IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 July 2023, accepted 2 September 2023, date of publication 7 September 2023, date of current version 13 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312729

RESEARCH ARTICLE

GastroNet: Gastrointestinal Polyp and Abnormal Feature Detection and Classification With Deep Learning Approach

FARHANA YASMIN[1], MD. MEHEDI HASSAN[2], (Member, IEEE), MAHADE HASAN[1], SADIKA ZAMAN[3], ANUPAM KUMAR BAIRAGI[2], (Senior Member, IEEE), WALID EL-SHAFAI[4], (Senior Member, IEEE), HASSAN FOUAD[5], AND YANG CHANG CHUN[1]

[1]School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu 213164, China
[2]Computer Science and Engineering Discipline, Khulna University, Khulna 9208, Bangladesh
[3]Department of Computer Science and Engineering, North Western University, Khulna 9100, Bangladesh
[4]Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt
[5]Applied Medical Science Department, Community College, King Saud University, Riyadh 11433, Saudi Arabia

Corresponding authors: Md. Mehedi Hassan (mehedihassan@ieee.org), Anupam Kumar Bairagi (anupam@ku.ac.bd), and Walid El-Shafai (eng.waled.elshafai@gmail.com)

This work was supported by King Saud University, Riyadh, Saudi Arabia, Researchers Supporting Project RSP2023R117.

ABSTRACT The early detection of digestive problems is essential for lowering the chance of acquiring any form of gastrointestinal cancer, including esophageal cancer. Endoscopy is the method that is used the majority of the time for the purpose of examining and taking photos of this sort of illness. The application of artificial intelligence is now proving to be very efficient in enhancing the identification of gastrointestinal polyps and other abnormal features located inside the gastrointestinal system. As a direct consequence of this development, the use of AI within this sector has seen substantial growth. In the framework of artificial intelligence, this research investigates how well various types of algorithms perform in terms of polyp and abnormal feature recognition accuracy, efficiency, and detection. And introduced a model in this work that is GastroNet. It is developed by doing hyperparameter fine-tuning on YOLOv5 in order to find specific polyps and abnormal characteristics, particularly esophagitis. In this method, a single neural network is used to do an analysis on the whole picture before it is disassembled into its component parts and the bounding boxes and probabilities for each one are calculated independently. The goal of the hyperparameter fine-tuning is to further enhance the overall optimization of the model. Two different methods of annotation were used on a collection of data that consisted of one thousand separate images that needed to be labelled. In addition to implementing the fine-tuned SSD model, this study used three distinct backbone networks: MobileNet v2, MobileNet v2 FPN Lite, and Resnet50 v1 FPN. Additionally, this study has used CSPdarknet53 to create the improved YOLOv4 model. The results of the studies demonstrate that the proposed model, GastroNet was effective in correctly recognizing polyps and aberrant characteristics, reaching a high mAP (mean Average Precision), F1 score, and precision with a value of 0.99 and recall with a value of 1.00. The findings of this research will be a great help to physicians in the proper identification and diagnosis of abnormal features and gastrointestinal polyps.

INDEX TERMS Artificial intelligence, detection and classification, GastroNet, machine learning, health care, gastrointestinal malignancies.

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

I. INTRODUCTION

In the year 2023, it is expected that there will be 1,958,310 newly diagnosed cases of cancer in the United States,

in addition to 609,820 deaths that would be caused by cancer. There will probably be 28.4 million newly diagnosed cases of cancer in the world in the year 2040 [1]. This is an indication that should raise concerns about the health and manner of life of the population. When compared to the total number of new instances of cancer that are diagnosed each year in the United States, the mortality toll attributable to esophageal cancer is disproportionately high. According to the statistical data that was compiled by the American Cancer Society in 2023, there will be 21,560 new instances identified in this year, and 16,120 individuals are anticipated to pass away because of the illness. It is anticipated that 17,030 men and 4,530 women would get a diagnosis of esophageal cancer in the United States in the year 2023. The total number of persons affected is 21,560. In the United States, deaths caused by esophageal cancer in men are the sixth most prevalent kind of cancer-related death overall. The risk of acquiring gastrointestinal malignancies of all types, including esophageal cancer, must be reduced by the early detection of gastrointestinal ailments.

According to the National Cancer Institute (NCI) [2], esophageal cancer ranks 17th among malignancies in incidence (new cases) in the United States, but it is the 11th largest cause of death due to cancer. It has a higher ranking in other parts of the world, coming in at number six on the list of major causes of death due to cancer and eighth on the list of most prevalent cancers.

Endoscopy is the medical expert's most suggested test for identifying esophageal cancer in its early stages. The care of esophageal cancer, particularly in Western nations, now centers on endoscopic surveillance of Barrett's esophagus (BE) [3], [4], [5].

According to study, high resolution endoscopy was the most accurate method for diagnosing early esophageal malignancies [6]. This can provide crucial details regarding the size and distribution of the tumour, which can be used to assess whether surgery can be utilized to remove it. This may assist to prevent the development of some esophageal cancers. Nevertheless, polyps may be eliminated based on their particular characteristics and the technical proficiency of the endoscopist. Approximately 27% of po1yps are removed, according to reports [7].

Artificial intelligence (AI) technologies are currently a prominent topic of research in clinical medicine, e.g. to detect covid-19, cancer, skin disease and mental illness as well. This study focused on notable occurrences in the most recent monkeypox dataset [8]. Where the prophet model was used, which performed consistently when compared to the state of the art. The authors [9] applied K-means algorithms for the elbow and silhouette. Numerous machine learning techniques using cluster-based datasets have been used to forecast diabetes. An unsupervised cluster-based feature grouping algorithm for early diabetes identification is demonstrated using of 520 patients.

AI has been applied in gastrointestinal oncology in many different methods, with significant improvements in computer-aided detection and diagnosis. The most popular focus has focused on colorectal cancer, although esophageal disorders have also been the subject of AI development. Numerous studies in this field have demonstrated excellent diagnostic properties, many of them may not have strong external validity. Mahmud et al. [10] describe the potential integration of computer vision and augmented reality into an endoscopic setup and foresee a number of unique applications. However, there are a very few methods that can accurately detect polyps and abnormal features like esophagitis.

Yu et al. [11] improved anchor construction technique allowed them to produce more boxes for detection and a better anchor box form. A separate backbone is utilised for polyp detection to compensate for the considerable time cost brought on by dense anchor box regression. Where 0.90 represented the F1 score that fell short of expectations. Hoang et al. [12] the method uses the YOLOv3 deep learning algorithm, which has a fairly poor precision score of 85% on average yet can identify polyps. Durak et al. [13] applied a deep learning-based computer-aided diagnosis system YOLOv4 for the detection of gastric polyps, and it scored poorly with a mean average precision of 87.95 percent.

In recent years, the field of medical imaging and diagnostics has witnessed remarkable advancements in the application of artificial intelligence (AI) for the early detection and diagnosis of various diseases. In the context of gastrointestinal (GI) disorders, including gastrointestinal polyps and abnormal features, AI has shown great potential in improving recognition accuracy and overall diagnostic efficiency. Several studies [15], [50], [51] have investigated the use of AI algorithms for identifying these anomalies within the gastrointestinal system. The study [49] compares seven traditional semantic segmentation models with ResNet50, MobineNetV2, and EfficientNet-B1 encoders. An integrated evaluation approach is suggested for selecting the best CNN model, combining subjective and objective information. The automated polyp-segmentation system is built using UNet++ with MobineNet v2 encoder. The integrated assessment technique is neutral and objective, and the semantic segmentation model has high clinical value in gastric polyp identification. This work [52] introduces a masked graph neural network model (MGNN) for real-time polyp detection in gastroscopic images in Healthcare 4.0. The model uses graph structure and convolution operations to extract spatial and semantic information, compensating for manual labeling. It has been tested on real gastroscope images.

However, edge blur between surrounding tissues will continue throughout the polyp detection procedure because of the inherent characteristics of the colorectal image, insufficient brightness, noise, contrast, and the technical limitations of the imaging equipment. Additionally, the criteria of practical application cannot be met by current artificial image feature selection methodologies in medical image processing and analysis. The latest available solution falls short of expectations. Therefore, the purpose of this

study is to offer an accurate and trustworthy method of polyp and abnormal feature detection.

GastroNet is a neural network developed by fine-tuning the YOLOv5 algorithm to increase the recognition accuracy and efficiency of polyps and abnormal characteristics, particularly esophagitis, within the context of artificial intelligence research. GastroNet introduces a novel strategy by employing a single neural network to analyze the entire image before deconstructing it into its component elements. This differs from conventional image processing methods, which typically divide an image into regions and process each separately. By analyzing the image as a whole, GastroNet is able to capture contextual information and the relationships between objects, resulting in more precise and reliable detection.

The procedure of hyperparameter fine-tuning is conducted to further optimize the model's performance. The following are the specific hyperparameters used by GastroNet:

- 100 epochs are used to send the entire training dataset through the model during training. More epochs enable the model to learn intricate patterns and enhance its precision.

- Batch Size: 16 is the number of training examples utilized per iteration. A larger sample size can speed up the training process, but it requires additional computational resources.

- Image Size: The input image resolution set to 416×416. The selected dimension strikes a balance between detection precision and computational efficacy.

- Weight Decay: A regularization parameter that is set to 0.0001 and inhibits overfitting. It controls the model's complexity and generalizes well to unseen data.

- Warmup Momentum: A parameter that, when set to 0.9, regulates the rate of change in the learning rate during initial training. It serves to accelerate and stabilize the training process.

The results of the studies demonstrate the accuracy with which GastroNet recognizes polyps and aberrant characteristics. The high mAP (mean Average Precision) indicates that GastroNet localizes and classifies objects precisely. The F1 score, precision, and recall values of 0.99 and 1.00 indicate that the model can detect these features with a high degree of accuracy and dependability.

The distinguishing characteristics and benefits of GastroNet include:

1. Contextual Analysis: By analyzing the entire image, GastroNet is able to capture global contextual information, resulting in a more precise and robust detection of polyps and aberrant characteristics.

The architecture of GastroNet enables the efficient processing of images by eliminating the need to analyze regions separately. This leads to quicker inference periods and improved scalability.

3. Hyperparameter Fine-Tuning: The fine-tuning procedure optimizes the performance of the model by selecting the most suitable hyperparameters. This further increases the detection's precision and dependability.

Significant prospective impact of GastroNet on detection accuracy improvement. In medical diagnostics and examinations, accurate and efficient detection of polyps and abnormal characteristics, such as esophagitis, is essential. The high precision, recall, and mAP scores of GastroNet indicate its potential to aid in the early detection, treatment planning, and monitoring of gastrointestinal diseases. GastroNet has the potential to improve patient outcomes, expedite medical workflows, and enhance the overall quality of gastroenterology care by leveraging the power of artificial intelligence.

In terms of artificial intelligence, this research discusses how well various types of methods perform in terms of polyp and aberrant feature detection accuracy, efficiency, and detection. The following are the article's primary contributions:

1. Labelling the dataset with two different annotation tools. One is MakeSense.ai, while the other is VGG Image Annotator.
2. Applied three models with variations in five distinct backbones.
3. To accurately detect polyps and abnormal features, propose a model by performing hyperparameter fine tuning on YOLOv5 and comparing with four other fine-tuned models.
4. Examined the performance of the suggested model and contrasted it with earlier research.

We have briefly addressed related earlier work in the related works section. Then describe the research technique we used. In this we discussed the proposed model, the applied models process, and the dataset characteristics. Then, during the result analysis phase, the conclusions of our suggested model were examined. Finally, we evaluated how well our model performed in comparison to four other applied models and with earlier models that were still in use.

II. RELATED WORKS

In the last 20 years, advances in artificial intelligence (AI) and deep learning (DL) have made it possible to quickly sort and analyze massive data sets. The field of gastroenterology can learn a lot from colorectal screening procedures. AI and deep learning algorithms that can automatically find and classify polyps can be developed using this data. It has been found that Convoluted Neural Networks (CNNs) are an effective technique for finding polyps.

Using attentive YOLOv5, Wan et al. [14] suggested a method for polyp detection from colorectum images. Regression theory is used to feed the complete image into the network, and at various locations throughout the image, the target frame of this position is immediately returned. Input, backbone network, neck, attention mechanism, and prediction are its five primary divisions. To extract information features from photos, YOLOv5 uses a backbone network. As the neck, it makes use of PANET for feature aggregation and SPP to improve the model's detection of objects of various scales. The model's information storage becomes richer as it contains more parameters. To combat

the problem of information overload, they added a self-attention module to the top layer of each step of the feature extraction backbone network. An attention-based YOLOv4 detector for polyp identification and localization is suggested by Sasmal et al. [15] in their study. The performance of the suggested algorithm significantly outperforms that of cutting-edge methods. The consistency of outcomes within and between datasets further demonstrates the generalizability and robustness of the strategy. The localized polyps are then categorized, which is essential for improving prognosis. Following SVM, they then suggest a triplet network built on the Siamese architecture. A deep learning object detection network is created by Cao et al. [16] by fusing the feature extraction and YOLOv3 network. It is possible to precisely identify and locate tiny polyps thanks to the feature map F's combination of low-level detailed texture information and high-level semantic information. The feature extraction and fusion module in the 36-layer network is made up of the feature fusion portion of the final three simple convolutional blocks from the backbone network and the feature pyramid. They divide Darknet-53 into six convolutional layers for feature extraction. To find small, medium, and large objects, they use three distinct scale detectors in the detecting head.

A CNN-based single image super resolution SRCNN model is suggested by Taş and Yılmaz [17] which is a collection of filters that enables the mapping of low-resolution inputs to high-resolution outputs. The Faster RCNN and SSD networks with Inception-v2 and ResNet-101 on the Tensor flow platform were then built using these high-resolution images as inputs. Compared to SSD Inception-v2 with SRCNN and Faster RCNN Inception-v2 with SRCNN, the Faster RCNN ResNet-101 with SRCNN performs better. According to Ellahyani et al. [18] they offer a cutting-edge method of computer-assisted diagnosis for the detection of polyps. After a first stage of pre-processing, deep features are extracted to carry out the detection of polyps using a fusion of two deep neural networks (DNNs) that were pre-trained on millions of tagged natural photographs (ImageNet). The Kvasir-seg dataset's images are used in the fine-tuning process. Additionally, the initial layer weights of the networks used in this work have been fixed. By combining the outputs of the models that had been adjusted until they were completely coupled, they ultimately performed the binary classification. Due to the significant prevalence of small polyps in various data sets, Li et al. [19] established a low-rank model with the human resources network as the backbone to perform accurate polyp segmentation.

By integrating the conventional VGGNets and Resnets models with global average pooling, Wang et al. [20] developed two separate lightweight network architectures, the VGGNets gap and Resnets gap, with good classification accuracy and fewer parameters. Haj-Manouchehri and Mohammadi [21] constructed a complete convolutional network and an effective post-processing technique for polyp segmentation after first developing a special convolutional neural network for polyp frame recognition based on the VGG network. To increase the effectiveness of detection, Bochkovskiy et al. [22] have presented a number of target detection algorithms based on the yolo series. An autonomous polyp recognition method with a decreased false positive rate was proposed by Guo et al. [23], and it is based on the yolov3 structure and active learning. The Yolov3 network was fused by Cao et al. [16] using a feature extraction and fusing module they created. Because it can combine the semantic information of a high-level feature map with a low-level feature map, this method surpasses earlier ones in the detection of small polyps. Yolov4-based real-time automatic polyp recognition was proposed by Pacal and Karaboga [24] in their study. They used the transformer block, mish activation function, Diou loss function, and CSPNet network throughout the entire architecture. In terms of accuracy and effectiveness, this approach performs better than prior ones. Taş and Yılmaz [17] proposed super-resolution convolutional neural network-based polyp identification pre-processing for colonoscopy pictures for the preparation stage of polyp detection in colonoscopy images. Polyps can be found with this pre-processing as image resolution increases. Transfer learning is advised for high-resolution colonoscopy polyp identification by Tang et al. [25]. The identification of the polyps was precise but not categorical. A transformer convolution network was recommended by Shen et al. [26] for end-to-end polyp identification. Transformer encoder and convolutional layers were interleaved for feature coding and recalibration after CNN features were extracted. The transformer decoder layer discovered objects, whereas the feedforward network discovered targets.

By combining the enhanced depth residual network, principal component analysis, and AdaBoost ensemble learning, Liew et al. [27] proposed an automatic colon polyp diagnosis method based on endoscopic images. Median filtering, picture thresholding, contrast enhancement, and normalization were used to train the classification model to eliminate visual interferences.

This study [45], evaluated deep learning's potential, constraints, and existing use in thyroid cancer imaging. They discussed about the most recent developments in deep learning-based thyroid cancer diagnosis as well as the numerous difficulties and practical barriers that might prevent it from developing and becoming part of the healthcare system. In order to construct a flexible and successful brain tumour segmentation system, this research [46], suggests a preprocessing technique that only affects a small portion of the image. This speeds up processing and eliminates overfitting in Cascade Deep Learning models. Because each slice has a smaller representation of the brain, a straightforward and effective Cascade Convolutional Neural Network (C-ConvNet/C-CNN) is suggested in the second stage. Two methods are used by this C-CNN model to mine local and global features. Compared to existing models, a unique Distance-Wise Attention (DWA) mechanism increases the accuracy of segmenting brain tumours. The brain and tumour centre are accounted for by the DWA mechanism of the

model. A breast cancer recognition system is proposed by Ranjbarzadeh et al. [47], that uses a variety of encoding techniques to create new images from input photographs. Each encoded image has distinct qualities required for texture detection. Pectoral muscle removal is achieved using encoded image features. A quick cascade that works well the pixels of 11 distinct photos are categorized by CNN. These 11 local patches are fed by 11 encoded pictures. For entirely connected layers, every gathered characteristic is concatenated to a vertical vector. The suggested model utilizes various mammography picture formats to interpret input material more effectively without a deep CNN model. Finally, thorough tests on two open datasets demonstrate the proposed framework's superior performance to a number of baselines. This study [48], uses a distinctive four-feature extraction CNN architecture to segment brain cancers in MRI images. T1, T2, T1-c, and FLAIR were the four imaging modalities used. To increase segmentation accuracy, they applied ResNet-50 weights and biases. To extract more features from each modality, they established two fundamental construction components. Two patch sizes are used to model local and global features.

III. METHODOLOGY

In the data collection block, we had an in-depth discussion about the various datasets and the sources of the datasets. The process of data labelling was then discussed in detail. Our dataset has been divided into two parts: a train set and a validation set. The train set accounts for 80% of the total, while the validation set accounts for the remaining 20%. After that, as is seen in the applied model block, we used our dataset to test five different models in order to determine which one performed the best when applied to our dataset. We carried out customized setup and fine-tuning procedures for each model on its own. And provided a model that was based on the one that had the greatest overall performance as a suggestion.

In addition, we have discussed our proposed model for recognizing and classifying coloured lifted polyp and esophagitis characteristics, which performs better than the alternatives we considered on the data we selected. In addition to that, the scores for accuracy, recall, mAP, and F1 in the results block were investigated.

This study put the concept to the test using colored lifted polyps and esophagitis. Esophageal inflammation is brought on by esophagitis. The first step in treating and preventing esophageal cancer is detection. Polyps are removed by locating Dyed Lifted Polyps. Damage to the gastrointestinal wall is reduced. This dataset is excellent for identifying and categorizing gastrointestinal anomalies including polyps. Five models successfully detected the intended problems, including esophagitis, and polyps. The f1 and mAP settings were optimum for our model. The changes to the model are outlined. We provide a model to recognize colored lifted polyps and esophagitis by tuning the hyperparameters of YOLOv5.

TABLE 1. Dataset description.

Name	Description
Total Number of Images	1000
Dimension	720x576 up to 1920x1072
Colour Grading	RGB
Total Number of Classes	2
esophagitis	500
dyed lifted polyp	500

The components of the entire are hyperparameter, input, backbone, neck, and prediction fine-tuning. Models for machine learning include variables that must be learnt from data. Training a model using previously acquired data is necessary for fitting model parameters. Hyperparameters, on the other hand, are not taught by a normal training method. Before training, they are corrected. With the appropriate hyper-parameters, neural networks learn more quickly and perform better. These criteria include picture size, batch size, and epoch. Weight decay and warmup momentum are taken into account by YOLOv5's default optimization method, stochastic gradient descent (SGD). In the end, we compared our model to some past research that had been done on the detection and categorization of polyps.

IV. DATASET

In this study, a total of 1000 endoscopic images from Kvasir v1 are evaluated. Endoscopic image dataset Kvasir v1 [28] has various classes. For the evaluation of our model, we have employed two classes: (a) esophagitis and (b) dyed lifted polyp. The illness known as esophagitis causes inflammation of the esophagus. In order to start therapy and prevent later repercussions like esophageal cancer, detection is necessary. The detection of Dyed Lifted Polyps is required for polyp removal. It lessens the possibility of gastrointestinal wall injury to the deeper layers. This dataset is ideal for our investigation in terms of identifying and categorizing the aberrant features and polyps that occur in the gastrointestinal tract. Because we were able to obtain images of a good quality from this dataset, which is of the utmost significance for training a model. Table 1 provides a summary of the dataset's description.

We have utilized two different annotator software for the data labelling. One is MakeSense.ai, while the other is VGG Image Annotator. Where MakeSense.ai annotates half of the images and VGG Image Annotator annotates the other half. The polygon annotation type was employed in the study's dataset annotation procedures to mark and label objects of interest. Both MakeSense.ai and VGG Image Annotator support the annotation of polygons (VIA), enabling annotators to more precisely identify and outline the boundaries of polyps and aberrant structures.

The dataset was manually annotated by expert annotators using the polygon annotation technique to precisely label polyps and abnormal features in gastrointestinal images.

Expert review ensured the accuracy and reliability of the annotations. While manual annotation offers accuracy, potential limitations were addressed through quality control measures. With the polygon annotation technique, annotators manually outline the desired items in the photos with polygonal forms. These polygons precisely encompass the study areas and offer precise and in-depth annotations for polyps and aberrant traits. When working with intricate and variable constructions, the polygon annotations make it possible to capture uneven shapes and curves.

The study's annotators use the polygon annotation type to provide more precise and fine-grained annotations, defining the precise boundaries of polyps and aberrant characteristics. This method increases the detection accuracy by allowing the trained models to understand and recognise the specific shapes and patterns connected to these objects.

The study made sure that polyps and aberrant features were consistently and accurately delineated across the annotated dataset by using the polygon annotation type in both MakeSense.ai and VGG Image Annotator. This annotation strategy enhances the dataset's dependability and utility for developing and testing machine learning models for polyp and abnormal feature recognition tasks. We divided the dataset into a train set and a validation set. For the training set, 800 images were randomly chosen, while the remaining 200 images were used for validation. The sample examples of each class with the ground-truths bounding box are shown in Figure 2 (a) esophagitis samples (b) ground-truths with bounding boxes. Figure 3 (a) dyed lifted polyp samples (b) ground-truths with bounding boxes.

Our dataset is now prepared to be utilized for training, so in the following part, we have implemented a several number of models.

V. APPLIED MODELS

To accurately identify and categorize targeted polyps and abnormal features like esophagitis, we employed a total of five models. Our suggested model had the highest f1 and mAP values among the four alternate models. Here is a brief explanation of how those models were fine-tuned.

A. SSD MOBILENET V2

SSD (Single Shot MultiBox Detector) with a unique depth wise separable convolution and a lean network, Mobilenet v2 is a one-stage object identification model. Convolutional neural network design called MobileNetV2 seeks to be mobile-friendly. It is based on a residual structure with links between bottleneck levels that is inverted [29]. The parameter fine-tuning of the model for our dataset is described in Table 2.

B. SSD MOBILENET V2 FPN-LITE

Model for object identification in SSD Mobilenet V2 using shared box predictor, focal loss, and FPN-lite feature extractor. A single-scale image of any size can be entered into the FPN (Feature Pyramid Network), a fully convolutional feature extractor, and it will output feature maps of the appropriate size at various layers. This process is unaffected by the underlying convolutional architectures [30]. Table 3 details how the model's parameters were fine tuned for our dataset.

TABLE 2. SSD mobilenet v2 model parameter fine tuning description.

Name	Description
Batch Size	8
Image Size	320×320
Fine Tune Checkpoint Type	Detection
Number of Step	10000
Warmup Step	1000

TABLE 3. SSD mobilenet v2 FPN-Lite model parameter fine tuning description.

Name	Description
Batch Size	16
Image Size	320×320
Fine Tune Checkpoint Type	Detection
Number of Step	10000
Warmup Step	1000

TABLE 4. SSD Resnet50 v1 FPN model parameter fine tuning description.

Name	Description
Batch Size	16
Image Size	640×640
Fine Tune Checkpoint Type	Detection
Number of Step	10000
Warmup Step	1000

TABLE 5. YOLOv4 model parameter fine tuning description.

Name	Description
Batch Size	64
Image Size	416×416
Subdivisions	16
Max Batches	4000
Steps	3200, 3600
Filters	21

C. SSD RESNET50 V1 FPN

A model for object detection is the SSD Resnet50 v1 FPN model. The SSD (Single Shot MultiBox Detector) is a method of item discovery for each location on a feature map at different aspect ratios and scales, it divides the output space of bounding boxes into a number of default boxes. A feature extractor FPN (Feature Pyramid Network) serves as the backbone of the 50-layer deep convolutional neural network known as ResNet-50 v1 FPN [30]. The parameter fine-tuning of the model for our dataset is described in Table 4.

FIGURE 1. Illustrates the full process of the proposed GastroNet model, with each phase represented by a block. Block (A) shows the original Kvasir v1 endoscopic dataset, (B) data labelling with two separate annotation tools, and (C) data split into train and validation sets. (D) A list of models that have been used with various feature extractors. (E) Custom configuration and fine tuning those models. (F) Model selection using the highest f1 and mAP value. (G) Analyse the proposed model's results and test it on a random dataset. Using coloured lifted polyps and esophagitis, we tested our theory. Five models, including ones for polyps and esophagitis, effectively identified the desired issues. For our model, the f1 and mAP settings were ideal. The model improvements are described in the proposed model section. By adjusting the YOLOv5 hyperparameters, we proposed a model named GastroNet to identify coloured lifted polyps and esophagitis.

D. YOLOV4

This well-known object detection backbone employs DarkNet-53 as its structural support. It uses a cross-stage hierarchy to combine two portions of a Dense Block feature map, according to the CSPNet technique [22]. Table 5 details how the model's parameters were modified for our dataset.

VI. PROPOSED MODEL

In order to enhance optimization and accurately identify dyed lifted polyps and esophagitis, we propose a model in this study that performs hyperparameter fine-tuning on YOLOv5. These five sections make up the whole: Fine-tuning of the hyperparameters, input, the backbone, the neck, and prediction. Figure 4 depicts the suggested model's architecture.

A. HYPERPARAMETER FINE TUNING

A machine-learning model is a mathematical construction with a number of parameters that must be learned from data. By using previously collected data to train a model, we can fit the model's parameters. On the other hand,

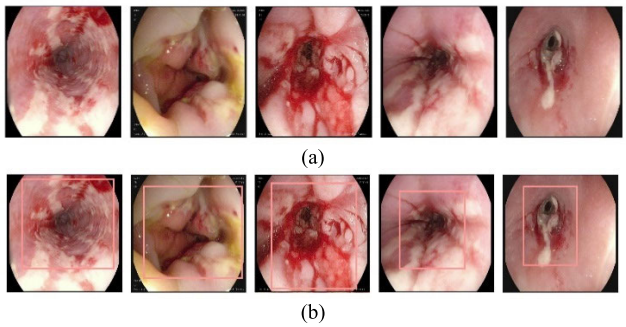

FIGURE 2. (a) Shows some esophagitis samples, (b) Shows the corresponding esophagitis ground-truths with bounding boxes that are denoted by the pink color rectangle.

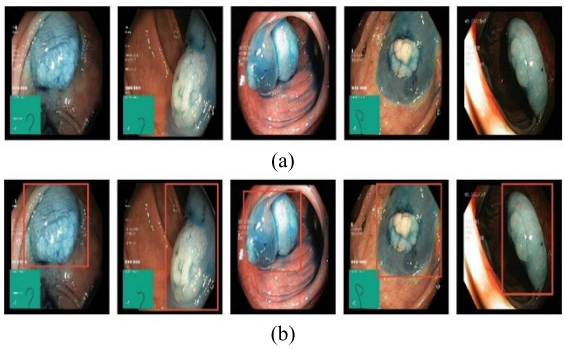

FIGURE 3. (a) Shows some dyed lifted polyp samples, (b) Shows the corresponding dyed lifted polyp ground-truths with bounding boxes that are denoted by the red color rectangle.

utilizing a typical training strategy does not immediately teach hyperparameters. They are normally corrected prior to the training process beginning. Neural networks can learn more quickly and perform better with the right selection of hyper-parameters during model training [31]. The epoch, batch size, and image size of the model are described by these parameters, among others. As a reason, YOLOv5 uses stochastic gradient descent (SGD) as the optimization function by default, weight decay, and warmup momentum additionally taken into account. The hyper parameter fine-tuning process applied to the YOLOv5 model involves optimizing the specific parameters of the model to improve its performance. This process aims to find the best combination of hyper parameters that leads to higher accuracy, efficiency, and detection capabilities. The following are some of the changes we made to the hyperparameters.

B. EPOCHS

The number of epochs since the start of the machine-learning algorithm's run is used to indicate how many times the machine-learning algorithm has traversed the training dataset. In other words, an epoch is a kind of hyperparameter that specifies the number of iterations of the machine learning algorithm that have been carried out throughout the whole of the training dataset [32]. Too few epochs may result in underfitting, while too many may lead to overfitting. Following a series of updates, including Epochs 50, 90, 100, and 120 respectively. GastroNet's fine-tuning set 100 epochs. The model can learn complex patterns and features by adding epochs. However, overfitting occurs when the model memorizes the training data and fails to generalize to new data.

C. BATCH SIZE

The amount of training samples that are employed during the course of one iteration before the internal model parameters are altered is referred to as the "batch size." In other words, it refers to the number of samples that are analysed before the model itself is changed. The number of samples included in a batch has to be more than or equal to one, but it can't be equal to the number of samples found in the training dataset [31]. Smaller batch sizes may increase training noise and improve generalization, while larger ones can speed up training but require more memory. We have conducted tests using batches of 8, 16, 32, and 64 in order to develop our model. Lastly, GastroNet fine-tuned with 16 batches. Batch size affects memory and computational efficiency. Processing more data in parallel speeds up training, but also requires more memory to record gradients and activations. Smaller batch sizes may increase training noise and improve generalization.

D. IMAGE SIZE

When the photos were initially gathered, their dimensions varied from 720 × 576 to 1920 × 1072. The dimensions of the images used during training can significantly impact the model's performance. We have revised its calculations to determine the size that works best with each distinct model. GastroNet used 416 × 416 images. Image size influences model correctness and computational efficiency. Smaller image sizes can speed inference but reduce detection accuracy, especially for little objects. Larger image sizes are more detailed but more computationally expensive. The size should balance accuracy and efficiency for the application

E. WEIGHT DECAY

Weight decay is a regularization method used in deep learning. It operates by adding a penalty term to the cost function of a neural network, which causes the weights to decrease during backpropagation. As a result, there is a lower chance that the network will over fit the training set and experience the exploding gradient problem [33]. It helps the model generalize to unseen cases by preventing it from becoming overly sensitive to the training data. In our model, the weight decay is 0.0001, which we gradually change to 0.0001-0.0005. However, Weight decay adds a penalty term to the loss function to prevent overfitting. GastroNet fine-tuned with 0.0001 weight decay. Weight decay simplifies and generalizes the model. It helps the model generalize to unseen cases by preventing it from becoming excessively sensitive to the training data.

F. WARMUP MOMENTUM

A method for accelerating gradient descent known as momentum builds up a velocity vector along persistent reductions in

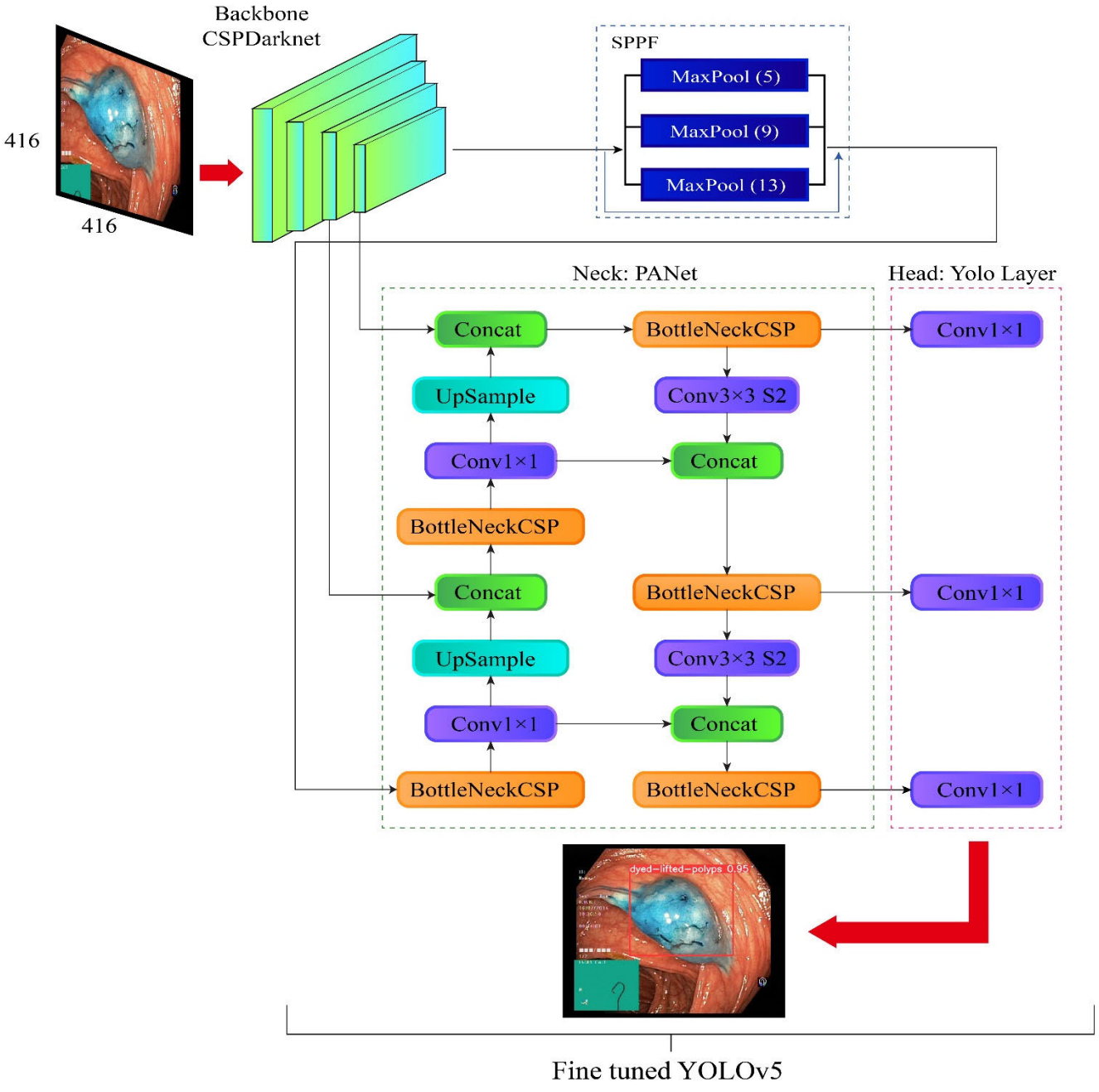

FIGURE 4. Architecture of proposed model where CSPDarknet used as backbone, PANet as neck and yolo layers as head.

the objective across iterations [34]. Momentum is added to the gradient descent optimization process. By lowering the number of function evaluations necessary to find the best solution or by enhancing the efficiency of the optimization algorithm, it aims to hasten the optimization process. To lessen the early training's priority effect. It gradually increases the learning rate from a lower beginning value to the desired value, which stabilizes and accelerates training, preventing the early training priority effect. GastroNet used 0.9 warmup momentum. Warmup momentum stabilizes and accelerates training by gradually increasing the learning rate from a lower beginning value to the desired value. It helps the model adapt to training data and optimize faster.

The summarized description of hyperparameter fine-tuning is shown in Table 6.

G. INPUT

It is simple to over fit the network because our training set only contains a small number of photos. In order to prevent this, we typically need to artificially boost certain image data

TABLE 6. Proposed model hyperparameter fine tuning description.

Name	Description
Epochs	100
Batch Size	16
Image Size	416×416
Weight Decay	0.0001
Warmup Momentum	0.9

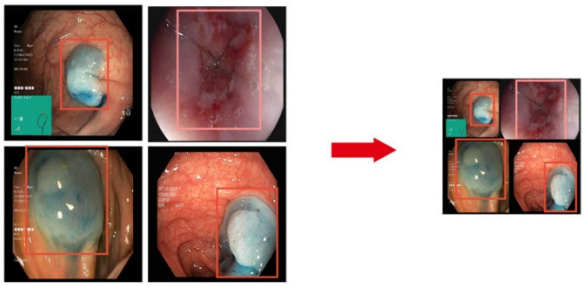

FIGURE 5. Mosaic data enhancement.

by image processing, which raises the total number of photos accessible and lowers the likelihood of overfitting. In order to increase the amount of training data, YOLOv5 leverages mosaic data improvement at the input. Four photographs are chosen at random for the mosaic, which then combines the four images with various semantic information, such as size modification, scaling, rotation, and zooming. The photos are eventually integrated with the frame to create a new image that is positioned in the four directions. Additionally, the frame associated with the new image is extracted. The label frame in each of the four shots is occluded or hidden by a number of other images throughout the splicing process. The image box needs to be eliminated if it touches the edges of the two photographs [26]. YOLOv5 builds a prediction frame based on the original anchor frame, compares it to the ground truth of the real frame, calculates the difference between the two, and then updates it in reverse to determine the best anchor set among multiple training sets. Figure 5 displays the mosaic data enhancement.

The enhanced image data are transmitted to the backbone segment. This is the output of the section based on the input, which is defined further down.

H. BACKBONE

Gradient values are practically non-existent in other circumstances, a phenomenon known as gradient fading. Focus layer and CSPDarknet are used by YOLOv5. The benefit of adopting a Focus layer is lower CUDA memory requirements, a smaller layer, and enhanced forward and back propagation [35].This backbone eliminates redundant gradient information seen in large backbones and incorporates gradient change into feature maps, which speeds up inference, improves accuracy, and shrinks the size of the model by lowering the parameter requirements [36]. CSPNet duplicates the feature map from the base layer and transmits it to the next level using a dense block. Using CSPNet, the F function is immediately optimized as follows.

$$y = M\left(\left[x_0', T\left(F\left(x_0''\right)\right)\right]\right) \tag{1}$$

where x_0 is divided in two along the channel and can be written as $[x_0', x_0'']$. The transition functions T and M are utilized to combine the two segmented components once the gradient flows have been truncated [37]. It aims to solve the problem of repeating network optimization gradient information in the backbone network. CSPNet's architecture reduces the number of model parameters and calculations by using a single-point-sharperference model.

I. NECK

In YOLOv5, SPPF (spatial pyramid pooling - fast) is utilized to obtain results that are mathematically identical to those of SPP, but with less FLOPS and faster rates [38]. It combines elements from every level, reducing the distance between the lowest and highest levels. Adaptive feature pooling is used to reconstruct the damaged information channel between each candidate region and all feature levels [39], which can also identify the various sizes of the targeted polyp and anomalous feature scale, in order to avoid arbitrary allocation. Figure 6 depicts an illustration of the PANet.

After obtaining the size of the targeted polyp, the following is an explanation of how the prediction of the bounding box and the prediction of the polyp class were carried out in order to produce accurate results.

J. PREDICTION

Similar to Yolov3 and Yolov4, YOLOv5 has the same head.

1) BOUNDING BOX PREDICTION

For each bounding box, the network predicts four coordinates: t_x, t_y, t_w, and t_h. If the cell is offset from the top left corner of the image by (c_x, c_y) and the bounding box prior has width and height p_w, p_h, then the predictions correspond to:

$$bx = \sigma(tx) + cx \tag{2}$$
$$by = \sigma(ty) + cy \tag{3}$$
$$b_w = p_w e^t{}_w \tag{4}$$
$$b_h = p_h e_h^t \tag{5}$$

During training, it employs sum of squared error loss. The gradient is equal to the ground truth-value minus the prediction: $t'*$- t$*$ if the ground truth for a coordinate prediction is $t'*$. The ground truth-value can be easily determined by inverting the aforementioned formulae. This technique forecasts an object class score for each bounding box using logistic regression. This value should be 1 if the bounding box prior overlaps a ground truth object more than any other bounding box prior. If the prediction overlaps a ground truth object by more than a threshold yet the bounding box prior is not the best, the prediction is dismissed. Each

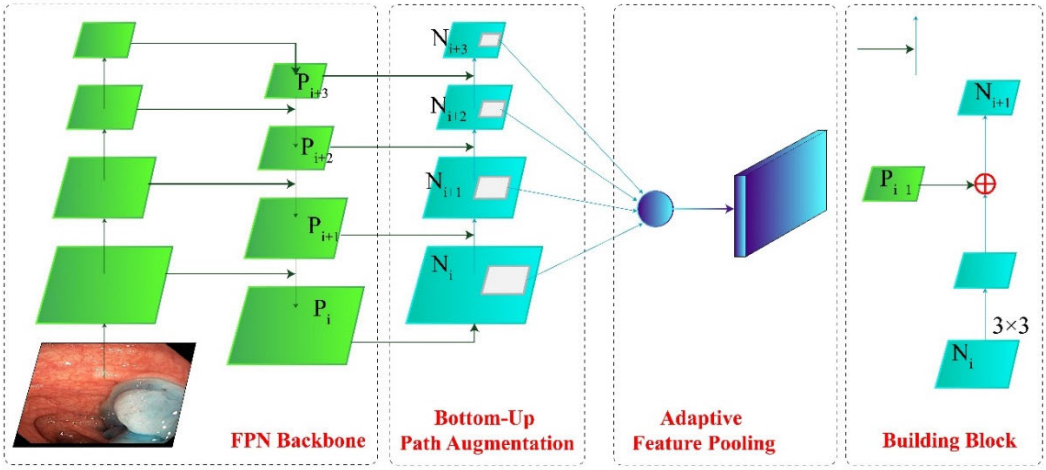

FIGURE 6. Illustration of PANet (Path aggregation network) in YOLOv5, showcasing the utilization of SPPF (spatial pyramid pooling - fast) and adaptive feature pooling. PANet combines features from multiple levels, reducing distance and reconstructing information channels to accurately identify various sizes of targeted polyps and anomalous features while avoiding arbitrary allocation.

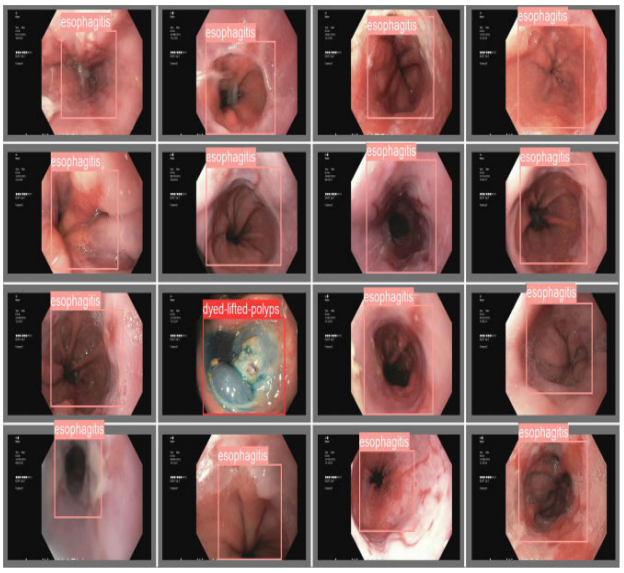

FIGURE 7. Bounding box and class prediction.

ground truth object is given a single bounding box before using a threshold of 0.5 [40].

2) CLASS PREDICTION

Each box forecasts the classes that the bounding box might include using multi-label classification. Since a softmax is not necessary for high performance, it depends on several logistic classifiers instead. During training, binary cross-entropy loss is used to generate the class predictions [40]. Figure 7 displays the bounding box and class prediction.

In the following section, we will discuss the outcome in further detail, taking into consideration our prognosis and classification of the polyp and the abnormal characteristic.

VII. RESULTS ANALYSIS

In image classification and detection tasks, performance evaluation metrics such as F1 score, precision, recall, and mean Average Precision (mAP) are frequently used to judge the performance of the models. These metrics help evaluate the precision of predictions and provide insight into numerous aspects of model performance.These metrics are essential for assessing the performance of image classification and detection models because they provide a quantitative evaluation of precision, dependability, and overall efficacy in identifying and localising objects within images. By analysing these indicators, this study compares multiple models, modifies parameters, and makes prudent judgements to improve system performance. Specific details on the evaluation metrics used and their calculation methodologies are as follows:

$$Precision = \frac{TruePositives}{(TruePositives + FalsePositives;} \tag{6}$$

$$Recall = \sqrt{\frac{TruePositives}{(TruePositives + FalseNegatives}} \tag{7}$$

$$F1Score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{8}$$

$$\overline{mAP} = \sqrt{\frac{1}{Numher\ of\ Classes}} \times \sqrt{\sum_{k=1}^{k=n} Average\ Precision} \tag{9}$$

A. PRECISION

A model's precision is a measure of how frequently it makes accurate predictions. It demonstrates the precision of the model's optimistic forecasts [41]. When the batch size was 16 and the epochs were 100. We obtained the highest precision value for our dataset. We tried other batch sizes and epoch values, but the batch size 16 and epochs

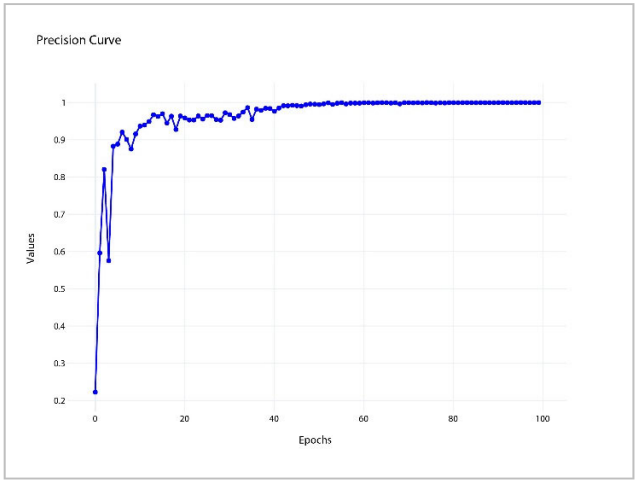

FIGURE 8. Per epoch precision value graph of proposed model.

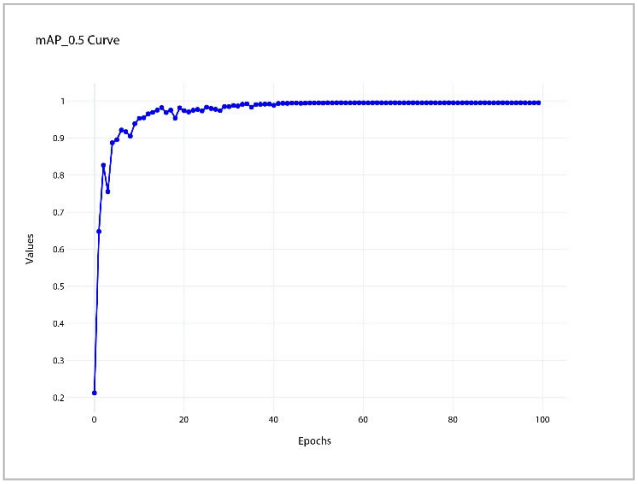

FIGURE 10. Per epoch mAP@0.5 graph of proposed model.

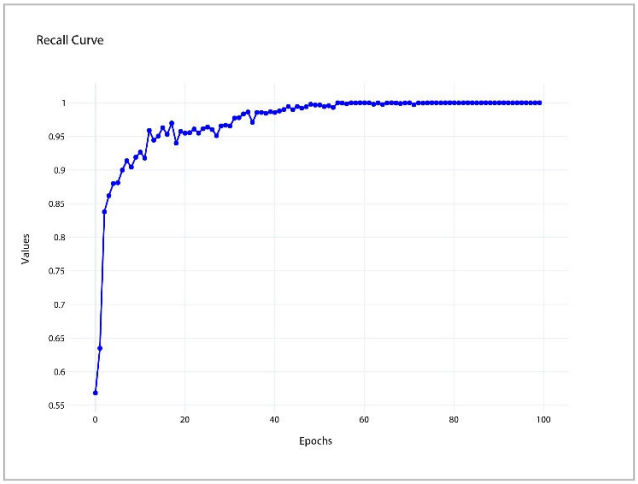

FIGURE 9. Per epoch recall value graph of proposed model.

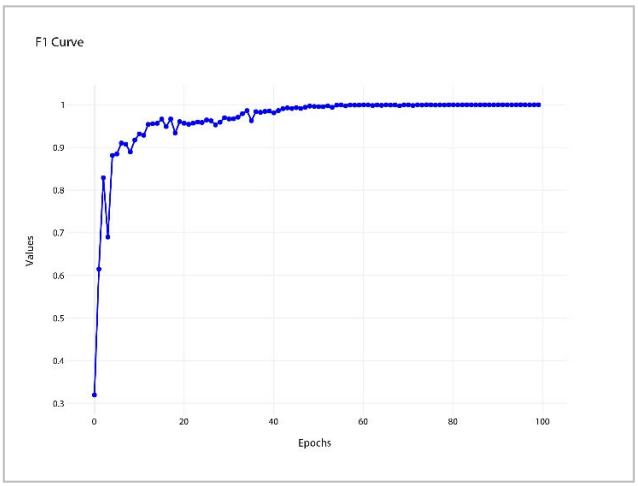

FIGURE 11. Per epoch F1 score graph of proposed model.

100 combination worked best. Figure 8 displays the per-epoch precision value of our proposed models.

B. RECALL

The number of relevant elements discovered is assessed by recall. Therefore it separates true positives by the number of relevant items [41]. When our batch size was 16 with epochs 100. We achieved the best recall value for our dataset. Despite experimenting with numerous batch sizes and epoch values, the batch size of 16 and epochs 100 combination proved to be the most efficient. Per epochs, recall value of our proposed model is displayed in Figure 9.

C. mAP

Models for object detection are evaluated using the mean average precision (mAP). By contrasting the detected box with the ground-truth bounding box, the mAP determines a score. The accuracy of the model's detection increases with score [42]. As previously mentioned, the batch size 16 and epochs 100 combination turned out to be the most effective for mAP value despite experiments with various batch sizes and epoch values. Figure 10 displays the per-epoch mAP value of our suggested model.

D. F1 SCORE

The F1 score is the harmonic mean of recall and precision. It is a metric for statistically assessing performance. In other words, a performance based on precision and recall is averaged to get an F1-score [41]. The F1 score, as is usually known, is the harmonic mean of recall and precision. When we reached our highest levels of precision and recall, we automatically received the best F1 score. Figure 11 displays the per-epoch F1 score for our suggested model.

Figure 12 illustrates how our classification loss for the validation dataset was significantly lower than for the train dataset. How accurately the algorithm predicts an object's class is shown by the classification loss. While validation objectness loss was far lower than validation bounding box loss. This proves that the bounding box was accurate and comprehensive. The precision and recall value still have a curved shape before 30 epochs. However, it performs

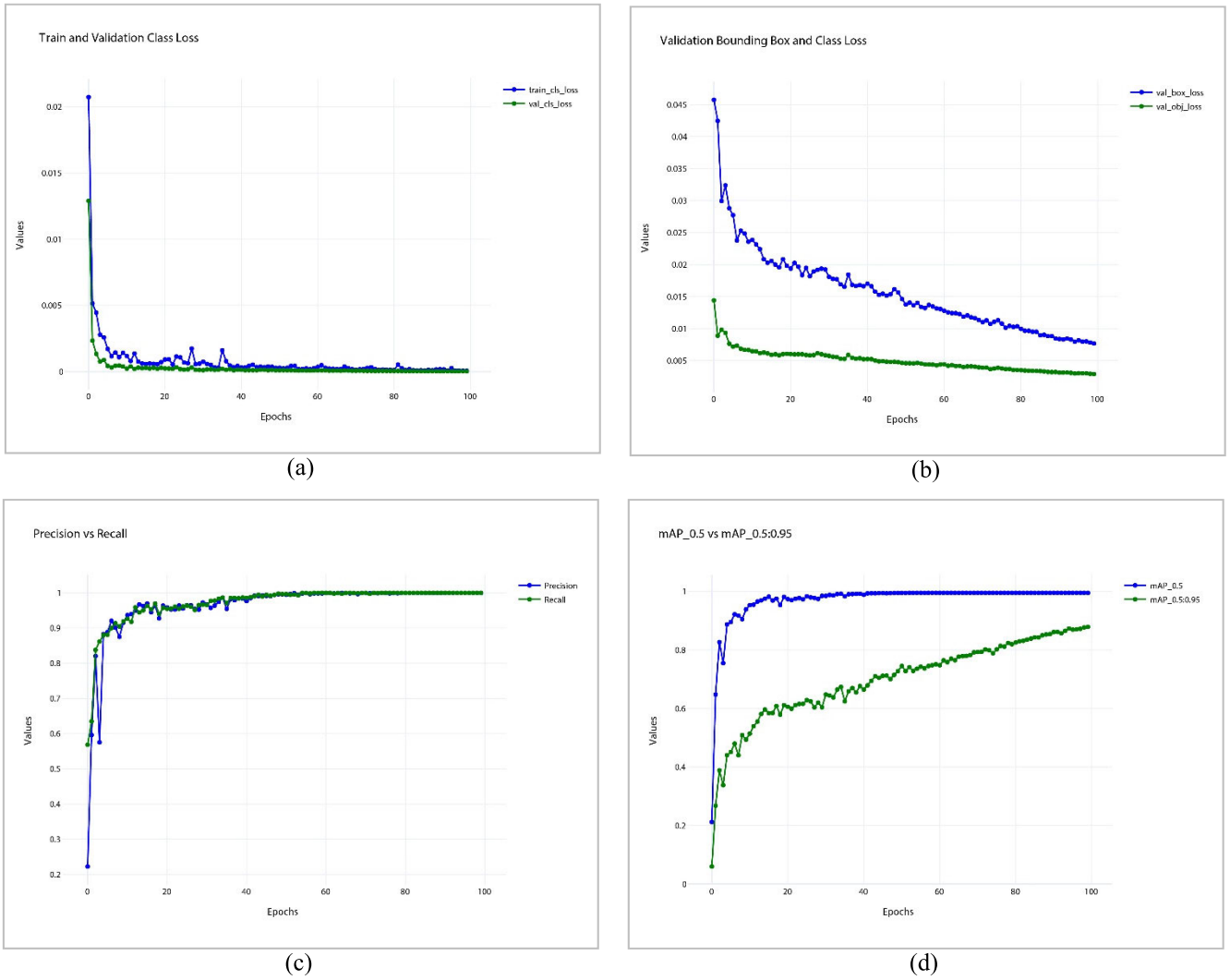

FIGURE 12. **(a) Shows the comparison between train and validation classification loss. (b) Shows the comparison between validation objectness loss and bounding box loss. (c) Shows the comparison between precision and recall. (d) Shows the comparison between validation mAP@0.5 and mAP@0.5:0.95.**

quite well after 50 epochs. Furthermore, mAP@0.5:0.95 was unsteady whereas mAP@0.5 had a smooth value of 0.99.

GastroNet is a novel model introduced in the paper for the purpose of polyp and abnormal feature recognition in gastroenterological imaging. It stands out due to several advancements and improvements compared to previous research. One notable aspect is the utilization of the CSP-Darknet backbone, which incorporates a cross-stage partial connection technique. This unique architecture enhances feature extraction capabilities, leading to improved accuracy in detecting polyps and abnormal features. The study also pays attention to the dataset annotation process by involving two different annotation software tools, MakeSense.ai and VGG Image Annotator, along with multiple annotators. This approach ensures diversity and consistency in the labeling process, minimizing biases and enhancing the reliability of the annotated data.

A comprehensive performance comparison is presented, demonstrating the superior performance of GastroNet compared to existing models. It outperforms models such as YOLOv4, SSD, and Faster RCNN in terms of precision, recall, F1 score, and mean Average Precision (mAP). This highlights the significant advancements achieved by GastroNet in accurately detecting and recognizing polyps and abnormal features. Furthermore, GastroNet is specifically designed and optimized for gastroenterological imaging datasets, with a particular focus on esophagitis. This targeted approach improves the model's performance by tailoring it to the specific characteristics and challenges of gastroenterological images. In summary, GastroNet's novelty and contribution lie in its adoption of the CSP-Darknet backbone, meticulous dataset annotation techniques, and superior performance compared to existing models. These advancements and improvements make GastroNet a

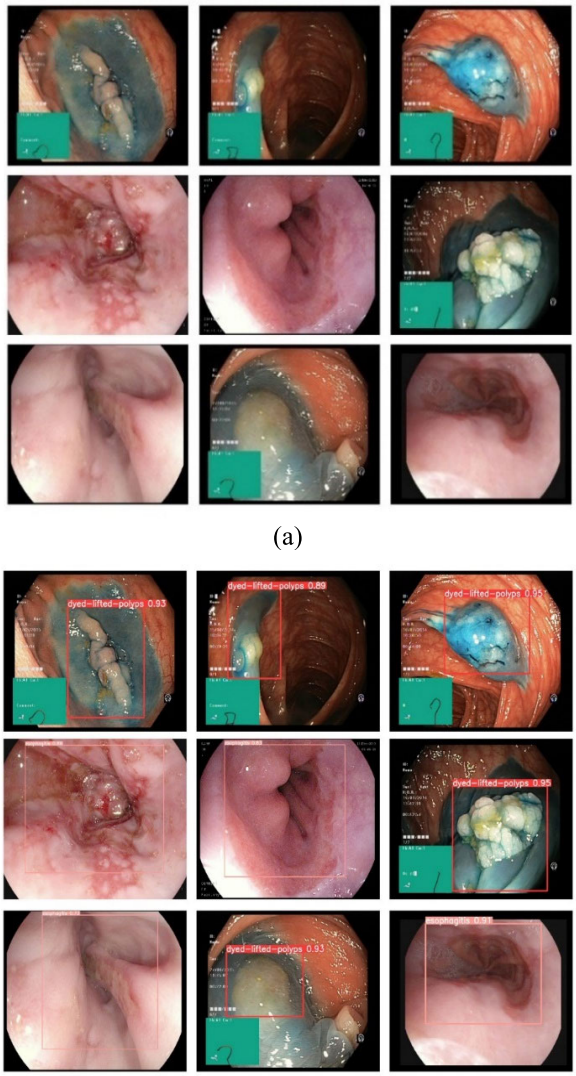

FIGURE 13. (a) Test input image of the Proposed Model; (b) Test output image of the proposed model.

significant step forward in the field of polyp detection and abnormal feature recognition in gastroenterological imaging.

VIII. TESTING MODEL

On some arbitrary data, we tried our proposed model. Our model performed remarkably well, with a high prediction score, in the detection and categorization of dyed lifted polyps and esophagitis. Figure 13 (a) and (b) demonstrate the tested input and output outcomes of the suggested model.

IX. DISCUSSION

The proposed GastroNet model demonstrates superior performance in gastrointestinal polyp and abnormal feature recognition compared to various backbone networks, including MobileNet v2, MobileNet v2 FPN Lite, Resnet50 v1 FPN, and the fine-tuned SSD model.

The selection of these specific backbone networks are attributed to several factors. MobileNet v2 is known for its efficiency in terms of computational resources, making it suitable for real-time applications with limited processing power. MobileNet v2 FPN Lite further enhances feature representation by introducing Feature Pyramid Networks (FPN), enabling multi-scale feature extraction and improving detection performance. Resnet50 v1 FPN, based on the ResNet architecture, is renowned for its depth and capability to capture complex features. By incorporating FPN, Resnet50 v1 enhances the representation of features at different scales, enabling the model to detect objects of varying sizes effectively. The fine-tuned SSD model likely serves as a baseline comparison in this study. SSD (Single Shot Multibox Detector) is a popular object detection model known for its speed and efficiency. Fine-tuning the SSD model allows to compare the performance of GastroNet against an existing well-established model.

The selection of these diverse backbone networks allows for a comprehensive evaluation of GastroNet's performance against different architectures, ensuring that the proposed model is versatile and adaptable to various medical imaging scenarios. Therefore, the choice of MobileNet v2, MobileNet v2 FPN Lite, Resnet50 v1 FPN, and fine-tuned SSD model as backbone networks for GastroNet is driven by considerations of computational efficiency, multi-scale feature extraction, depth, and established baseline comparison. These selections contribute to GastroNet's state-of-the-art performance in gastrointestinal abnormality recognition, making it a promising tool for early detection and diagnosis in gastroenterology.

On the basis of the results of our experiment (Table 7 and Figure 14), our hypothetical scenario is as follows. When compared to the other four possible iterations, our model achieves the best results overall. Precision values of 0.92, 0.92, 0.93, and 0.95 found in the YOLOv4 model, SSD Mobilenet v2, SSD Mobilenet v2 FPN-Lite, SSD ResNet50 v1 FPN, and SSD Mobilenet v2 FPN-Lite, respectively. The accuracy of the model that we suggested was, on the other hand, 0.99. SSD makes use of anchor prioritization boxes and applies the IoU methodology to achieve a score that is greater than 0.5. They are pre-calculated boxes of a given size, quite similar to the ground-truth boxes used in the beginning. Our model, on the other hand, takes advantage of auto-learning bounding boxes, which helps to improve the accuracy of the algorithm as a whole.

In our particular instance, the model numbers SSD Mobilenet v2, SSD Mobilenet v2 FPN-Lite, SSD ResNet50 v1 FPN, and YOLOv4 have been assigned the recall numbers 0.63, 0.64, 0.66, and 0.92 respectively. The recall value for the model that we suggested was 1. Despite the fact, that YOLOv4 and our model share a number of architectural similarities. The key differentiating factor that contributes to our model's superiority is the incorporation of a focus layer, which boosts the overall performance of our representation.

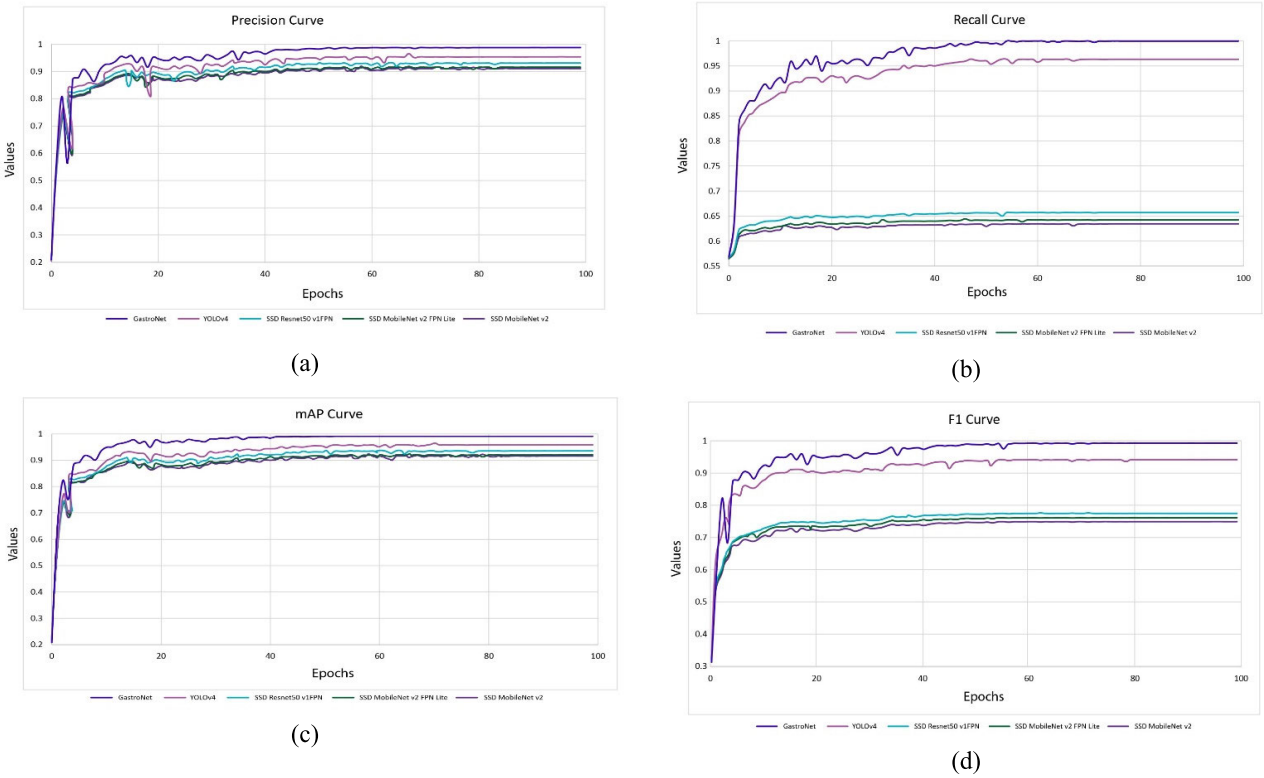

FIGURE 14. Comparison graph of all models where precision is shown in (a), Recall is shown in (b), mAP is shown in (c) and F1 is shown in (d). Additionally GastroNet model is denoted by blue colour line, Yolov4 is denoted by megenta colour, SSD Resnet50v1 FPN is denoted by sky colour, SSD MobileNet v2 FPN lite is denoted by green colour and SSD MobileNet v2 is denoted by purple colour.

TABLE 7. Summarized description of applied models performance comparison.

Method	Backbone	Precision	Recall	mAP @.50IOU	F1	Learning Rate
SSD	MobileNet v2	0.92	0.63	0.92	0.75	0.00145
SSD	MobileNet v2 FPN Lite	0.92	0.64	0.92	0.76	0.00014
SSD	Resnet50 v1FPN	0.93	0.66	0.93	0.77	0.03808
YOLOv4	CSPDarknet-53	0.95	0.92	0.95	0.94	0.00001
Proposed Model	CSPDarknet	0.99	1	0.99	0.99	0.00044

After making a number of changes to the batch size, the epochs, and the training step, the mAP value settled in at the following. The mAP for the SSD Mobilenet v2, SSD Mobilenet v2 FPN-Lite, SSD ResNet50 v1 FPN, and YOLOv4 models are respectively 0.92, 0.92, 0.93, and 0.95. However, the value of the mAP for the model that we suggested was 0.99.

It has come to our attention that YOLOv4 possesses the highest learning rate. Our model, on the other hand, was able to acquire the best possible F1 score because to its remarkable precision and recall value. We get F1 scores of 0.75 for the SSD Mobilenet v2 model, 0.76 for the SSD Mobilenet v2 FPN-Lite model, 0.77 for the SSD ResNet50 v1 FPN model, and 0.94 for the YOLOv4 model. In contrast, the model that we thought would work best had an F1 score of 0.99. CSPDarknet53 is an enhanced version of the Darknet-53 backbone that incorporates the Cross Stage Partial (CSP) module. The CSP module improves information flow and feature representation within the network, enabling better detection performance. It enhances the accuracy and recall of the YOLOv4 model, contributing to its superior precision, recall, mAP, and F1 Score compared to other models in the Table 7. The improved YOLOv4 model introduces additional optimizations, such as PANet, Mish activation, and CIoU loss, to enhance object detection performance. These optimizations improve the

TABLE 8. Performance comparison description with existing work.

Method	Description	Dataset	Precision	Recall	F1	mAP
Wan et al., (2021) [14]	YOLOv5-attention mechanism	Kvasir-SEG Dataset	0.92	0.90	0.91	-
		WCY Dataset	0.91	0.92	0.92	-
Sasmal et al., (2021) [15]	YOLOv4-attention mechanism	Kvsir-SEG and SUN Colonoscopy	0.93	1.00	0.97	-
Cao et al., (2021) [16]	YOLOv3 with feature extraction and fusion module	CVC-CLINIC&ETIS-LARIB	0.93	0.88	0.90	-
Taş et al., (2021) [17]	Faster RCNN-ResNet-101 with SRCNN based pre-processing	ETISLARIB	0.71	0.84	0.77	-
Liu et al., (2019) [43]	SSD-ResNet50	CVC-CLINIC&ETIS-LARIB	0.73	0.80	0.76	-
	SSD-InceptionV3	CVC-CLINIC&ETIS-LARIB	0.74	0.80	0.77	-
	SSD-VGG16	CVC-CLINIC&ETIS-LARIB	0.62	0.76	0.68	-
Gao et al., (2022) [44]	YOLOv5x-CG	Digestive Endoscopy Center, Lingang Hospital, Shanghai Sixth People's Hospital, China	-	-	-	0.92
Proposed Model	**GastroNet**	**Kvasir v1** [32]	**0.99**	**1.00**	**0.99**	**0.99**

accuracy of bounding box regression and overall detection capabilities.

In the proposed GastroNet model, the combination of CSPDarknet as the backbone and the improved YOLOv4 optimizations contributes to its exceptional performance. GastroNet achieves a precision, mAP, F1 score of 0.99 and recall of 1, indicating its ability to accurately detect polyps and abnormal features. The use of CSPDarknet enhances feature representation and information flow, while the improved YOLOv4 optimizations enhance object detection performance, resulting in a highly accurate and reliable model for detecting and recognizing polyps and abnormal features. Overall, the contributions of CSPDarknet53 and the improved YOLOv4 model significantly affect the performance of the proposed GastroNet model, enabling it to achieve state-of-the-art accuracy and precision in polyp and abnormal feature recognition.

X. PERFORMANCE COMPARISON WITH EXISTING WORK

The results obtained from the proposed GastroNet model showcase its superior performance compared to existing methods or benchmarks, as outlined in Table 8. GastroNet outperforms other models in terms of precision, recall, F1 score, and mAP, demonstrating its effectiveness in polyp detection and recognition.

Wan et al. [14] utilized the YOLOv5-attention mechanism on the Kvasir-SEG and WCY datasets. While they achieved good precision and recall values of 0.92 and 0.90 for Kvasir-SEG, and 0.91 and 0.92 for WCY, respectively, GastroNet surpasses these results with precision, recall, and F1 score of 0.99, 1, and 0.99 for the Kvasir v1 dataset.

Sasmal et al. [15] applied the YOLOv4-attention mechanism on the Kvsir-SEG and SUN Colonoscopy datasets, achieving high precision 0.93 and recall 1 values, resulting in an impressive F1 score of 0.97. GastroNet's precision, recall, and F1 score of 0.99, 1, and 0.99, respectively, demonstrate its superiority over their results.

Cao et al. [16] employed YOLOv3 with a feature extraction and fusion module on the CVC-CLINIC&ETIS-LARIB dataset, achieving precision, recall, and F1 score of 0.93, 0.88, and 0.90. GastroNet's significantly higher precision, recall, and F1 score of 0.99, 1, and 0.99 highlight its improved performance.

Taş and Yılmaz [17] utilized Faster RCNN-ResNet-101 with SRCNN-based pre-processing on the ETISLARIB dataset, achieving precision, recall, and F1 score of 0.71, 0.84, and 0.77, respectively. GastroNet's precision, recall, and F1 score of 0.99, 1, and 0.99 indicate its superiority in polyp detection.

Liu et al. [43] employed SSD-ResNet50, SSD-InceptionV3, and SSD-VGG16 on the CVC-CLINIC&ETIS-LARIB dataset, achieving various precision, recall, and F1 scores. GastroNet outperforms all three models with its precision, recall, and F1 score of 0.99, 1, and 0.99.

Gao et al. [44] proposed YOLOv5x-CG for lesion diagnosis, attaining a mAP score of 0.92. In contrast, GastroNet achieves a significantly higher mAP score of 0.99, indicating its superior performance in accurately detecting and localizing abnormalities.

Overall, the comprehensive analysis and comparison highlight the superiority of the GastroNet model. It achieves exceptional precision, recall, F1 score, and mAP, surpassing existing methods and benchmarks. GastroNet's ability to accurately detect and recognize polyps and abnormal features in medical imaging demonstrates its potential for improving diagnostic accuracy and assisting medical professionals in gastroenterological applications.

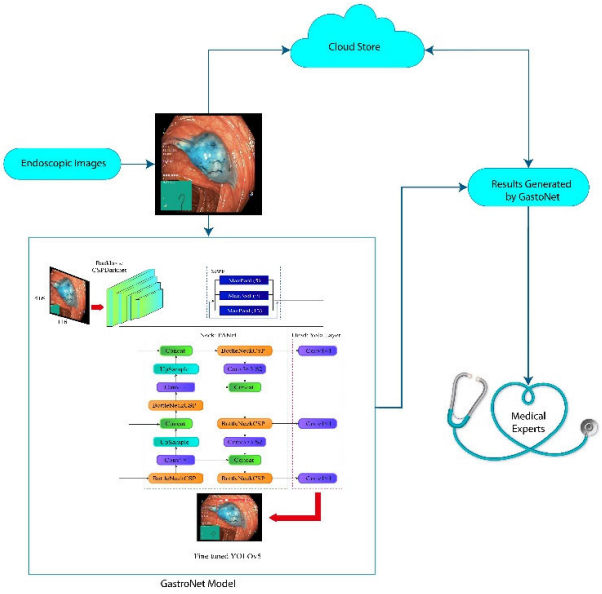

FIGURE 15. The illustration of proposed "GastroNet Diagnostic" application.

XI. PROPOSED APPLICATION

This study proposed the "GastroNet Diagnostic" application, a cloud-based platform that utilizes the GastroNet model for image classification and diagnosis in the field of gastroenterology. GastroNet Diagnostic simplifies the process of image analysis, supports accurate diagnosis, and facilitates efficient collaboration between healthcare providers. Figure 15 shows the illustration of proposed "GastroNet Diagnostic" application.

Key features of the GastroNet Diagnostic application:

- Image Upload and Cloud Storage: Healthcare providers can securely upload patient images, such as endoscopic or radiological images, to the cloud storage within the application. These images are stored in a centralized and secure environment, ensuring accessibility and data integrity.
- Integration with GastroNet Model: GastroNet Diagnostic seamlessly integrates with the powerful GastroNet model to analyze the uploaded images. The GastroNet model, specifically trained for gastroenterology, leverages deep learning techniques to classify and identify specific gastrointestinal conditions or abnormalities present in the images.
- Image Classification and Diagnostic Support: GastroNet Diagnostic processes the uploaded images using the GastroNet model. The model analyzes the images and provides diagnostic support, generating classification results that indicate the presence of abnormalities, specific conditions, or potential areas of concern within the gastrointestinal system.
- Result Communication and Collaboration: The diagnostic results produced by the GastroNet model are promptly communicated to healthcare providers via the application. The platform facilitates seamless collaboration by enabling discussions, annotations, and comments on the results, allowing for interdisciplinary consultations and knowledge-sharing among medical professionals.
- Result Storage and Retrieval: GastroNet Diagnostic securely stores the diagnostic results in the cloud, associating them with the patient's profile or medical record. This ensures that the results are easily accessible for future reference, follow-up visits, or consultations, providing a comprehensive history of the patient's imaging studies and outcomes.
- Privacy and Security: The application prioritizes patient data privacy and security. It adheres to stringent data protection standards, including encryption and access controls, to safeguard patient information throughout the image uploading, analysis, and storage processes.
- Integration with Electronic Health Records (EHR): GastroNet Diagnostic seamlessly integrates with existing EHR systems, allowing healthcare providers to access patient data, medical history, and imaging results within a unified interface. This integration streamlines the workflow and ensures the seamless flow of information across different healthcare IT systems.

GastroNet Diagnostic aims to streamline and enhance the accuracy of image-based diagnostics in gastroenterology using the powerful GastroNet model. By leveraging cloud-based infrastructure and enabling collaboration, the application empowers healthcare professionals with an advanced tool for efficient and reliable image analysis, contributing to improved patient outcomes and collaborative decision-making with affordable and cost-effective image analysis for Gastroenterology.

XII. CONCLUSION

This study concludes by emphasizing the significant contribution of artificial intelligence, specifically the GastroNet model, to the enhancement of early identification and diagnosis of digestive issues, such as gastrointestinal polyps and abnormal characteristics such as esophagitis. Utilizing cutting-edge algorithms and hyperparameter fine-tuning techniques, the GastroNet model demonstrates superior precision, efficiency, and detection capabilities. The study employs numerous backbone networks, including the enhanced YOLOv4 model with CSPdarknet53, MobileNet v2, MobileNet v2 FPN Lite, Resnet50 v1, and MobileNet v2 FPN. The results demonstrate the effectiveness of the GastroNet model, with mean Average Precision (mAP), F1 score, precision, and recall values of 0.99 and 1.00, respectively. As a result of these discoveries, physicians will be better able to recognize and diagnose gastrointestinal disorders and polyps. It effectively reduces false positives and negatives, leading to more accurate treatment plans and better patient outcomes. GastroNet enables timely interventions and tailored management plans, potentially offering less invasive treatment options and better prognoses, especially for gastrointestinal cancers. The model's accurate detection

capabilities streamline the diagnostic process, saving time and resources for physicians to focus on patient care and complex decisions. Utilizing AI in this sector has the potential to enhance early detection efforts, reduce the risk of gastrointestinal cancers, especially esophageal cancer, and improve patient outcomes. This paper presents a novel and beneficial paradigm for clinical endoscopists. However, there is a severe lack of high-quality endoscopic image databases for the identification and classification of gastrointestinal polyps and anomalous features. Because they are more cost-effective, endoscopic images of average quality are more prevalent than those of superior quality. Optimization of a model is only possible with a high-quality dataset. GastroNet has some limitations and potential challenges that need to be addressed. These include scalability issues with large-scale datasets, the interpretability of the model's decisions, the generalizability to different populations and imaging modalities, data availability and quality, and ethical considerations related to patient privacy and data security. Overcoming these challenges requires ongoing research, collaboration, and validation efforts to ensure the reliability and applicability of GastroNet in clinical practice. However, we will continue to examine current issues in the future.

REFERENCES

[1] *Cancer Statistics, 2023—Siegel—2023—CA: A Cancer Journal for Clinicians—Wiley Online Library*. Accessed: May 8, 2023. [Online]. Available: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21763

[2] Cancer.Net. (Jun. 25, 2012). *Esophageal Cancer—Statistics*. Accessed: May 8, 2023. [Online]. Available: https://www.cancer.net/cancer-types/esophageal-cancer/statistics

[3] J. W. van Sandick, J. F. W. M. Bartelsman, J. J. B. van Lanschot, G. N. J. Tytgat, and H. Obertop, "Surveillance of Barrett's oesophagus: Physicians' practices and review of current guidelines," *Eur. J. Gastroenterol. Hepatol.*, vol. 12, no. 1, p. 111, Jan. 2000. Accessed: Aug. 10, 2022. [Online]. Available: https://journals.lww.com/eurojgh/Abstract/2000/12010/Surveillance_of_Barrett_s_oesophagus__physicians_.20.aspx

[4] G. W. Falk, T. M. Ours, and J. E. Richter, "Practice patterns for surveillance of Barrett's esophagus in the United States," *Gastrointestinal Endoscopy*, vol. 52, no. 2, pp. 197–203, Aug. 2000, doi: 10.1067/mge.2000.107728.

[5] C. P. Gross, M. I. Canto, J. Hixson, and N. R. Powe, "Management of Barrett's esophagus: A national study of practice patterns and their cost implications," *Amer. J. Gastroenterol.*, vol. 94, no. 12, pp. 3440–3447, Dec. 1999, doi: 10.1016/S0002-9270(99)00665-6.

[6] A. May, E. Günter, F. Roth, L. Gossner, M. Stolte, M. Vieth, and C. Ell, "Accuracy of staging in early oesophageal cancer using high resolution endoscopy and high resolution endosonography: A comparative, prospective, and blinded trial," *Gut*, vol. 53, no. 5, pp. 634–640, May 2004, doi: 10.1136/gut.2003.029421.

[7] S. B. Ahn, D. S. Han, J. H. Bae, T. J. Byun, J. P. Kim, and C. S. Eun, "The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies," *Gut Liver*, vol. 6, no. 1, pp. 64–70, Jan. 2012, doi: 10.5009/gnl.2012.6.1.64.

[8] F. Yasmin, M. M. Hassan, S. Zaman, S. T. Aung, A. Karim, and S. Azam, "A forecasting prognosis of the monkeypox outbreak based on a comprehensive statistical and regression analysis," *Computation*, vol. 10, no. 10, p. 177, Oct. 2022, doi: 10.3390/computation10100177.

[9] M. M. Hassan, S. Mollick, and F. Yasmin, "An unsupervised cluster-based feature grouping model for early diabetes detection," *Healthcare Anal.*, vol. 2, Nov. 2022, Art. no. 100112, doi: 10.1016/j.health.2022.100112.

[10] N. Mahmud, J. Cohen, K. Tsourides, and T. M. Berzin, "Computer vision and augmented reality in gastrointestinal endoscopy," *Gastroenterol. Rep.*, vol. 3, no. 3, pp. 179–184, Aug. 2015, doi: 10.1093/gastro/gov027.

[11] J. Yu, H. Wang, and M. Chen, "Colonoscopy polyp detection with massive endoscopic images," Feb. 2022, *arXiv:2202.08730*.

[12] M. C. Hoang, K. T. Nguyen, J. Kim, J.-O. Park, and C.-S. Kim, "Automated bowel polyp detection based on actively controlled capsule endoscopy: Feasibility study," *Diagnostics*, vol. 11, no. 10, p. 1878, Oct. 2021, doi: 10.3390/diagnostics11101878.

[13] S. Durak, B. Bayram, T. Bakırman, M. Erkut, M. Doğan, M. Gürtürk, and B. Akpınar, "Deep neural network approaches for detecting gastric polyps in endoscopic images," *Med. Biol. Eng. Comput.*, vol. 59, nos. 7–8, pp. 1563–1574, Aug. 2021, doi: 10.1007/s11517-021-02398-8.

[14] J. Wan, B. Chen, and Y. Yu, "Polyp detection from colorectum images by using attentive YOLOv5," *Diagnostics*, vol. 11, no. 12, p. 2264, Dec. 2021, doi: 10.3390/diagnostics11122264.

[15] P. Sasmal, A. Paul, M. Bhuyan, Y. Iwahori, N. Ogasawara, and K. Kasugai, "An automated framework for detection, localization, and classification of colonic polyp using deep learning," Tech. Rep., 2021.

[16] C. Cao, R. Wang, Y. Yu, H. Zhang, Y. Yu, and C. Sun, "Gastric polyp detection in gastroscopic images using deep neural network," *PLoS ONE*, vol. 16, no. 4, Apr. 2021, Art. no. e0250632, doi: 10.1371/journal.pone.0250632.

[17] M. Taş and B. Yılmaz, "Super resolution convolutional neural network based pre-processing for automatic polyp detection in colonoscopy images," *Comput. Electr. Eng.*, vol. 90, Mar. 2021, Art. no. 106959, doi: 10.1016/j.compeleceng.2020.106959.

[18] A. Ellahyani, I. E. Jaafari, S. Charfi, and M. E. Ansari, "Fine-tuned deep neural networks for polyp detection in colonoscopy images," *Pers. Ubiquitous Comput.*, vol. 27, no. 2, pp. 235–247, Mar. 2022, doi: 10.1007/s00779-021-01660-y.

[19] W. Li, C. Yang, J. Liu, X. Liu, X. Guo, and Y. Yuan, "Joint polyp detection and segmentation with heterogeneous endoscopic data," in *Proc. 3rd Int. Workshop Challenge Comput. Vis. Endoscopy (EndoCV), Co-Located With 18th IEEE Int. Sympo Biomed. Imag. (ISBI)*, Apr. 2021, pp. 69–79. Accessed: Aug. 10, 2022. [Online]. Available: https://scholars.cityu.edu.hk/en/publications/joint-polyp-detection-and-segmentation-with-heterogeneous-endoscopic-data(e0605078-08c6-4c0d-98e6-2e4d8ed50db1).html

[20] W. Wang, J. Tian, C. Zhang, Y. Luo, X. Wang, and J. Li, "An improved deep learning approach and its applications on colonic polyp images detection," *BMC Med. Imag.*, vol. 20, no. 1, p. 83, Jul. 2020, doi: 10.1186/s12880-020-00482-3.

[21] A. Haj-Manouchehri and H. M. Mohammadi, "Polyp detection using CNNs in colonoscopy video," *IET Comput. Vis.*, vol. 14, no. 5, pp. 241–247, Aug. 2020, doi: 10.1049/iet-cvi.2019.0300.

[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," Apr. 2020, *arXiv:2004.10934*. Accessed: Aug. 10, 2022.

[23] Z. Guo, R. Zhang, Q. Li, X. Liu, D. Nemoto, K. Togashi, S. M. I. Niroshana, Y. Shi, and X. Zhu, "Reduce false-positive rate by active learning for automatic polyp detection in colonoscopy videos," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1655–1658, doi: 10.1109/ISBI45749.2020.9098500.

[24] I. Pacal and D. Karaboga, "A robust real-time deep learning based automatic polyp detection system," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104519, doi: 10.1016/j.compbiomed.2021.104519.

[25] C.-P. Tang, K.-H. Chen, and T.-L. Lin, "Computer-aided colon polyp detection on high resolution colonoscopy using transfer learning techniques," *Sensors*, vol. 21, no. 16, p. 5315, Aug. 2021, doi: 10.3390/s21165315.

[26] Z. Shen, R. Fu, C. Lin, and S. Zheng, "COTR: Convolution in transformer network for end to end polyp detection," in *Proc. 7th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2021, pp. 1757–1761, doi: 10.1109/ICCC54389.2021.9674267.

[27] W. S. Liew, T. B. Tang, C.-H. Lin, and C.-K. Lu, "Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches," *Comput. Methods Programs Biomed.*, vol. 206, Jul. 2021, Art. no. 106114, doi: 10.1016/j.cmpb.2021.106114.

[28] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D. T. Dang-Nguyen, M. Lux, P. T. Schmidt, and M. Riegler, "KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proc. 8th ACM Multimedia Syst. Conf.*, New York, NY, USA, Jun. 2017, pp. 164–169, doi: 10.1145/3083187.3083212.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[30] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[31] T. Yu and H. Zhu, "Hyper-parameter optimization: A review of algorithms and applications," Mar. 2020, *arXiv:2003.05689*. Accessed: Aug. 10, 2022.

[32] F. Gaillard, "Epoch (machine learning) | radiology reference article | radiopaedia.org," Radiopaedia, USA, Tech. Rep. 6. Accessed: Aug. 10, 2022. [Online]. Available: https://radiopaedia.org/articles/epoch-machine-learning

[33] Z. Xie, I. Sato, and M. Sugiyama, "Understanding and scheduling weight decay," Sep. 2021, *arXiv:2011.11152*. Accessed: Aug. 10, 2022.

[34] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. 30th Int. Conf. Mach. Learn.*, May 2013, pp. 1139–1147. Accessed: Aug. 10, 2022. [Online]. Available: https://proceedings.mlr.press/v28/sutskever13.html

[35] GitHub. *Releases · Ultralytics/YOLov5*. Accessed: Aug. 10, 2022. [Online]. Available: https://github.com/ultralytics/yolov5/releases

[36] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs," *Sensors*, vol. 22, no. 2, p. 464, Jan. 2022, doi: 10.3390/s22020464.

[37] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "CSPNet: A new backbone that can enhance learning capability of CNN," Nov. 2019, *arXiv:1911.11929*, doi: 10.48550/arXiv.1911.11929.

[38] Ultralytics. (Aug. 10, 2022). *Ultralytics/YOLOv5*. Accessed: Aug. 10, 2022. [Online]. Available: https://github.com/ultralytics/yolov5

[39] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," Sep. 2018, *arXiv:1803.01534*, doi: 10.48550/arXiv.1803.01534.

[40] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Apr. 2018, *arXiv:1804.02767*, doi: 10.48550/arXiv.1804.02767.

[41] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, New York, NY, USA, Jun. 2006, pp. 233–240, doi: 10.1145/1143844.1143874.

[42] T. Abbas, A. Razzaq, M. A. Zia, I. Mumtaz, M. A. Saleem, W. Akbar, M. A. Khan, G. Akhtar, and C. S. Shivachi, "Deep neural networks for automatic flower species localization and recognition," *Comput. Intell. Neurosci.*, vol. 2022, p. e9359353, Apr. 2022, doi: 10.1155/2022/9359353.

[43] M. Liu, J. Jiang, and Z. Wang, "Colonic polyp detection in endoscopic videos with single shot detection based deep convolutional neural network," *IEEE Access*, vol. 7, pp. 75058–75066, 2019, doi: 10.1109/ACCESS.2019.2921027.

[44] J. Gao, Q. Xiong, C. Yu, and G. Qu, "White-light endoscopic colorectal lesion detection based on improved YOLOv5," *Comput. Math. Methods Med.*, vol. 2022, Jan. 2022, Art. no. e9508004, doi: 10.1155/2022/9508004.

[45] S. Anari, N. T. Sarshar, N. Mahjoori, S. Dorosti, and A. Rezaie, "Review of deep learning approaches for thyroid cancer diagnosis," *Math. Problems Eng.*, vol. 2022, Aug. 2022, Art. no. e5052435, doi: 10.1155/2022/5052435.

[46] *Brain Tumor Segmentation Based on Deep Learning and an Attention Mechanism Using MRI Multi-Modalities Brain Images | Scientific Reports*. Accessed: Jun. 4, 2023. [Online]. Available: https://www.nature.com/articles/s41598-021-90428-8

[47] R. Ranjbarzadeh, S. J. Ghoushchi, N. T. Sarshar, E. B. Tirkolaee, S. S. Ali, T. Kumar, and M. Bendechache, "ME-CCNN: Multi-encoded images and a cascade convolutional neural network for breast tumor segmentation and recognition," *Artif. Intell. Rev.*, vol. 56, no. 9, pp. 10099–10136, Feb. 2023, doi: 10.1007/s10462-023-10426-2.

[48] N. T. Sarshar, R. Ranjbarzadeh, S. J. Ghoushchi, G. G. de Oliveira, S. Anari, M. Parhizkar, and M. Bendechache, "Glioma brain tumor segmentation in four MRI modalities using a convolutional neural network and based on a transfer learning method," in *Proc. 7th Brazilian Technol. Symp. (BTSym)*, in Smart Innovation, Systems and Technologies, Y. Iano, O. Saotome, K. Vásquez, G. L. C. C. Pezzuto, R. Arthur, and G. G. de Oliveira, Eds. Cham, Switzerland: Springer, 2023, pp. 386–402, doi: 10.1007/978-3-031-04435-9_39.

[49] T. Yan, Y. Y. Qin, P. K. Wong, H. Ren, C. H. Wong, L. Yao, Y. Hu, C. I. Chan, S. Gao, and P. P. Chan, "Semantic segmentation of gastric polyps in endoscopic images based on convolutional neural networks and an integrated evaluation approach," *Bioengineering*, vol. 10, no. 7, p. 806, Jul. 2023, doi: 10.3390/bioengineering10070806.

[50] I. Varela-Rey, D. de la Iglesia, A. S. Bruno-Ruz, R. Mejuto-Fernández, L. Monteserín-Ron, J. López-Diaz, P. García-Salom, A. González-Cantalapiedra, J. M. Brea, R. Piña-Márquez, V. Díaz-Tomé, M. González-Barcia, I. Zarra-Ferro, C. Mondelo-García, J. E. Dominguez-Muñoz, F. J. Otero-Espinar, and A. Fernández-Ferreiro, "Design and biopharmaceutical preclinical characterisation of a new thermosensitive hydrogel for the removal of gastric polyps," *Int. J. Pharmaceutics*, vol. 635, Mar. 2023, Art. no. 122706, doi: 10.1016/j.ijpharm.2023.122706.

[51] S. Massironi, A. Elvevi, C. Gallo, A. Laffusa, A. Tortorella, and P. Invernizzi, "Exploring the spectrum of incidental gastric polyps in autoimmune gastritis," *Digestive Liver Disease*, vol. 55, no. 9, pp. 1201–1207, Sep. 2023, doi: 10.1016/j.dld.2023.02.008.

[52] J. Huang, S. N. Saw, T. He, W. Feng, and C. K. Loo, "A masked graph neural network model for real-time gastric polyp detection in Healthcare 4.0," *J. Ind. Inf. Integr.*, vol. 34, Aug. 2023, Art. no. 100467, doi: 10.1016/j.jii.2023.100467.

FARHANA YASMIN received the B.Sc. degree from the Department of Computer Science and Engineering, Eastern University, Bangladesh. She is currently pursuing the M.Sc. degree in computer application and technology with Changzhou University, China. She is an expert in computer vision, biomedical engineering, and data science, all of which fall under the broad umbrella of artificial intelligence (AI). Two of her primary research interests are in predictive analysis and the development of expert systems. She works with scientists from all around the world to broaden her horizons and increase the depth of her expertise.

MD. MEHEDI HASSAN (Member, IEEE) received the B.Sc. degree in computer science and engineering from North Western University, Khulna, Bangladesh, in 2022, where he excelled in his studies and demonstrated a strong aptitude for research. He is currently pursuing the Master of Science (M.Sc.) degree in computer science and engineering with Khulna University, Khulna. He is a dedicated and accomplished researcher. As the Founder and the CEO of The Virtual BD IT Firm and VRD Research Laboratory, Bangladesh, he has established himself as a highly respected leader in the fields of biomedical engineering, data science, and expert systems. His research interests are broad and include important human diseases, such as oncology, cancer, and hepatitis, as well as human behavior analysis and mental health. He is highly skilled in association rule mining, predictive analysis, machine learning, and data analysis, with a particular focus on the biomedical sciences. As a Young Researcher, he has published 31 articles in various international top journals and conferences, which is a remarkable achievement. Additionally, he serves as a reviewer for 22 prestigious journals. He has filed more than three patents out of which two are granted to his name. He is a member of the prestigious Institute of Electrical and Electronics Engineers (IEEE). His work has been well-received by the research community and has significantly contributed to the advancement of knowledge in his field. Overall, he is a highly motivated and skilled researcher with a strong commitment to improving human health and well-being through cutting-edge scientific research. His accomplishments to date are impressive, and his potential for future contributions to his field is very promising.

MAHADE HASAN received the B.Sc. degree in computer science and engineering from Eastern University, Bangladesh. He is currently pursuing the M.Sc. degree in computer application and technology program with Changzhou University, China. He is an expert in all three of the artificial intelligence subfields of computer vision, image processing, and data analysis. His primary areas of interest in research are expert system design and image analysis. He has written substantially on both subjects.

SADIKA ZAMAN is currently pursuing the bachelor's degree in computer science and engineering with North Western University, Khulna, Bangladesh. In addition to her academic pursuits, she is actively involved in resolving social issues and is committed to having a positive impact on society. She is a youthful Researcher and has already made significant contributions to her discipline, publishing nine articles in prestigious international journals and conferences. Her research interests reside in the field of biomedical engineering, where she focuses on devising innovative solutions to enhance the health and well-being of humans. She is a creative and analytical thinker with strong problem-solving and attention-to-detail abilities. She is exceptionally proficient in numerous research methodologies, including data analysis, machine learning, and statistics. Her aptitude and commitment as a Researcher are exemplified by her ability to synthesize complex information and derive meaningful insights from her research.

ANUPAM KUMAR BAIRAGI (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science and engineering from Khulna University (KU), Bangladesh, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea. He is a Professor with the Computer Science and Engineering Discipline, KU. He has authored and coauthored around 60 publications, including refereed IEEE/ACM journals and conference papers. His research interests include wireless resource management in 5G and beyond, healthcare, the IIoT, cooperative communication, and game theory. He has served as a technical program committee member at different international conferences.

WALID EL-SHAFAI (Senior Member, IEEE) was born in Alexandria, Egypt. He received the B.Sc. degree (Hons.) in electronics and electrical communication engineering from the Faculty of Electronic Engineering (FEE), Menoufia University, Menouf, Egypt, in 2008, the M.Sc. degree from the Egypt–Japan University of Science and Technology (E-JUST), in 2012, and the Ph.D. degree from the Faculty of Electronic Engineering, Menoufia University, in 2019. Since January 2021, he has been a Postdoctoral Research Fellow with the Security Engineering Laboratory (SEL), Prince Sultan University (PSU), Riyadh, Saudi Arabia. He is currently a Lecturer and an Assistant Professor with the Department of Electronics and Communication Engineering (ECE), FEE, Menoufia University. His research interests include wireless mobile and multimedia communications systems, image and video signal processing, efficient 2D video/3D multi-view video coding, multi-view video plus depth coding, 3D multi-view video coding and transmission, quality of service and experience, digital communication techniques, cognitive radio networks, adaptive filters design, 3D video watermarking, steganography, encryption, error resilience and concealment algorithms for H.264/AVC, H.264/MVC, and H.265/HEVC video codecs standards, cognitive cryptography, medical image processing, speech processing, security algorithms, software-defined networks, the Internet of Things, medical diagnoses applications, FPGA implementations for signal processing algorithms and communication systems, cancellable biometrics and pattern recognition, image and video magnification, artificial intelligence for signal processing algorithms and communication systems, modulation identification and classification, image and video super-resolution and denoising, cybersecurity applications, malware and ransomware detection and analysis, deep learning in signal processing, and communication systems applications. He also serves as a reviewer for several international journals.

HASSAN FOUAD received the B.Sc. and M.Sc. degrees in mechanical engineering from Helwan University, Egypt, in 1990 and 1996, respectively, and the joint Ph.D. degree in biomedical polymers from the University of Leeds, U.K., and Helwan University. He was a Research Scientist with the Vienna University of Technology. He is currently a Full Professor of biomedical engineering, biomaterials, and tissue engineering with King Saud University and Helwan University.

YANG CHANG CHUN is a Professor with the School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu, China. His current research interests include database systems, machine learning, recommendation systems, and data mining.