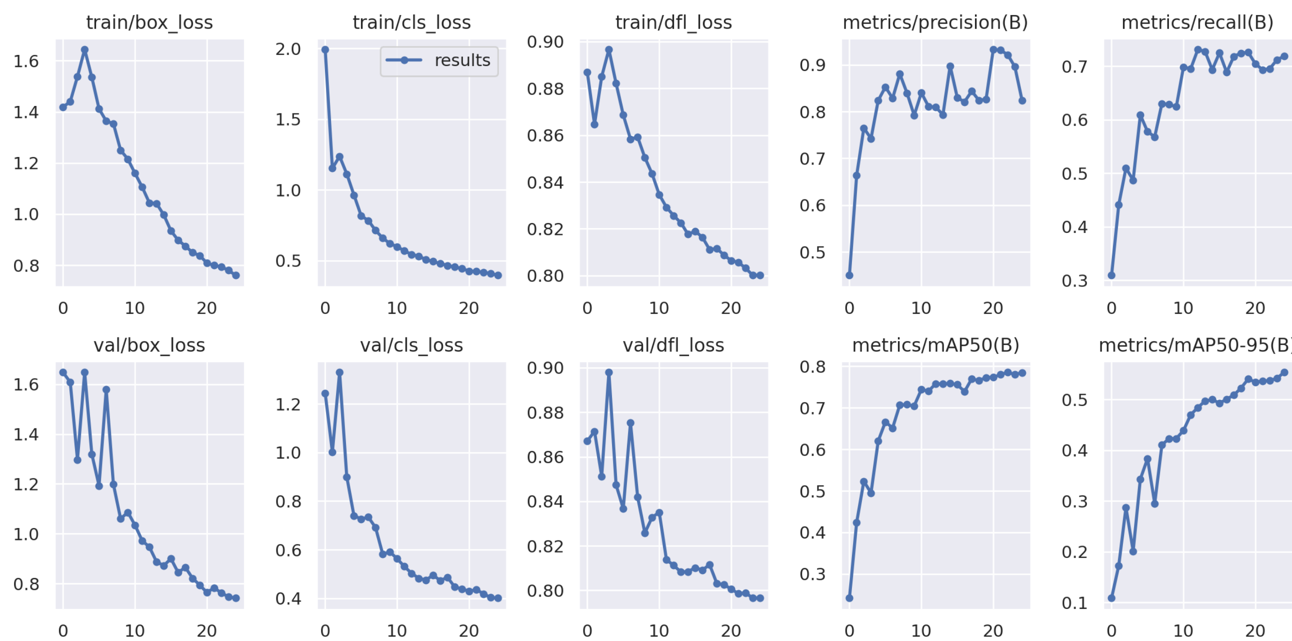


Figure 6. Analysis of metrics for 25 epochs for training and validation

the normal and abnormal regions. Both classes had good and noticeably better mAP ratings for precision and recall. Figure. **??** shows detection of normal and abnormal class with level of confidence on test images. Figure 6 shows the various performance metrics after the model runs for 25 epochs. It has both the training and validation values for reference. The loss is significantly reducing as the training progresses. The precision and recall also improved in both training and validation. Faster RCNN (Region based Convolutional Neural Networks) and Mask RCNN belong to the family of RCNN. RCNN creates region proposals and subsequently processes them through convolutional layers to execute object detection and classification. Yolo uses a one-stage methodology to conduct end-to-end object detection in a single pass. YOLO models outperform RCNN in terms of speed, making them the preferred choice for real-time object detection. RCNN, simultaneously, possesses the capability to achieve higher levels of accuracy. Consequently, a comparative analysis offers a comprehensive overview of the comparison between object detection algorithms, highlighting the performance of YOLOv8 in relation to other leading models in the field. Figure 8 provides the comparative analysis of YOLOv8 with the other state of the art object detection model such as YOLOv7, Faster R-CNN

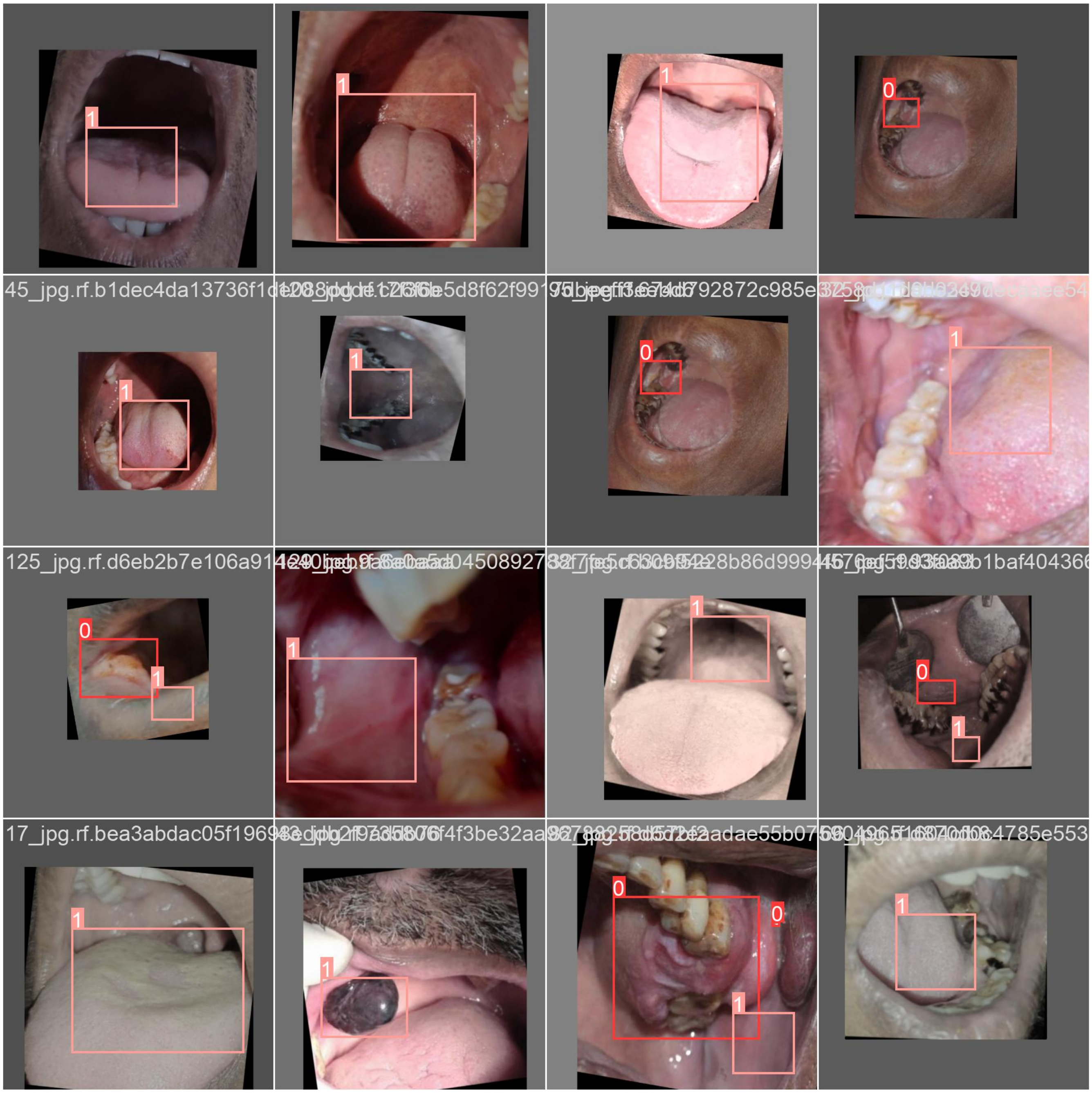
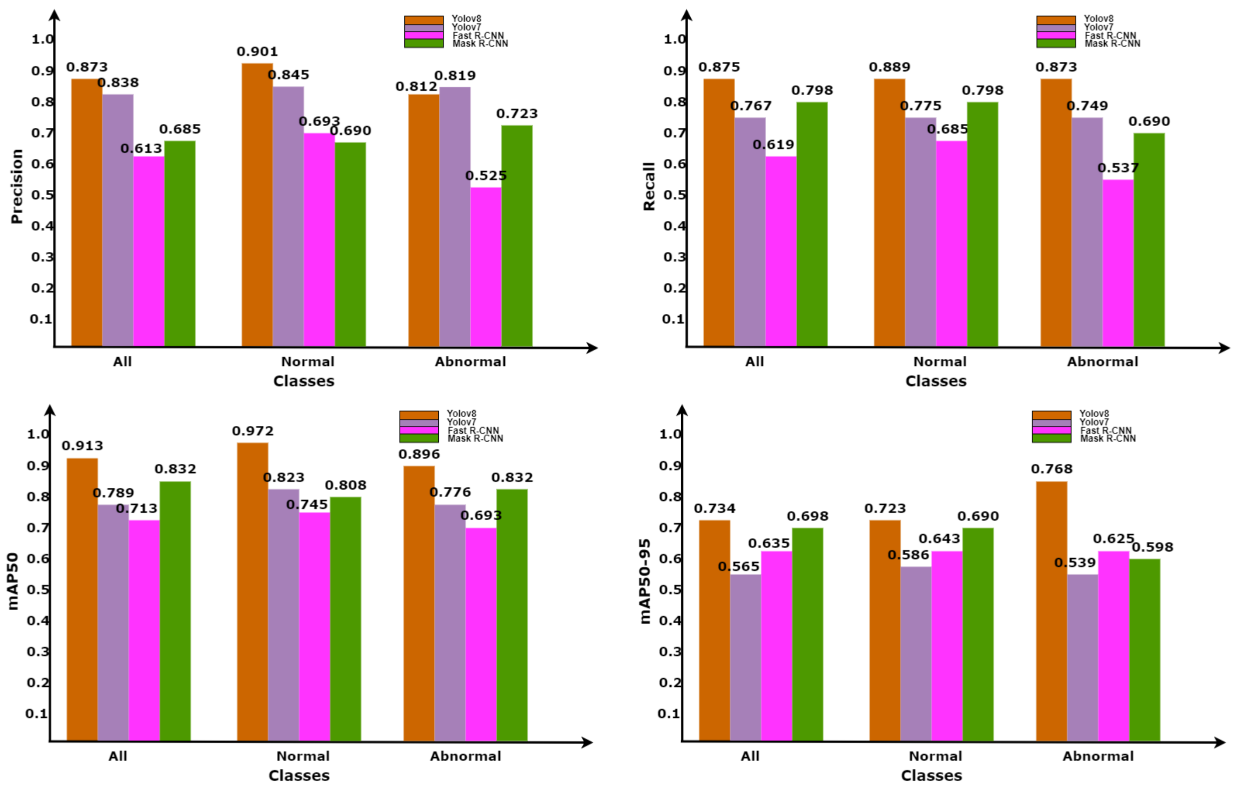


Figure 7. Detection of Normal and Abnormal regions for the Validation dataset

Figure 8. Comparison of YOLOv8 with YOLOv7, Mask R-CNN and Faster R-CNN



and Mask R-CNN. The metrics like precision, recall and mAP is determined for the classes Normal, Abnormal and all. Precision, recall, mAP are all improved by the Yolov8 model, which surpasses the competing model.

6. Conclusion and Future Work

In this work, a sophisticated approach for identifying and classifying oral cancer was created. This system is useful for experts to diagnose oral cancer from the photographic image using deep learning. The roboflow provides a framework to perform the annotation of images. The augmentation of images improved the quality of the dataset. The pretrained YOLOv8 model is used to train 573 images. The model is validated using the test data and it is found that it could able to detect the normal and abnormal part in the given test image with a better confidence level. The system's outcomes show that it is possible to successfully identify and categorise normal and abnormal parts in a picture using an artificial intelligence system that is assisted by the Yolo object detection method. A comparison analysis is provided with the other object detection model with the YOLOv8 is providing the better results. In the future, we intend to generate a data set with approximately 10,000 images that are to be collected from various hospitals to increase the system's resiliency. Additionally, we intend to build a unique model based on the CNN architecture for the identification and classification of oral cancer.

References

[1] R. Serra, C. S. de Oliveira, S. Roque, F. Herrera, and H. Arco, "Oral hygiene care and the management of oral symptoms in patients with cancer in palliative care: a mixed methods systematic review protocol," *JBI Evidence Synthesis*, vol. 22, no. 4, pp. 673–680, 2024.

[2] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024." *CA: a cancer journal for clinicians*, vol. 74, no. 1, 2024.

[3] H. A. Helaly, M. Badawy, and A. Y. Haikal, "A review of deep learning approaches in clinical and healthcare systems based on medical image analysis," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 36 039–36 080, 2024.

[4] A. Vijayakumar and S. Vairavasundaram, "Yolo-based object detection models: A review and its applications," *Multimedia Tools and Applications*, pp. 1–40, 2024.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[6] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia computer science*, vol. 199, pp. 1066–1073, 2022.

[7] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using yolo: Challenges, architectural successors, datasets and applications," *multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[8] L. Gao, T. Xu, M. Liu, J. Jin, L. Peng, X. Zhao, J. Li, M. Yang, S. Li, and S. Liang, "Ai-aided diagnosis of oral x-ray images of periapical films based on deep learning," *Displays*, vol. 82, p. 102649, 2024.

[9] Y. Hsu, C.-Y. Chou, Y.-C. Huang, Y.-C. Liu, Y.-L. Lin, Z.-P. Zhong, J.-K. Liao, J.-C. Lee, H.-Y. Chen, J.-J. Lee *et al.*, "Oral mucosal lesions triage via yolov7 models," *Journal of the Formosan Medical Association*, 2024.

[10] S. Mammeri, M. Amroune, M.-Y. Haouam, I. Bendib, and A. Corrêa Silva, "Early detection and diagnosis of lung cancer using yolo v7, and transfer learning," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 30 965–30 980, 2024.

[11] F. Prinzi, M. Insalaco, A. Orlando, S. Gaglio, and S. Vitabile, "A yolo-based model for breast cancer detection in mammograms," *Cognitive Computation*, vol. 16, no. 1, pp. 107–120, 2024.

[12] C.-K. Chou, R. Karmakar, Y.-M. Tsao, L. W. Jie, A. Mukundan, C.-W. Huang, T.-H. Chen, C.-Y. Ko, and H.-C. Wang, "Evaluation of spectrum-aided visual enhancer (save) in esophageal cancer detection using yolo frameworks," *Diagnostics*, vol. 14, no. 11, p. 1129, 2024.

[13] M. E. Salman, G. Ç. Çakar, J. Azimjonov, M. Kösem, and İ. H. Cedimoğlu, "Automated prostate cancer grading and diagnosis system using deep learning-based yolo object detection algorithm," *Expert Systems with Applications*, vol. 201, p. 117148, 2022.

[14] 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," *Computers in biology and medicine*, vol. 141, p. 105031, 2022.

[15] A. Baccouche, B. Garcia-Zapirain, C. C. Olea, and A. S. Elmaghraby, "Breast lesions detection and classification via yolo-based fusion models." *Computers, Materials & Continua*, vol. 69, no. 1, 2021.

[16] 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," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106884, 2022.

[17] A. Karaman, I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, S. Coskun, O. Sahin, and D. Karaboga, "Robust real-time polyp detection system design based on yolo algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (abc)," *Expert systems with applications*, vol. 221, p. 119741, 2023.

[18] R. Nersisson, T. J. Iyer, A. N. Joseph Raj, and V. Rajangam, "A dermoscopic skin lesion classification technique using yolo-cnn and traditional feature model," *Arabian Journal for Science and Engineering*, vol. 46, no. 10, pp. 9797–9808, 2021.

[19] G. Hamed, M. A. E.-R. Marey, S. E.-S. Amin, and M. F. Tolba, "Deep learning in breast cancer detection and classification," in *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*. Springer, 2020, pp. 322–333.

[20] 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," *Computer methods and programs in biomedicine*, vol. 200, p. 105823, 2021.

[21] Y. Nie, P. Sommella, M. O'Nils, C. Liguori, and J. Lundgren, "Automatic detection of melanoma with yolo deep convolutional neural networks," in *2019 E-Health and Bioengineering Conference (EHB)*. IEEE, 2019, pp. 1–4.

[22] S. O. Manoj, K. R. Abirami, A. Victor, and M. Arya, "Automatic detection and categorization of skin lesions for early diagnosis of skin cancer using yolo-v3-dcnn architecture," *Image Analysis and Stereology*, vol. 42, no. 2, pp. 101–117, 2023.

[23] Z. Ji, J. Zhao, J. Liu, X. Zeng, H. Zhang, X. Zhang, and I. Ganchev, "Elct-yolo: an efficient one-stage model for automatic lung tumor detection based on ct images," *Mathematics*, vol. 11, no. 10, p. 2344, 2023.

[24] V. Patel, M. Kanojia, and V. Nair, "Exploring the potential of resnet50 and yolov8 in improving breast cancer diagnosis: A deep learning perspective," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 16, no. 3, pp. 16–16, 2024.

[25] ultrlytics, ""yolov8 model is used from https://docs.ultralytics.com/."," *NA*, 2023.

[26] ——, "Oral cancer image dataset is used from https://universe.roboflow.com/sagari-vijay/oral-cancer-data/dataset/1." *NA*, 2023.

[27] G. Jocher, A. Stoken, J. Borovec, L. Changyu, A. Hogan, A. Chaurasia, L. Diaconu, F. Ingham, A. Colmagro, H. Ye *et al.*, "ultralytics/yolov5: v4. 0-nn. silu () activations, weights & biases logging, pytorch hub integration," *Zenodo*, 2021.