**Background**

Colorectal cancer (CRC) remains a major health problem, with the latest statistics from the American Cancer Society (ACS) demonstrating its broad prevalence. According to the American Cancer Society's most recent publication, Cancer Statistics, 2024, colorectal cancer is expected to cause 152,810 new cases in 2024, with males accounting for 81,540 and women for 71,270. Of them, 106,590 will be categorized as colon cancer, and 46,220 as rectal cancer. CRC is the second highest cause of cancer-related fatalities in the United States, with a projected 53,010 deaths in 2024, up slightly from 52,550 deaths the previous year[1].

Traditionally, colorectal cancer (CRC) is diagnosed using screening methods such as colonoscopy, sigmoidoscopy, and imaging tests like CT colonography. Colonoscopy, considered the gold standard for CRC detection, allows for direct visualization of the colon’s interior and facilitates the removal of polyps for biopsy. However, the complexity of tissue structures and the potential for human error in identifying and classifying abnormal cells through histopathological analysis present ongoing challenges. These constraints often lead to diagnostic delays and variability in accuracy, highlighting the need for more reliable and faster diagnostic approaches, particularly as CRC in younger populations tends to be diagnosed at more advanced stages [2].

The increasing complexity and volume of medical data have further fuelled the demand for more efficient diagnostic tools in healthcare. In response, deep learning and artificial intelligence (AI) have emerged as transformative technologies capable of automating the diagnostic process and significantly improving cancer detection accuracy. AI-driven models, particularly those using deep learning frameworks, can process vast amounts of medical imaging data to identify malignant tissues with greater precision, reducing human error and accelerating diagnosis times. These advancements are especially crucial in complex diseases like cancer, where early detection is key to successful treatment and better patient outcomes. By integrating AI into CRC diagnostics, healthcare systems can achieve more consistent and accurate results, improving early detection and ultimately enhancing patient care [3-4]. This section discusses deep learning applications in CRC detection, the specific benefits of the YOLOv8 model, the significance of interpretability in AI-driven healthcare systems, and the performance measures used to assess these models.

**2.1 Automated Diagnosis with Deep Learning Models**

The increasing complexity of medical data, particularly in medical imaging, has necessitated the development of automated technologies to improve diagnostic accuracy and speed. Colorectal cancer (CRC), one of the most frequent and deadly tumors in the world, benefits substantially from the use of artificial intelligence (AI) and deep learning algorithms. These technologies can automate the diagnostic procedure, eliminating the difficulties and constraints of manual histopathology investigation.

Traditionally, pathologists used histopathology slides to identify malignant tissues, which was time-consuming and prone to human error. This reliance on manual interpretation frequently causes diagnostic variability and delays. Deep learning models, including Convolutional Neural Networks (CNNs), have emerged as effective methods for automating these tasks. CNNs and similar AI models have demonstrated enormous promise in diagnosing malignant tissues from histopathology photos with greater accuracy and consistency than human pathologists. Studies have shown that AI systems can dramatically enhance cancer detection, which is crucial for better patient outcomes in CRC cases [5]. AI-powered models improve diagnostic accuracy while simultaneously shortening diagnosis time. This is especially significant in cases of CRC, where early identification is critical for effective therapy. Automated diagnostic systems powered by AI can scan vast amounts of medical data, recognize trends, and diagnose anomalies more quickly and efficiently than traditional approaches. Furthermore, these methods help reduce diagnostic variability by offering standardized assessments of histological pictures [7].

The incorporation of AI into CRC diagnostics is revolutionizing the healthcare environment by giving doctors tools that improve diagnostic accuracy and efficiency. As technology advances, it has the potential to transform cancer diagnosis by offering faster, more reliable, and more accessible detection approaches. While numerous AI models have been used to diagnose CRC, object detection and classification models, such as YOLOv8, have emerged as critical tools for automated CRC classification, notably in recognizing and classifying malignant spots in histopathology images. The shift from general deep learning models to more complex object detection frameworks represents a significant advancement in the automation of CRC diagnoses[7].

**2.2 YOLOv8 Model for CRC Classification**

Building on the automation of CRC diagnosis outlined in Subsection 2.1, YOLOv8 (You Only Look Once, version 8) is a cutting-edge object detection model that is essential for detecting and categorizing CRC in medical imaging. Unlike standard CNNs, which focus on picture classification, YOLOv8 is intended for real-time object recognition, making it ideal for detecting specific malignant spots in high-resolution histopathology images. This capacity enables more exact localization and classification of tumor tissues, hence increasing the overall accuracy and speed of CRC diagnosis. YOLOv8 was chosen for CRC classification because of its capacity to effectively analyze high-resolution histopathological pictures, detect several regions of interest, and accurately identify malignant tissues.

YOLOv8 provides significant enhancements over its predecessors, including an anchor-free detection mechanism and strengthened convolutional layers, which enable the precise recognition of small and irregular malignant tumors even in complex histological pictures. This feature is crucial for CRC classification. Early diagnosis of tiny lesions can considerably improve patient outcomes by enabling timely therapy [8].

The capability to process vast amounts of high-resolution data in real time makes YOLOv8 an excellent choice for CRC classification. Its ability to handle massive medical datasets allows for rapid tumor detection and classification, decreasing diagnostic delays. YOLOv8's improved architecture enables accurate localization and classification of various tissue types in the colon and rectum.

YOLOv8 is distinguished from other deep learning models such as Convolutional Neural Networks (CNNs) and Vision Transformers by its ability to identify in real time. While CNNs are effective in picture classification, they frequently lack the object identification and real-time processing features that YOLOv8 provides. Although powerful for some picture classification tasks, vision transformers often require greater computer resources and longer processing periods, making them less viable for fast, real-time applications like CRC detection[9].

YOLOv8 has significant advantages in colorectal cancer (CRC) classification, including the ability to detect and categorize small, irregular tumor areas with high accuracy. It also has faster training and inference times than other models, making it ideal for real-time applications. Furthermore, its architecture is designed to easily handle huge, high-resolution histopathology images, which are essential for precise medical diagnosis. However, these benefits are not without costs. YOLOv8 takes significant computational resources, particularly during the training phase, which may limit its usability in environments with less advanced infrastructure. Furthermore, the model's success is heavily reliant on the availability of big, annotated datasets, which are notoriously difficult to get in medical applications. Despite these challenges, YOLOv8's ability to balance speed, accuracy, and efficiency makes it an excellent alternative for CRC classification, as it overcomes many of the constraints of manual and slow diagnostic methods by providing exact, real-time detection of malignant regions.

**2.3 Interpretability with EigenCAM**

As artificial intelligence (AI) models become more crucial to medical diagnosis, the demand for interpretability has increased dramatically. Understanding how AI models make judgments is critical in high-stakes domains such as colorectal cancer (CRC) detection so that clinicians can trust and act on the model's output. Interpretability tools, such as EigenCAM, offer visual insights into the decision-making process, allowing healthcare practitioners to more easily test and evaluate model predictions. EigenCAM was used for this investigation since it is simple and effective. It creates heatmaps that visually highlight areas of an image that are most important to a model's classification conclusion, providing intuitive feedback on the model's attention during diagnosis. These heatmaps are particularly effective in histological pictures for identifying tumour areas, which are crucial in CRC diagnosis. This transparency allows physicians to understand which areas the model thinks significant, which increases trust in the AI's conclusions[10].

*Pros and Cons of EigenCAM*: EigenCAM works perfectly with current deep learning models and requires no architectural changes or retraining, making it simple to apply. Furthermore, the visual explanations it provides are simple and basic, making them easier for medical experts to interpret. However, EigenCAM has less granularity than more advanced techniques such as GradCAM, which can provide layer-wise insights into model decisions. Furthermore, EigenCAM can be computationally expensive, especially when dealing with large, high-resolution medical images, affecting real-time applications [11-12].

*Comparison with other tools*: Compared to GradCAM and LIME, EigenCAM strikes a balance between usability and interpretability. GradCAM gives more extensive, layer-by-layer explanations, making it suited for in-depth understanding of a model's activities, but it takes more computer capacity. LIME, while model-agnostic and versatile, is less intuitive for deep learning models and may be slower in medical applications with huge datasets. EigenCAM's simplicity and effective visual feedback make it ideal for CRC classification, increasing model transparency and trust in AI-powered diagnostics[13].

**2.4 Performance Evaluation Metrics**

The YOLOv8 model's efficacy for colorectal cancer (CRC) diagnosis is evaluated using a combination of standard and advanced measures that quantify its effectiveness across multiple aspects. In medical applications like CRC diagnosis, where datasets are frequently unbalanced, relying entirely on standard measures might be deceptive. As a result, enhanced metrics enable a more comprehensive examination[13].

*Standard Metrics: Accuracy, Precision, Recall, and Specificity*: Accuracy, precision, recall, and specificity are frequently used to assess model performance. Accuracy examines the proportion of correct forecasts, whereas precision evaluates the proportion of real positive results among all positive predictions. Recall examines the model's capacity to detect all true positives, whereas specificity assesses how well the model identifies non-cancerous regions. While metrics can provide significant insights, they may not accurately represent performance when dealing with imbalanced datasets.

*Advanced Metrics*: Intersection over Union (IoU), Balanced Accuracy (BAC), and Matthews Correlation Coefficient (MCC): further advanced metrics, such as Intersection over Union (IoU), Balanced Accuracy (BAC), and Matthews Correlation Coefficient (MCC), provide further insights when evaluating models on imbalanced datasets. IoU quantifies the overlap between predicted bounding boxes and ground truth, which is critical for object detection tasks such as CRC classification. BAC balances recall across all classes, minimizing biases towards the dominant class. MCC gives a thorough evaluation, especially beneficial in imbalanced environments.

*Comparison of Metrics*: Standard metrics are beneficial, but they are less trustworthy when dealing with imbalanced datasets, which are typical in CRC diagnoses. Advanced measures like MCC and BAC are more suited for such scenarios because they provide a more balanced perspective of model performance, allowing for correct evaluation of minority classes like malignant regions. These measurements provide a comprehensive evaluation of the YOLOv8 model, ensuring efficiency and reliability in medical diagnostics.

[1] Siegel, R. L., Giaquinto, A. N., & Jemal, A. (2024). Cancer statistics, 2024. CA: a cancer journal for clinicians, 74(1).

*[2] Davri, A., Birbas, E., Kanavos, T., Ntritsos, G., Giannakeas, N., Tzallas, A. T., & Batistatou, A. (2022). Deep learning on histopathological images for colorectal cancer diagnosis: a systematic review. Diagnostics, 12(4), 837.*

[3] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine, 13, 1-17.

[4] Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep learning for medical image-based cancer diagnosis. Cancers, 15(14), 3608.

[5] Mitsala, A., Tsalikidis, C., Pitiakoudis, M., Simopoulos, C., & Tsaroucha, A. K. (2021). Artificial intelligence in colorectal cancer screening, diagnosis and treatment. A new era. Current Oncology, 28(3), 1581-1607.

[6] Uchikov, Petar, Usman Khalid, Krasimir Kraev, Bozhidar Hristov, Maria Kraeva, Tihomir Tenchev, Dzhevdet Chakarov et al. "Artificial intelligence in the diagnosis of colorectal cancer: a literature review." Diagnostics 14, no. 5 (2024): 528.

[7] Yang, X., Song, E., Ma, G., Zhu, Y., Yu, D., Ding, B., & Wang, X. (2023). YOLO-OB: An improved anchor-free real-time multiscale colon polyp detector in colonoscopy. arXiv preprint arXiv:2312.08628.

[8] Terven, Juan, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas." Machine Learning and Knowledge Extraction 5, no. 4 (2023): 1680-1716.

[9] Sohan, Mupparaju, Thotakura Sai Ram, Rami Reddy, and Ch Venkata. "A review on yolov8 and its advancements." In *International Conference on Data Intelligence and Cognitive Informatics*, pp. 529-545. Springer, Singapore, 2024.

[10] Bany Muhammad, M., & Yeasin, M. (2021). Eigen-CAM: Visual explanations for deep convolutional neural networks. SN Computer Science, 2(1), 47.

[11] Muhammad, Mohammed & Yeasin, M.. (2020). Eigen-CAM: Class Activation Map using Principal Components. 1-7. 10.1109/IJCNN48605.2020.9206626.

[12] <https://www.datature.io/blog/understanding-your-yolov8-model-with-eigen-cam>

[13] Kakogeorgiou, I., & Karantzalos, K. (2021). Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *103*, 102520.

[13] <https://yolov8.org/unveiling-the-secrets-of-yolov8-metrics/>

**2. Related Work**

**2.1 Deep Learning Models (YOLOv8, Transformers, and DNN) in Colorectal Cancer Diagnosis**

This section reviews the use of deep learning models like YOLOv8, Vision Transformers, and Deep Neural Networks (DNNs) in colorectal cancer (CRC) diagnosis. It will look at how these models have been applied to medical imaging, namely histology, to improve CRC detection and classification. The segment will highlight the effectiveness of these models in recognizing malignant tissues, as well as their progress in improving the accuracy and speed of CRC diagnosis in medical research.

**2.1.1 YOLOv8 in Medical Imaging**:

The first YOLO was presented by Joseph Redmon et al. at CVPR 2016. The YOLO acronym stands for "You Only Look Once," and it refers to the ability to look at an image as if it were a human and instantly recognize the object, what they are doing, and where they are. YOLO, a regression-based object recognition system, outperformed typical deep learning-based models in terms of accuracy and speed. YOLO is a powerful one-stage object identification system that has grown over time and produced multiple variants. The YOLO algorithm divides the input image into 𝑆×𝑆 grids and forecasts B bounding boxes of the same class, as well as the confidence of each grid for 𝐶 other classes, instead of the standard sliding window technique. Each bounding box forecasts five values: (𝑥,𝑦,𝑤,ℎ,𝑐), representing its position, size, and confidence, accordingly. Each grid predicts (𝐵×5+𝐶) values and uses Non-Maximum Suppression (NMS) to remove duplicate detections.

2.1.2 **YOLOv8**

The YOLOv5 team launched YOLOv8[6] in January 2023, however an official article is still to be published. The primary improvements are as follows: The backbone of YOLOv8 is CSPDraknet-53. As with YOLOv5, YOLOv8's C3 module is replaced by the C2f module with richer gradient flow, and different channel counts are adjusted for different scale models to achieve even greater lightweight. Furthermore, YOLOv8 retains the SPPF module used in YOLOv5. Head: The Head section contains two notable upgrades over YOLOv5. First, it is replaced with the current standard Decoupled-Head algorithm, which separates classification and detection. Second, it has been modified from anchor-based to anchor-free. Loss: YOLOv8 abandons the previously used IOU matching or unilateral ratio distribution methods in favor of the Task-Aligned Assigner positive and negative sample matching method. Furthermore, YOLOv8 introduces Distribution Focal Loss (DFL) [35]. Data augmentation can increase model performance, but using mosaic augmentation during training may have negative consequences. YOLOv8 disables Mosaic augmentation in the latest 10 epochs, improving accuracy. To suit the needs of various scenarios, YOLOv8 offers a variety of size models in N, S, M, L, and X scales.

Several target identification and classification methods based on YOLOv8 are employed in medical picture diagnosis. [42] employed YOLOv8 for binary categorization of breast cancer histopathology images from the BreakHis dataset. [43] proposed a method for breast cancer segmentation utilizing YOLOv8x-seg, a customized variation of the YOLO model suited for semantic segmentation. [44] presented a polyp detection method based on artificial intelligence and the YOLO-V8 network. This strikes a compromise between accuracy and computational efficiency. [45] proposed a breast cancer pathological picture classification approach based on wavelet transform and YOLOv8. [46] provided a strategy for detecting breast cancer from ultrasound images. The models used are YOLOv6, YOLOv7, and YOLOv8. [47] proposed using deep learning in medical imaging to study the usefulness of the YOLOv8 algorithm for diagnosing several cancer types, including Acute Lymphoblastic Leukemia, Cervical, Lung, Colon, Oral, and Skin cancers. The YOLOv8 method, known for its real-time object detection capabilities, is a good contender for automating the detection and classification of malignant spots in medical pictures.

Several target detection methods based on the YOLO series have been designed to improve polyp detection accuracy. Guo et al. introduced an automatic polyp detection system that uses the YOLO-V3 structure in conjunction with active learning to reduce false positive rates in polyp detection [29]. Cao et al. combine the feature extraction and fusion module with the YOLO-V3 network to extract both high- and low-level feature maps [30]. Pacal et al. proposed a real-time automatic polyp identification approach using YOLO-V4. They added the cspnet network into the architecture, as well as the mish activation function, Diou loss function, and transformer block, to achieve improved accuracy and performance [31]. These advances in target detection algorithms based on the YOLO series show promise for more efficient and precise polyp detection during colonoscopy. Lee and colleagues [32] developed a real-time system for polyp detection based on YOLO-V4. The technique used a multiscale mesh to identify tiny polyps. Performance improvements were accomplished by incorporating advanced data augmentation techniques and using various activation functions. Wan and colleagues [12] proposed a YOLO-V5-based model for real-time polyp detection that included a self-attention mechanism. This strategy strengthens good traits while weakening less relevant ones, resulting in improved polyp identification performance. Pacal et al. [33] used the Scaled YOLO-V4 method to evaluate novel datasets, SUN and PICCOLO. Durak and colleagues [34] trained on cutting-edge object identification algorithms, including YOLO-V4 [35], CenterNet, EfficientDet [36], and YOLO-V3 [37], for automatic gastric polyp diagnosis. Qian et al. [38] proposed a method for detecting polyps that combines GAN structures and the YOLO-V4 object detection algorithm. Gabriel implements YOLO-V4 at multiple precision levels (FP32, FP16, and INT8) for polyp detection. [39]. In [40], Ahmet Karaman and Ishak Pacal provide a pioneering integration of the ABC method with YOLO-based object identification techniques, with a focus on improving activation functions and hyperparameters—an unprecedented exploration in the literature. The suggested method is adaptable and may be applied to any dataset or YOLO-based algorithm, with parameters tailored to enhance performance [41].

**2.1.3 Transformers in CRC Detection**:

In [39], Zeid et al. employed Vision Transformers to perform multiclass tissue classification on a colorectal cancer histology image dataset. The authors trained two Transformer variants: the Vision Transformer model and the Compact Convolutional Transformer model. Transformers, on the other hand, typically require a considerable amount of training data and computer resources to reach optimal performance, which can be a challenge when working with smaller medical datasets. In another work [40], Yiqing Shen et al. proposed a method named MoViT on a small dataset, achieving virtually full-dataset performance with only 1%-3% of the data. This methodology not only outperformed other methods, but it also worked with smaller datasets and substantially shorter training times each epoch. However, relying on such a small portion of the dataset raises concerns about the model's generalizability to new data. In addition, the limited training samples may pose a performance risk due to overfitting. Zhilong et al. [41] described a unique technique to colorectal cancer survival analysis that combined histopathological images and genomic data. They employed the NCT biobank dataset with a transformer-based model called TransSurv. This integrated strategy intended to improve the forecast accuracy of patient survival rates using visual and genomic data, reaching a Concordance value (C-index) of 0.822. In [42], the researchers suggested a unique strategy that combined the strengths of Convolutional Neural Networks and Transformers for pathological picture categorization. This method employed CNNs for local feature extraction and Transformers for extracting global contextual information from picture patches in histopathology images. The model was tested using two datasets: PCam and NCT-CRC. In this investigation, the class token layer was removed in favor of a sequence pooling module. The findings show an improvement over current state-of-the-art approaches, emphasizing the possibility of integrating CNNs and Transformers in medical picture processing. However, the removal of the class token layer and the addition of a sequence pooling module are novel, but they add layers of complexity that may affect model training and interpretability. The authors of [43] proposed an enhanced method for categorizing colorectal cancer histopathology pictures that uses the ResNet-50 model as well as transfer learning and fine-tuning techniques. Using the NCT-CRC-HE-100K and CRC-VAL-HE-7K datasets for training and validation, respectively. This achievement highlights the model's effectiveness in CRC image categorization, which is a major improvement over earlier studies. However, this model did not do well on the external dataset. In [44], the proposed ''TransNetV'' model uses the local feature extraction capabilities of CNNs and then serially passes these features via the global contextual understanding of the Transformer's attention mechanism. This methodology takes advantage of CNNs' weight-sharing capabilities and Transformers' capacity to understand the broader context of spatial relationships in large-scale patterns, making it applicable to datasets of varied sizes and sophisticated feature analysis. [45] suggested a colorectal cancer detection network (CCDNet) that incorporates coordinate attention transformer and atrous convolution. CCDNet first denoises the input histopathology image with a Wiener-based Midpoint weighted non-local means filter (WMW-NLM) to ensure accurate diagnoses while preserving image features. In addition, a novel atrous convolution with coordinate attention transformer (AConvCAT) is presented, which successfully combines the advantages of two networks to categorize colorectal tissue at different scales by gathering local and global information. Furthermore, the coordinate attention model is used with a Cross-shaped window (CrSWin) transformer to capture small changes in colorectal tissue from numerous viewpoints. The study [46] compared the performance of models ViT-B/16, ViT-B/32, and ViTL/16 with YOLv8n-cls, YOLOv8s-cls, and YOLOv8m-cls, respectively. While all of the data show that the YOLOv8 model outperformed the Vision Transformer (ViT) model in both the training and testing sets. This performance disparity could be due to a variety of variables. Overfitting is one possible cause for the ViT model's problems. This problem arises when the model learns specific patterns, including noise and superfluous features, in the training data extremely well. As a result, the model performs poorly on unlabeled data because it lacks the generalizable characteristics required for accurate classification. The second reason ViT's performance is inferior than YOLOv8 is that its hyperparameter configuration is not ideal. The YOLOv8 models utilized in the study had pretrained hyperparameters that were optimized for a variety of training conditions. However, because to their inherent complexity, ViT models may rely more on hyperparameter tweaks than YOLOv8 models do. This means that little hyperparameter changes can have a big impact on the performance of ViT models. Another factor contributing to YOLO's higher efficiency over ViT is its considerable complexity. Transformer-based models are primarily intended for use in Natural Language Processing (NLP) activities, although their complicated architecture may present a difficulty in computer vision applications. Because of the great complexity of the models, a large dataset for training is typically required to achieve optimal accuracy. In contrast, modestly sized training datasets can assist low-complexity YOLOv8 models in improving accuracy.

**2.1.4 Deep Neural Networks (DNN)**:

CNN models were created and tested on the Kather colorectal cancer histology dataset to categorize and grade colorectal cancer tissues [14]. Many researchers have been able to assess the efficacy of various models and identify viable ways for improving the accuracy and dependability of CRC detection. Janowczyk and Madabhushi16 used the Kather colorectal cancer dataset to train and test multiple CNN models, including AlexNet [17], VGGNet18, and GoogLeNet19. Cruz-Roa et al. [20] also used the CRC dataset to train and test a CNN model for identifying invasive ductal carcinoma in breast cancer histology images. Furthermore, Wang et al. [21] used the CRC dataset to train an attention-based CNN model for grading colorectal cancer histology images.

In [22], Rizalputri et al. used multiple approaches to examine the Kather colon histological dataset, including K-Nearest Neighbor [23], CNN, Random Forest, and Logistic Regression. They wanted to assess the efficacy of each technique and choose the best algorithm. Based on their data, CNN was found to be the most effective strategy. In [24], they proposed a technique that uses Kather colorectal cancer histological assessment to automatically identify eight tissues observed in CRC examination. Furthermore, it was implemented to transfer knowledge based on CNN architectures by them. CNNs' structures have been adapted to extract information from images and feed it into machine learning algorithms including k-Nearest Neighbours, Multilayer Perceptrons, Random Forest, Naive Bayes, and Support Vector Machines (SVM). DenseNet169 had the best performance using SVM (RBF). In [25], we developed a strategy for determining the enormous geometric variation of histological images by extracting four local characteristics: local architectural information, local geometric information, local energetic information, and local patterns. These features are derived from the Riesz transform and monogenic local binary patterns. They used their approach to two multiclass histology image datasets (Kather and Kimiapath [24]).

Several studies have focused on the application of transfer learning in CNN models. For example, in reference [26], a deep learning model including transfer learning and attention mechanisms was developed to estimate electromyography hand movements. This model consists of a feature extraction system with focus modules for extracting relevant features, a three-layer fully interconnected label classifier, and a gesture estimator that uses a threshold voting algorithm to provide gesture estimation results before the hand motion is completed. . The featured research in [27], the optimization framework, provides a brief introduction to the recently created concept of TL methods for a collection of functions. The authors emphasize the need of addressing the topic of where to transfer in addition to the three critical questions of what, how, and when to relocate. It is mentioned and clarified that the first step in transfer learning is to consider the question, "From where to transfer?" In [28], the authors proposed a method for categorizing CRC tissues using multispectral HI. They discovered three types of tissue associated with CRC varieties: benign hyperplasia (BH), intraepithelial neoplasia (IN), and carcinoma. The authors of [29] proposed a reliable CAD technique for metastatic lymph nodes (LNM) in CRC by combining HI analysis and feature assessment. To distinguish CRC tissue from various data sets, the researchers developed a deep learning model based on the CNN structure.

The real-time object detection capabilities of YOLOv8 set it apart from other deep learning models such as Convolutional Neural Networks (CNNs) and Vision Transformers. Ultralytics has made significant improvements to YOLOv8, making it more effective and user-friendly than YOLOv5. It is an enhanced model that builds on the success of YOLOv5 by making changes that increase its power and usability in a variety of computer vision tasks. The improvements include a changed backbone network, an anchor-free detecting head, and a revised loss function. Furthermore, it includes built-in capability for image classification. YOLOv8 is unique in that it provides unrivaled speed and accuracy performance while keeping a simplified design that makes it ideal for a wide range of applications and easily adaptable to diverse hardware platforms.

**2.1.5 YOLOv8 Architectural Advancements**

*Anchor-Free Detection*: Like YOLOv6 and YOLOv7, YOLOv8 is a model that does not rely on anchors. This means that it forecasts an object's center directly rather than calculating the offset from a known anchor box. Anchor boxes were a well-known challenge for early YOLO models (YOLOv5 and before), as they could capture the target benchmark's box distribution but not the distribution of the custom dataset. The use of anchor-free detection reduces the amount of box predictions, which accelerates Non-Maximum Suppression (NMS), a sophisticated post-processing phase that sorts through potential detections after inference. [30]

*New Convolution layer:* The YOLO architecture's convolutional (conv) layers recognize features in input images using learnable filters. These layers detect features at multiple scales and resolutions, allowing the network to identify objects of various sizes and forms. The output of these layers is then sent into further layers, which provide bounding boxes and class predictions for each object detected in the image. Unlike YOLOv5, YOLOv8 employs a separate convolution layer known as C2f. This new layer replaces the C3 layer from YOLOv5. The C2f layer in YOLOv8 concatenates the outputs of all the Bottleneck layers, whereas YOLOv5's C3 layer just uses the output of the last Bottleneck layer. In the system's neck, the features are concatenated without requiring the same channel dimensions, reducing the number of parameters and total tensor size. YOLOv8's improved accuracy and speed can help medical practitioners make faster and more accurate diagnoses, resulting in better patient outcomes. Overall, YOLO v8 has tremendous potential as an object detection model that can improve real-time detection capabilities.

**2.2 Interpretability Tools in Colorectal Cancer Diagnosis**

This subsection examines at the relevance of interpretability tools in AI-driven colorectal cancer (CRC) detection, highlighting the significance of knowing how AI models create predictions. It will concentrate on technologies like EigenCAM, demonstrating how they provide visual insights into model decisions by emphasizing crucial locations in medical imaging, such as tumor areas. The segment will also compare EigenCAM to other interpretability approaches, such as GradCAM and LIME, analyzing their benefits and drawbacks in the context of CRC classification, as well as their impact on clinical trust and model validity.

**2.2.1 Usage of Eigen-CAM in Medical imaging:**

In [73], they developed Eigen-CAM, which calculates and visualizes the principal components of learned features/representations from convolutional layers. Empirical studies were conducted to compare Eigen-CAM against state-of-the-art algorithms (such as Grad-CAM, Grad-CAM++, and CNNfixations) using benchmark datasets such as weakly supervised localization and object localization in the presence of adversary noise. Eigen-CAM has been shown to be robust to classification mistakes caused by fully connected layers in CNNs, and it does not rely on gradient backpropagation, class relevance score, maximal activation locations, or any other kind of feature weighting. Furthermore, it is compatible with all CNN models and does not require any layer modifications or model retraining.

Gradient-weighted Class Activation Mapping (Grad-CAM) uses the gradients of any target concept (say, 'dog' in a classification network or a sequence of words in a captioning network) flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept (77).

[78] compared several Yolo architectures, including YoloV3, YoloV5, and YoloV5-Transformer. In addition, Eigen-CAM was used for model introspection and explanation, showing all worrisome regions of concern within the mammography.   
In [79], Eigen-CAM was used with YOLOv8 for alignment-free bacterium identification using optical scattering to obtain insight into the model's decision-making processes during prediction.   
In [80], the Eigen-CAM approach is shown with pictures. This investigation showed that the model used the position of the distal femur and proximal tibia to detect the probable fracture site and then determine whether a Segond fracture existed in that area.

According to [73] Eigen-CAM is robust and reliable in producing consistent visual explanations and outperforms state-of-the-art methods. The easy to use and intuitive Eigen-CAM only needs the learned representations at the final convolution layer, making it independent of classification layers. Eigen-CAM can be used with any CNN-based DL models without any modification. The EigenCAM [81] is a simple and efficient method for calculating the principle components of the learned convolutional activations, but it is not class-aware.

**2.2.2 CAM (Class Activation Map) and comparison between different interpretability tools:**

CAM (Class Activation Map) is a technique used in convolutional neural networks (CNNs) to determine which portions of an image are relevant for classification. Visual explanations of CNN are expected to provide a high resolution class discriminative interpretation for a variety of purposes. A variety of approaches, including CAM [19], Grad-CAM [20], Grad-CAM++ [21], and CNN-fixations [22], have produced results of varied degrees of success. The Eigen-CAM creates visual explanations by using the principal components of the learnt representations from the convolutional layers. Eigen-CAM improves on this approach by including a visual explanation for each class prediction. The numbers show the importance of certain locations in the image for the appropriate class prediction. These values represent how each region contributes to the classification decision. Higher levels often represent more important contributions. Understanding these heat maps helps comprehend model decisions, especially in medical diagnostics, where exact localization of diseased characteristics is critical. By analyzing Eigen-CAM outputs across grades and forecasts, physicians can obtain insight into the model's reasoning process and potentially find patterns indicative of disease progression or severity. This visual explanation increases transparency and trust in the model's predictions, encouraging artificial-intelligence collaboration. The key goals of implementing activation mapping methods in the selected publications were classified as follows: (1) Using CAM techniques exclusively as explanatory tools for CNN. (2) Applying CAM approaches for CNN explanation and model validation, assessment, and comparison. (3) Use CAM approaches to improve the interpretability of freshly produced models. (4) Creating a framework for developing novel activation mapping approaches. (5) Using them in an innovative way to pick features.

In [76], they selected eight different CAM methods: Grad-CAM, GradCAM++, XGradCAM [72], AblationCAM, EigenCAM [73], EigenGrad-CAM, Layer-CAM [74], and FullGrad [75]. Grad-CAM is currently the most widely used CAM method in the medical profession, with the remaining seven being variants of it.

|  |  |
| --- | --- |
| Comparison of Different CAM Methods. | |
| **Technique** | **Description** |
| Grad-CAM | Uses convolutional layer gradients to create heatmaps, highlighting crucial image regions for decision-making. |
| GradCAM++ | Improved Grad-CAM with refined localization and gradient weighting for multi-class tasks. |
| XGradCAM | Adjusts Grad-CAM gradients and activation maps for better consistency and interpretability. |
| Ablation-CAM | Generates activation maps by ablating convolutions; focuses on feature importance for decision-making. |
| EigenCAM | Uses PCA for gradient-independent visualization of key features in convolutional layers. |
| EigenGradCAM | Hybrid of EigenCAM’s PCA and Grad-CAM’s gradients for comprehensive activation maps. |
| LayerCAM | Accumulates information across layers for a broader perspective on model decisions. |
| FullGrad | Combines input image gradients and convolutions for detailed class activation maps. |

The heatmaps were created utilizing four different models (DenseNet121, EfficientNet-B3, ResNet50, and GoogleNet) and eight CAM approaches for breast cancer classification tasks. These findings clearly show that different CAM approaches have distinct interpretative consequences on various models. Specifically, utilizing the DenseNet121 model, the EigenCAM technique achieved the greatest IoU score (0.372), demonstrating higher interpretative capabilities on this model. GradCAM and XGradCAM approaches both received the highest score (0.307) in the EfficientNet-B3 model, proving their effectiveness. AblationCAM outperformed the ResNet50 model, achieving an IoU score of 0.271. Finally, with the GoogleNet model, the EigenGradCAM approach performed the best, scoring 0.192.