IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 June 2024, accepted 31 July 2024, date of publication 5 August 2024, date of current version 26 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3438799

RESEARCH ARTICLE

Gastro Intestinal Disease Classification Using Hierarchical Spatio Pyramid TranfoNet With PitTree Fusion and Efficient-CondConv SwishNet

V. SHARMILA, AND S. GEETHA, (Senior Member, IEEE)
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India
Corresponding author: S. Geetha (geetha.s@vit.ac.in)

This work was supported in part by Vellore Institute of Technology, Chennai, India.

ABSTRACT Early detection of Gastrointestinal (GI) tract diseases is essential for effective healthcare management, treatment, and prevention, ultimately lowering morbidity and mortality rates worldwide. Current classification models lack spatial feature arrangement consideration, diminishing discriminative power and leading to misdiagnosis of esophagitis and ulcerative colitis due to overlapping visual characteristics with other GI diseases. Hence, a novel Hierarchical Spatio Pyramid TranfoNet featuring a Spatial Transformer Network (STN) with spatial pyramid pooling is introduced, which enhances discriminative power in distinguishing between overlapping disease characteristics. Enhancing classification models for Dyed Lifted Polyps (DLP) and Dyed Resection Margins (DRM) in endoscopy images is critical for precise gastrointestinal diagnosis, tackling challenges posed by spatial complexity and inter-class confounders. Hence, a novel PitTree Fusion Algorithm, combining Minimum Spanning Tree (MST) analysis and Kudo's pit pattern analysis is introduced to accurately locate and differentiate normal tissue from dyed regions like DLPs and DRMs in endoscopy images. Then, a novel Efficient-CondConv SwishNet is introduced to enhance GI disease classification by extracting informative features from endoscopic images, utilizing EfficientNet-CondConv with Swish activation. After classification, heatmaps highlighting influential regions are produced via gradient-weighted class activation mapping, or Grad-CAM, which provides information about classification decisions. The proposed Hierarchical Spatio Pyramid TranfoNet with PitTree Fusion and EfficientNet-CondConv SwishNet achieved a classification accuracy of 98.2%. The proposed framework is tested on 8000 images using the Kvasir dataset, which is publicly available in Kaggle, consisting of eight classes. The results show that the suggested model outperforms the current models showing increased accuracy, precision, recall, sensitivity, specificity, F1 score, and reduced loss rate.

INDEX TERMS Dyed lifted polyps, dyed resection margins, gastrointestinal tract disease, grad-CAM, minimum spanning tree, spatial transformer, swish activation.

I. INTRODUCTION

Food and waste products are ingested, broken down, absorbed, and expelled by the human gastrointestinal (GI) tract, also known as the digestive system. It is a sophisticated and essential organ system. The GI tract, which is comprise of the mouth, esophagus, stomach, small intestine, large intestine (colon), rectum, and anus, is crucial for preserving fluid balance, nutritional balance, and metabolic homeostasis in general. Nevertheless, a wide range of illnesses and conditions can affect this complex system, compromising both general health and its capability to function. A broad spectrum of conditions, from benign to life threatening, fall under the category of gastrointestinal diseases. These conditions

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

include peptic ulcers, Gastro Esophageal Reflux Disease (GERD), inflammatory bowel diseases (like Crohn's disease and ulcerative colitis), gastrointestinal cancers, infectious gastroenteritis, and functional GI disorders (like irritable bowel syndrome). Numerous factors, including the origin, anatomical location, pathophysiology, and clinical symptoms of these disorders, can be used to classify them [1], [2], [3].

Esophagitis and ulcerative colitis are two different disorders of the gastrointestinal system, each with special traits and effects on the well-being of the patient. Esophagitis is the medical term for inflammation of the esophagus, which is frequently brought on by illnesses including GERD, infections, or specific drugs. Chest pain, difficulty swallowing, and heartburn are common symptoms. If left untreated, esophagitis can result in Barrett's esophagus, a condition that is a precursor to esophageal cancer, or esophageal strictures. Conversely, ulcerative colitis is a subtype of inflammatory bowel disease (IBD) that causes ulcers in the colon and rectum as well as persistent inflammation. Abdominal pain, diarrhea, rectal bleeding, and weight loss are signs of ulcerative colitis. Furthermore, one of the most important tasks in gastrointestinal endoscopy is to recognize Dyed Lifted Polyps (DLPs) and Dyed Resection Margins (DRMs) in endoscopic images. This is especially true during techniques like chromoendoscopy, when dyes are employed to improve tissue visualization [4], [5], [6], [7].

DLPs are elevated lesions that may represent precancerous or cancerous polyps, while DRMs indicate the margins of tissue that have been resected during endoscopic submucosal dissection (ESD) or endoscopic mucosal resection (EMR) procedures. Reducing the chance of missing lesions or recurrence and guaranteeing total removal of aberrant tissue depend on accurately identifying these traits. Deep learning algorithms-based Computer Aided Detection (CAD) systems have demonstrated potential in automatically identifying DLPs and DRMs in endoscopic pictures, helping endoscopists recognize lesions in real time and increasing diagnostic precision. The classification of these gastrointestinal tract disorders has seen the emergence of strong tools based on deep learning, which hold promise for quick and precise diagnosis. These methods automatically extract information from medical pictures, such as endoscopic movies or histological slides, and categorize them into several illness classifications by utilizing sophisticated neural network designs. Because Convolutional Neural Networks (CNNs) can capture spatial hierarchies of data, they are frequently used to identify subtle patterns that may indicate a variety of gastrointestinal disorders [8], [9], [10], [11].

Sequential data analysis is also performed using Recurrent Neural Networks (RNNs) and their derivatives, such as Long Short Term Memory (LSTM) networks, to analyze dynamic changes in gastrointestinal pictures over time. By utilizing pre-trained networks on sizable datasets, transfer learning approaches improve the effectiveness of deep learning models even more and get around the drawbacks of a lack of labelled medical data. Deep learning-based categorization techniques for gastrointestinal tract disorders have various drawbacks despite their enormous potential. Large annotated datasets are essential for medical imaging, but they are frequently hard to come by because of privacy issues, restrictions on data access, and the high expense of having skilled physicians annotate data. Insufficient diversity in datasets may result in overfitting and restricted model generalization to actual clinical scenarios. Moreover, deep learning models are occasionally viewed as ''black boxes,'' which makes it challenging to comprehend the decisions they make. This is particularly problematic in vital healthcare applications, where explainability is essential to winning both patients' and physicians' trust. Furthermore, these techniques could perform poorly in identifying uncommon or novel illness patterns that are not well-recognized in the training set [12], [13], [14], [15]. Thus, the section highlights the potential of deep learning techniques and emphasizes the importance of early identification and accurate diagnosis of gastrointestinal disorders.

The manuscript is structured as follows: Literature review in the field of gastrointestinal detection is presented in Section II; Section III below explains the methodology of the proposed Hierarchical Spatio Pyramid TranfoNet with PitTree Fusion and Efficient-CondConv SwishNet model, its working process and its significance; Section IV presents the experimental results and the findings of the proposed model; Section V concludes the work indicating future works.

II. LITERATURE SURVEY

Hosain et al. [16] used a transfer learning technique based on the DenseNet201 and Vision Transformer architecture to distinguish between normal colon images and three gastrointestinal disorders: ulcerative colitis, polyps, and esophagitis. The top layer of DenseNet201 was eliminated because there were four classes in this classification task. It was then replaced with a dense layer that contained 512 neurons with Relu activation function and an output layer that contained four neurons with softmax activation function. A softmax activation function was employed in the output layer because of the need for multiclass classification. Images were split into patches using pretrained supervised machine learning, and those patches were subsequently given keys or tokens. However, the scalability and efficiency of the models employed have been influenced by the resource constraints.

Khan et al. [17] revealed a fully automated method for identifying stomach illnesses that relied on the combination and selection of deep learning features. By manually assigning ulcer images, it provided a saliency-based ulcer identification approach. The previously learned VGG16 deep learning model was then retrained using transfer learning, which assembled from two consecutive completely linked layers using an array-based technique. In addition, the mean value-based fitness function and the PSO metaheuristic were used to choose the top candidates. Ultimately, Cubic SVM was used to identify the selected individuals. However, the

current approaches' integration of manually created and CNN characteristics lengthens the system's overall execution time, which affects its speed and efficiency. Haile et al. [18] proposed a neural network model for the diagnosis of gastrointestinal disorders that was built by linking the extracted features of the VGGNet and InceptionNet networks. Training VGGNet and InceptionNet, two deep convolutional neural networks, enabled the extraction of features from the supplied endoscopic images. After being extracted, support vector machines, Random Forest, k-Nearest Neighbor, Softmax, and other machine learning classification methods were used to concatenate and classify these features. SVM, or support vector machines, was one of these techniques that produced superior outcomes. However, it relied on a limited set of imaging features extracted from endoscopy images and thus faced difficulties in distinguishing between different gastrointestinal diseases. Obayya et al. [19] created the Modified Salp Swarm Algorithm for Endoscopic Image-based Classification of Gastrointestinal Tract Disease (MSSADL-GITDC), whose foundation is deep learning. The primary goal of the MSSADL-GITDC strategy was to use the median filtering (MF) technique for image smoothing in order to evaluate WCE images for GIT classification. The class attention layer (CAL) in the outlined MSSADL-GITDC strategy modified the enhanced capsule network (CapsNet) model for extracting features. A Deep Belief Network with an Extreme Learning Machine (DBNELM) was used for GIT classification. Finally, backpropagation was used to supervise and refine the DBN-ELM model. However, the large number of layers and parameters increase the risk of overfitting, especially if not properly regularized or validated. Mohapatra et al. [20] presented a system that used a CNN technique and hybrid EWT to diagnose GI diseases from endoscopic images. Before the dataset was subjected to additional processing, a few steps were taken to help get it ready, such as image scaling, image pre-processing, and image augmentation. The following step involved using EWT as an image feature pattern extractor, which assisted in breaking down the images into their IMF. An input image was initially categorized as abnormal or normal in the first level; only the abnormal class image was subsequently classified into a particular illness class in the second level. However, the performance of EWT-based feature extraction was limited by its ability to represent complex spatial patterns or textures present in endoscopic images, potentially affecting the model's discriminative power. Ramamurthy et al. [21] suggested an innovative approach that used convolutional neural networks (CNN) to focus on feature mining for the categorization of endoscopic images. The model that was being given was constructed by merging a specially designed CNN architecture called Effimix with a cutting-edge architecture, namely EfficientNet B0. For accurate gastrointestinal disease classification, the squeezing and excitation layers and the self-normalising activation layers were combined in the suggested Effimix model. Experiments conducted on the HyperKvasir dataset validate that the suggested architecture was efficient in classifying endoscopic images. However, the model struggles with high level of complexity, which makes the model less interpretable. Nass et al. [22] suggested a unique categorization scheme based on a commonly used surgical instrument, called adverse events in GI endoscopy (AGREE). For endoscopy, the Clavien-Dindo classification of surgical adverse events was modified. To confirm the unique categorization, ten pairwise comparisons were made to determine whether the severity of adverse events (AEs) reported by ten endoscopists, ten endoscopy nurses, and ten patients matched the AGREE classification's severity grading. Additionally, the relationship between the American Society for Gastrointestinal Endoscopy (ASGE) classification and the AGREE classification was assessed. Using a global survey, the tolerability of the AGREE classification was assessed. However, the AE classification does not include a subjective opinion. Fati et al. [23] proposed various multi-methodologies for detecting and classifying features. The first system used hybrid features extraction using three algorithms: gray level co-occurrence matrix, local binary pattern, and fuzzy color histogram. In second system, the pre-trained CNN models, such as AlexNet and GoogLeNet, were employed. These models were founded on the deep feature map extraction and their remarkably accurate categorization. The third used a hybrid technique with SVM for deep feature map classification and CNN models for feature map extraction. The hybrid characteristics of CNN models and GLCM, LBP, and FCH algorithms were the basis for the artificial neural network and FFNN employed in the fourth system. However, the CNN models were able to extract more dimensions of deep feature maps. Iqbal et al. [24] suggested a customized DCNN architecture to effectively recognize anomalies in the human gastrointestinal system from endoscopic images. To increase efficiency and performance, it was developed with numerous pathways, different image resolutions, and multiple convolutional layers. The Kvasir dataset's specificity, recall, AUROC, and other metrics were provided as the outcomes of deep learning-based approach. Thus, it provided a novel and feasible means of reducing time and effort while also expediting and organizing the categorization of gastrointestinal anomalies in humans. However, the proposed method suffers with computational complexity. Sharma et al. [25] provided a method for GI tract organs segmentation, large intestine, and small intestine into segments to help radiologists treat cancer patients more rapidly and precisely. The U-Net model was created from the scratch and was used to more effectively extract local features by segmenting tiny images. Additionally, the U-Net topology was supported by six transfer-learning models such as Inception V3, VGG19, ResNet50, InceptionResNetV2, DenseNet121, and EfficientNet B0. The proposed model was examined using IoU, dice coefficient, and model loss. However, the difference in input image size affects the models performance.

From the analysis, it is determined that in [16], the scalability and efficiency of the models employed have been influenced by the resource constraints, in [17], the execution time is high, which affects model's speed and efficiency and [18] faces difficulties in distinguishing between different gastrointestinal diseases. In [19], the large number of layers and parameters increase the risk of overfitting, in [20], the performance is limited by its ability to represent complex spatial patterns and [21] struggles with high level of complexity, which makes the model less interpretable. In [22], the AE classification does not include a subjective opinion, in [23], the size of deep feature maps that CNN models were able to extract was greater, [24] suffers with computational complexity, and in [25], the difference in input image size affects the models performance. Hence, there is a need to develop a novel Gastrointestinal tract disease classification model to solve the problems occurred in existing techniques.

A. MOTIVATION FOR THE RESEARCH

GI tract disorders affects millions of people globally and result in high rates of morbidity and mortality, placing a heavy cost on global healthcare systems. Timely and accurate identification of these diseases is crucial for operative management, treatment, and inhibition strategies. Detecting and distinguishing between esophagitis and ulcerative colitis, as well as identifying DLPs and DRMs in endoscopy images, holds paramount importance in gastrointestinal healthcare. Esophagitis is often mistaken for ulcerative colitis or vice versa leading to incorrect diagnoses and inappropriate treatment strategies, impacting patient outcomes. The visual characteristics of esophagitis and ulcerative colitis, such as mucosal inflammation, ulceration, and tissue damage, overlap with other gastrointestinal tract diseases, comprising Crohn's disease, gastroesophageal reflux disease (GERD), and infectious colitis. Distinguishing between these conditions based on imaging findings alone is challenging for the existing classification models as most of them treat individual pixels or regions of interest in isolation i.e., each pixel's color, intensity, or texture features are analysed independently without considering how these features relate to nearby pixels or the overall spatial arrangement of features within the image leading to reduced discriminative power.

Improving the effectiveness of classification models in identifying DLPs and DRMs in endoscopy images is important for enhancing the accuracy and efficiency of gastrointestinal diagnostics and therapeutic measures. Dyed lifted polyps and resection margins often occur within complex spatial contexts, such as anatomical folds, creases, or areas of tissue overlap. Existing classification models struggle to accurately localize and distinguish between the normal tissues from the dyed regions within such complicated anatomical sites. This is because the dye pooling in anatomical crevices and the artifacts from dyed tissue manipulation resemble genuine dyed regions thus the DLPs and DRMs share the same visual characteristics called Inter-class Confounders with other anatomical structures or pathological conditions, leading to potential confusion and misclassification by existing classification models.

Hence, there is a need for a novel deep-learning based gastrointestinal tract disease classification model to distinguish between esophagitis and ulcerative colitis and identify DLPs and DRMs in endoscopy images.

B. MAJOR CONTRIBUTIONS OF THIS RESEARCH

In summary, this paper makes the following major research contributions:

- To address the challenges posed by overlapping visual characteristics of gastrointestinal diseases, a novel Hierarchical Spatio Pyramid TranfoNet is presented, which focus on relevant informative regions and adaptively adjust to different spatial configurations.
- To accurately localize and distinguish between normal tissue and dyed regions, such as DLPs and DRMs, a novel PitTree Fusion Algorithm is introduced, which improves the accuracy of classification.
- To efficiently classify GI diseases, a novel Efficient-CondConv SwishNet is proposed, which permits the network to capture more informative and discriminative features from endoscopic images of the GI tract.

III. HIERARCHICAL SPATIO PYRAMID TranfoNet WITH PitTree FUSION AND EFFICIENT-CondConv SwishNet

Gastrointestinal (GI) tract diseases, affecting millions globally, require timely detection for effective management, treatment, and prevention. Identifying esophagitis, ulcerative colitis, DLPs, and DRMs in endoscopy images is crucial in gastrointestinal healthcare. Hence, a novel Advanced GI Disease Classification using Hierarchical Spatio Pyramid TranfoNet with PitTree Fusion and Efficient-CondConv SwishNet is proposed to effectively distinguish between esophagitis and ulcerative colitis and improve the identification accuracy of DLPs and DRMs in endoscopy images. Here, a novel Hierarchical Spatio Pyramid TranfoNet, utilizing a Spatial Transformer Network (STN) with spatial pyramid pooling in the encoding layer is developed to address challenges posed by overlapping visual characteristics in gastrointestinal diseases, which enables feature extraction at different granularity levels, allowing for a comprehensive understanding of feature relationships across different spatial scales. Then, a new PitTree Fusion Algorithm is introduced, which combines Minimum Spanning Tree (MST) analysis in conjunction with Kudo's pit pattern analysis to accurately distinguish between normal tissue and dyed regions in endoscopy images. MST analysis connects anatomical features, while Kudo's analysis captures abnormal pit patterns, identifying neoplastic lesions and polyps as DLPs and irregular pit patterns as DRMs. Finally, a novel approach Efficient-CondConv SwishNet is introduced for classifying GI diseases based on extracted features. This EfficientNet-CondConv with Swish activation function enhances feature capture from endoscopic images. CondConv efficiently extracts features, promotes reuse, and

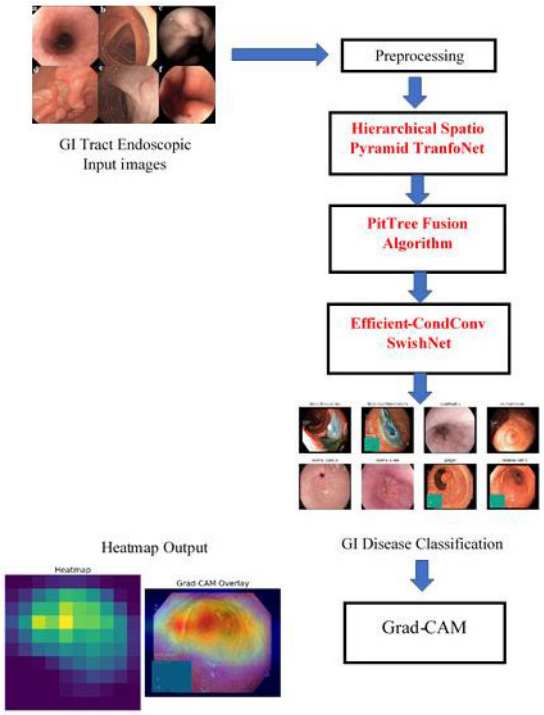

FIGURE 1. The Functional diagram of the proposed framework.

uses swish activation for complex relationship learning, ensuring accurate classification. Post-classification, a Grad-CAM produces heatmap, visualizes influential areas, offering insight into classification factors.

The functional diagram of the suggested model is shown in Figure 1. Firstly, the input images from the dataset are preprocessed then send to the Hierarchical Spatio Pyramid TranfoNet to address the challenges posed by overlapping visual characteristics of gastrointestinal diseases. The output from this network is sent as input to the PitTree Fusion Algorithm, which accurately localizes and distinguishes between normal tissue and dyed regions. Further, Efficient-CondConv SwishNet is used for accurate classification of the diseases. Lastly, Grad-CAM is used to create a heatmap that shows the sections of the source image that most influenced the classification decision made by the model. The following subsections provide the clear description of each of these elements in the proposed model.

A. HIERARCHICAL SPATIO PYRAMID TranfoNet

An inventive approach called the Hierarchical Spatio Pyramid TranfoNet was created to address the difficulties caused by overlapping visual features in GI disorders such as ulcerative colitis and esophagitis. Accurate classification of these disorders is challenging since their visual characteristics in endoscopic images are frequently comparable. In the encoding layer of the model, a Spatial Transformer Network (STN) with spatial pyramid pooling is incorporated to overcome this problem.

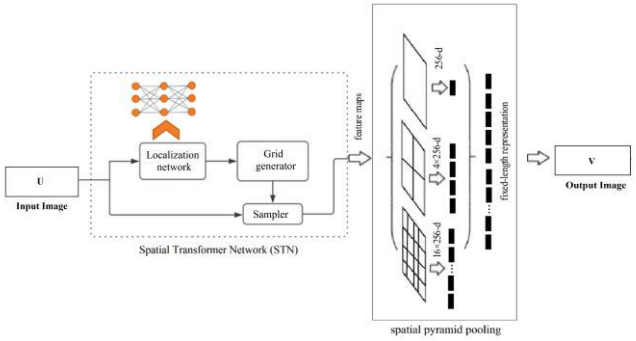

FIGURE 2. Hierarchical Spatio Pyramid TranfoNet of the proposed model.

Figure 2 shows the Hierarchical Spatio Pyramid TranfoNet of the proposed model. The source image is sent to the STN network, which provides a learnable mechanism to spatially transform input images and then send them to the Spatial Pyramid Pooling, which divides the input images (feature maps) 'U' into multiple spatial bins at different scales and sends the output image, 'V'.

The crucial part of the model is the STN, which provides a mechanism for learning to spatially transform input images. This allows the model to concentrate on relevant informative regions within the image and adaptively adjust to different spatial configurations. STN basically aids in the capacity of the model to dynamically resize and rearrange the input images so that it is able to efficiently extract the most discriminative features for classification. The STN generates as its output a localization of the input image. This output is fed into the transformer encoder network after converting it into patches. There are three parts in the localization module: (i) localization network, (ii) grid generator, and (iii) sampler. The localization network generates θ parameters, which are learned as affine transforms for input 'U' with width W, height H, and channel C. The affine transformation is given by equation (1).

$$\begin{pmatrix} x_j^s \\ y_j^s \end{pmatrix} = T_\theta\left(G_j\right) = A_\theta \begin{pmatrix} x_j^t \\ y_j^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_j^t \\ y_j^t \\ 1 \end{pmatrix} \tag{1}$$

In addition, the encoding layer of the STN uses spatial pyramid pooling. Using this method, the input images (or feature maps) are divided into several spatial bins at various scales. By doing this, the network is able to extract features at different granularities, ranging from fine details to more comprehensive contextual data. Equation (2) codes the spatial pyramid pooling operation,

$$SPP(I) = [max_{pool}(I_1), max_{pool}(I_2), \ldots, max_{pool}\,(I_n) \tag{2}$$

where, I is the input feature map and [max]_pool is the max pooling operation. This hierarchical representation of spatial context enables the model to consider not only individual pixels or regions but also their relationships and interactions

Algorithm 1 Hierarchical Spatio Pyramid TranfoNet

Input : Endoscopic images
Output : Transformed and pooled feature maps

1: Apply STN to input images to learn spatial transformations.
2: Perform spatial pyramid pooling on transformed images to extract hierarchical features.
3: Train the model using the hierarchical features.
4: Transform input images using learned transforma tions and perform spatial pyramid pooling.
5: The model distinguishes between overlapping visual characteristics of gastrointestinal diseases

across different spatial scales. As a result, the model gains a more comprehensive understanding of the overall spatial arrangement of features within the image.

By employing an STN with spatial pyramid pooling, the model can learn to spatially transform input images to focus on the informative regions relevant to distinguishing between these diseases. The hierarchical representation of spatial context enables the model to capture not only individual pixels or regions but also their relationships and interactions across different spatial scales. This comprehensive understanding of spatial arrangements aids in differentiating between diseases with overlapping visual features. Algorithm for the proposed Hierarchical Spatio Pyramid TranfoNet method is shown in Algorithm 1 below,

Thus, the Hierarchical Spatio Pyramid TranfoNet improves the discriminative capacity of the model by including the STN with spatial pyramid pooling into the model architecture. It improves the accurateness of disease classification by helping the model differentiate more clearly between the overlapping visual traits of GI diseases. This is accomplished by efficiently capturing and utilizing spatial data at multiple scales, which results in a more reliable representation of the underlying features present in the endoscopic images. Next, a novel PitTree Fusion Algorithm is introduced to distinguish between normal tissue and dyed regions, which is explained in the next section.

B. PitTree FUSION ALGORITHM

The PitTree Fusion Algorithm is a novel method developed to solve the problems associated with precisely localizing and differentiating between dyed regions and normal tissue inside complex anatomical sites in endoscopic images. These dyed regions, such as DLPs and DRMs, pose particular difficulties due to their subtle visual differences and the presence of confounding factors. To overcome these challenges, the algorithm employs a combination of two techniques: Minimum Spanning Tree (MST) analysis and Kudo's pit pattern analysis. This is expressed general in equation (3) as follows,

$$PitTree\ Algorithm = MST(G, W) Kudo's\ Analysis(P) \quad (3)$$

where, G represents the graph structure of the pits, W denotes the weights associated with the edges in the graph, reflecting the importance of connections between pits, and P signifies the pit patterns identified within the endoscopic images.

First, the structural characteristics found in the mucosal surface of the endoscopic images are captured using the MST analysis. MST minimizes the total edge weight while constructing a tree structure that links every region in the image. The MST serves to identify regions of interest (ROI) that are closely connected or exhibit abnormal patterns, which correspond to dyed regions such as DLPs or DRMs. The MST analysis aids in the identification of possible regions of pathology or tissue manipulation by concentrating on the structural organization of the mucosal surface.

Kudo's pit pattern analysis is then used on the selected ROIs that the MST identified, for evaluating the morphological appearance of the mucosal surface using Kudo's pit pattern analysis. It involves identifying particular glandular structures or pit patterns that point to pathological changes. Kudo's pit pattern analysis aids in this context by enabling the distinction of various dyed sections. Specifically, neoplastic lesions, adenomatous, and hyperplastic polyps are characterized by certain pit patterns and are classified as DLPs. On the other hand, irregular, disrupted, or distorted pit patterns, indicative of pathological changes or previous tissue resection, are identified as DRMs.

Hence, by integrating MST analysis with Kudo's pit pattern analysis, the algorithm combines information about structural organization (from MST) with morphological appearance (from pit patterns) to differentiate between inter-class confounders and genuine dyed tissues. MST analysis helps identify regions of interest based on anatomical structures, while Kudo's pit pattern analysis provides insights into the morphological characteristics of these regions, enabling the algorithm to distinguish between normal tissue and dyed regions accurately. Algorithm 2 below illustrates the algorithm for the suggested PitTree Fusion Algorithm.

Thus, the PitTree Fusion Algorithm combines information about the morphological appearance of specific regions (from pit patterns) with the structural organization of the mucosal surface (from MST) by merging MST analysis with Kudo's pit pattern analysis. By separating inter-class confounders from genuine dyed tissues, this integration enhances the accuracy of DLPs and DRMs classification in endoscopic images. In essence, the algorithm leverages both structural and morphological features to enhance the detection and depiction of dyed regions within the gastrointestinal tract. Finally, a novel Efficient-CondConv SwishNet is introduced for the classification of GI diseases based on the extracted features, which is explained in sub section III-C.

C. EFFICIENT-CondConv SwishNet

The introduction of Efficient-CondConv SwishNet marks a significant advancement in the classification of GI diseases, leveraging extracted features from preceding steps. This model employs EfficientNet-CondConv architecture

Algorithm 2 PitTree Fusion Algorithm

Input : Endoscopy image containing the GI tract.
Output : Labeled image corresponding to normal tissue, DLPs, and DRMs

1: Input Endoscopic Image.
2: Perform MST Analysis to identify ROIs.
3: Apply Kudo's Pit Pattern Analysis to each ROI.
4: Fuse MST and Pit Pattern Analysis results.
5: Output labeled image of ROIs as DLP or DRM.

combined with Swish activation function, facilitating the capture of more informative and discriminative features from endoscopic GI images.

EfficientNet-CondConv's adaptive nature plays a crucial role, enabling effective feature extraction irrespective of input resolution variations. This adaptability ensures that relevant features are extracted efficiently, allowing for adaptive feature reuse across different input resolutions. The convolutional kernel in a CondConv layer is calculated as a function of the input example, which is provided by equation (4),

$$Output(y) = \sigma((\alpha_1 \cdot W_1 + \ldots + \alpha_m \cdot W_m) * y) \quad (4)$$

where m is the number of experts, σ is an activation function, and each $\alpha_i = r_i(y)$ is an example-dependent scalar weight generated using a routing function with learnt parameters. When a convolutional layer is modified to utilize CondConv, every kernel W_i has the same dimensions as the original convolutional kernel. Additionally, the Swish activation function enhances non-linearities within feature representations. This enhancement makes it possible for the model to gain more complex knowledge and complicated associations in the data, which results in a deeper comprehension of the underlying patterns shown in the endoscopic images. The Swish function is computed as in equation (5),

$$f(y) = y \times sigmoid(\beta y) \quad (5)$$

where, y is the input to the function and β is a scaling parameter. Thus, by leveraging Efficient-CondConv SwishNet, the model gains the capability to discern more complex and fine-grained patterns within the endoscopic image data. This capability is particularly vital for accurately classifying GI tract diseases, where subtle visual cues may be indicative of specific pathologies.

The combination of EfficientNet-CondConv with Swish activation function enables the model to capture more informative and discriminative features from the endoscopic images. These features encode the characteristics relevant for disease classification, leading to improved diagnostic accuracy. Algorithm 3 below illustrates the algorithm for the suggested Efficient-CondConv SwishNet method,

Algorithm 3 Efficient-CondConv SwishNet

Input: Endoscopic GI images to be classified.
Output: Predicted class label for the input images.

1: Input endoscopic GI images.
2: Extract features using EfficientNet architecture with conditional convolutions.
3: Apply Swish activation function to enhance feature representations.
4: Pass features through a classification head to predict GI disease classifications.
5: Train the model on labeled data.
6: During inference, use the trained model to predict disease classifications for unseen images.

D. EXPLAINABILITY WITH VISUALIZATION USING GRAD-CAM

Following classification, a heatmap visualization technique is employed to highlight the locations of the input image that added most significantly to the model's classification decision. Grad-CAM, which calculates the gradient of the model's output prediction regarding the feature mappings of the final convolutional layer, is used to generate this heatmap. By analyzing these gradients, Grad-CAM identifies the regions of the input image that had the highest impact on the model's decision-making process.

The resulting heatmap provides important information about the features or patterns that influenced the classification outcome. By visualizing the regions of the image that the model deemed most relevant for its decision, clinicians and researchers gain a deeper understanding of the model's reasoning process. This visualization aids in interpreting and validating the model's classifications, ultimately enhancing trust and confidence in its diagnostic capabilities. Algorithm for the heatmap generation using Grad-CAM method is shown in Algorithm 4.

Overall, this comprehensive approach improves the identification accuracy of GI diseases by leveraging advanced techniques in feature extraction, spatial analysis, pit pattern analysis, and heatmap visualization. Section IV gives a thorough examination of the empirical findings and the suggested methodology.

IV. RESULTS AND DISCUSSION

These findings include the performance of the suggested Hierarchical Spatio Pyramid TranfoNet with PitTree Fusion and Efficient-CondConv SwishNet system, as well as a comparison section to verify the suggested system's suitability for classifying GI diseases.

A. SYSTEM CONFIGURATION

This section offers a thorough explanation of the implementation findings and the performance of the proposed system, which is simulated in Python. It also includes a comparison section to verify the effectiveness of the suggested system. The tests were conducted on a Windows 10, 64-bit machine equipped with 32 GB of RAM and a 1 TB Hard Drive. The coding was executed using the Python programming language, with all required packages incorporated.

Algorithm 4 Heatmap Generation Using Grad-CAM

Input : Input image for classification
Output : Heatmap visualization highlighting regions of the input image

1: Perform a forward pass through the network to obtain the predicted class score and the final convolutional feature maps.
2: Compute the gradients of the predicted class score in relation to the final convolutional layer's feature mappings.
3: Compute the importance weights for each feature map by global average pooling of the gradients obtained in the previous step.
4: Weight the feature maps by the importance weights computed in step 3 to create the class activation map.
5: Create a heatmap overlaying the input image with the class activation map to show the places that significantly influence the classification decision.

B. DATASET DESCRIPTION

The Kvasir dataset [26], which includes images from inside the gastrointestinal (GI) tract, is used for the empirical study of the proposed model. Two major image categories associated with endoscopic polyp removal are included and one miscellaneous (all other categories) are included.

(i) Z-line, Pylorus, Cecum, and other anatomic landmarks are examples of pathological findings;
(ii) Esophagitis, Polyps, and Ulcerative Colitis are examples of anatomic landmarks.
(iii) Furthermore, other sets of images related to lesion removal were offered, such as ''Dyed lifted polyps'', the ''Dyed resection margins'', etc.

The images contained in the dataset range in resolution from 720 × 576 to 1920 × 1072 pixels. It is arranged into several folders titled in accordance with its content. Total images used for training is 8000 and testing is 800 with a ratio of 80:20.

C. EXPERIMENTAL RESULTS

The suggested model's experimental outcomes for GI disease classification are discussed in this section from the initial setup.

Figure 3 displays input images extracted from the dataset. It includes eight distinct types of images utilized in the process. Among these, three represent normal images, namely (a) normal-cecum, (b) normal-pylorus, and (c) normal-z-line, and five images showcasing various infections: (d) dyed-lifted-polyps, (e) dyed-resection-margins, (f) esophagitis, (g) polyps, and (h) ulcerative-colitis. These input images undergo initial preprocessing before proceeding to subsequent stages of processing.

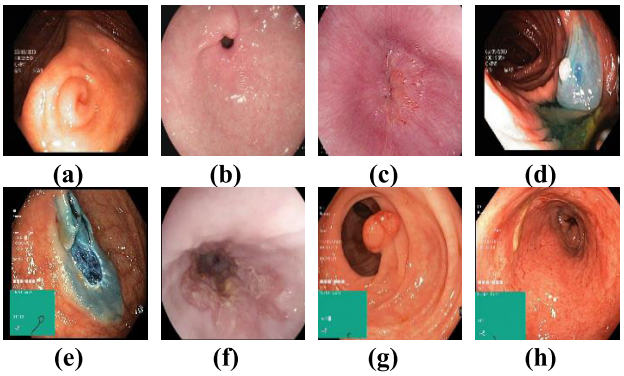

FIGURE 3. Input image from dataset.

Figure 4 shows the preprocessed image (right) of the input data for different categories such as (a) normal-cecum, (b) normal-pylorus, (c) normal-z-line, (d) dyed-lifted-polyps, (e) dyed-resection-margins, (f) esophagitis, (g) polyps, and (h) ulcerative-colitis. Preprocessing involve tasks such as data cleaning, normalization, feature selection, and data transformation to prepare the data for classification algorithms. By using this method, training time is shortened, model performance is enhanced, and data quality and consistency are guaranteed.

Figure 5 displays the predicted output results of the proposed system for different categories such as (a) normal-cecum, (b) normal-pylorus, (c) normal-z-line, (d) dyed-lifted-polyps, (e) dyed-resection-margins, (f) esophagitis, (g) polyps, and (h) ulcerative-colitis. This is a critical component in many disciplines where models are taught to forecast outcomes based on hidden data. This entails evaluating the model's performance by comparing predictions to actual results and utilizing error rate metrics.

Figure 6 displays the heatmap image using Grad-CAM for different categories such as (a) dyed-lifted-polyps, (b) dyed-resection-margins, (c) esophagitis, (d) polyps, and (e) ulcerative-colitis. Grad-CAM is used to create heatmaps that highpoint the areas of an image that are most vital for a certain class prediction. It operates by examining the gradients that enter the last convolutional layer of a network.

D. PERFORMANCE METRICS OF THE PROPOSED SYSTEM

The experimental part for the proposed method is assessed using several measures. These measurements include the accuracy, precision, recall, F1 score, sensitivity, specificity, Matthews's correlation coefficient (MCC), Area Under Curve (AUC), Positive Predictive Value (PPV), Loss and False Positive Rate (FPR). Performance measures are determined using the formulas found in equations (6)–(15),

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FN} + \text{TN} + \text{FP})} \tag{6}$$

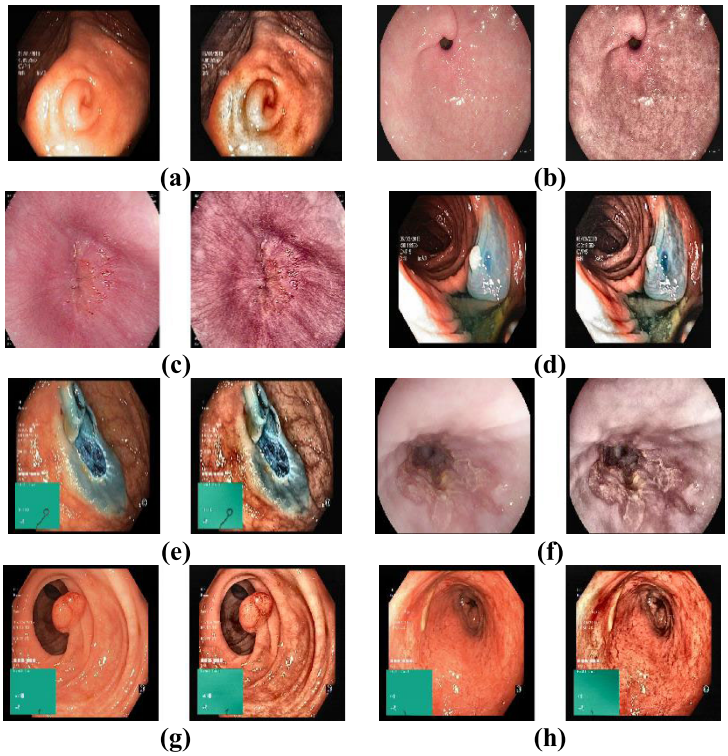

FIGURE 4. Preprocessed Image of the proposed model.

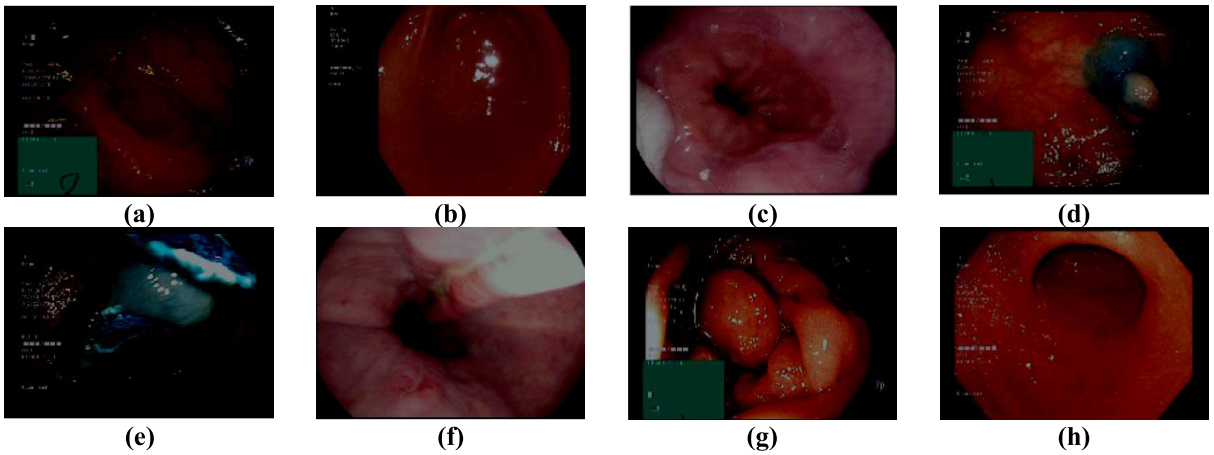

FIGURE 5. Predicted Output of the proposed method.

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

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

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \tag{10}$$

$$\text{MCC} = \frac{(\text{TP.TN}) - (\text{FP.FN})}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{11}$$

$$\text{AUC} = \frac{\text{True Positive Rate}}{\text{False Positive Rate}} \tag{12}$$

$$\text{PPV} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{13}$$

$$\text{Loss} = -\sum_{n=1}^{k} (L_i \log(p_i)) \tag{14}$$

$$\text{FPR} = \frac{\text{FP}}{(\text{FP} + \text{TN})} \tag{15}$$

where L is the calculated loss of each class and p is the probability determined by the Soft function. The number

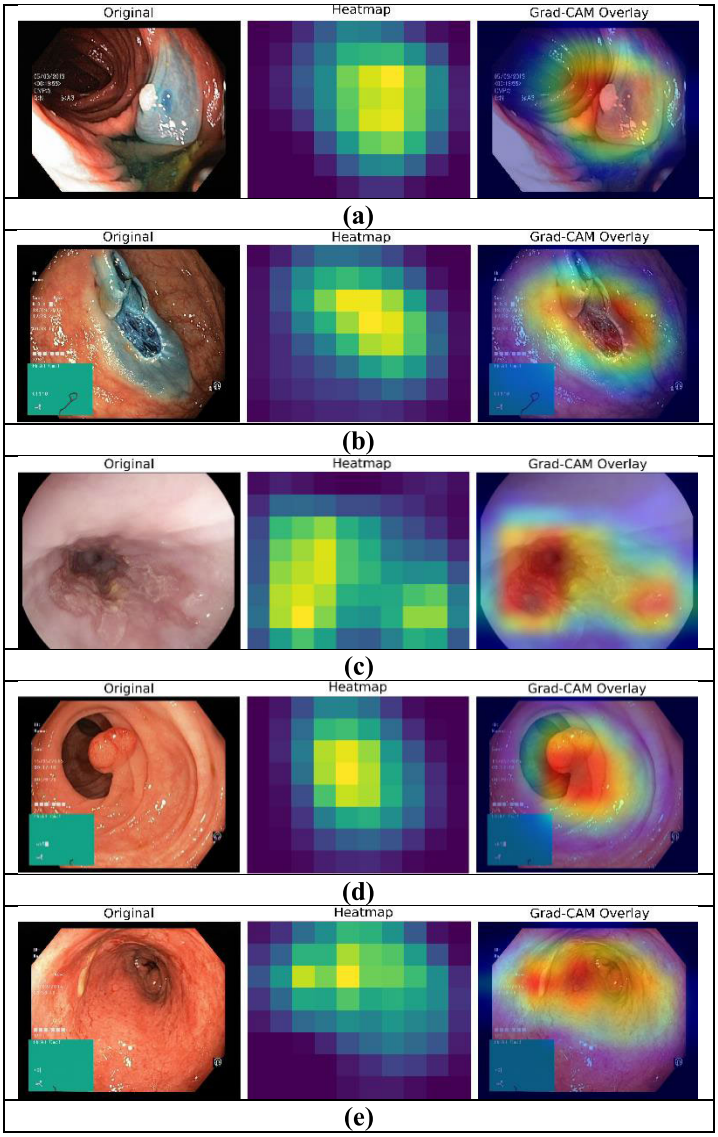

FIGURE 6. Heatmap of the proposed model.

of true positives, true negatives, false positives, and false negatives are allocated to the TP, TN, FP, and FN parameters, respectively.

The confusion matrix for the classification results is displayed in Figure 7. It shows the true labels and predicted labels for different categories such as dyed-lifted-polyps, dyed-resection-margins, esophagitis, normal-cecum, normal-pylorus, normal-z-line, polyps, and ulcerative-colitis. The matrix displays the number of correct predictions (true positives) for each category and shows some misclassifications. Thus, the model performs well with high true positive rates and few misclassifications.

The training and testing accuracy of the suggested model with different epochs is shown in Figure 8. The analysis shows that the accuracy increases as the epoch increases. The training set obtains a maximum accuracy of 98.1% and the testing set achieves a maximum accuracy of 98.2% when the epoch value is 100, and when the epoch is 20, the training set achieves a minimum accuracy of 82% and the testing set achieves a minimum accuracy of 77%. By combining information about structural organization from MST and morphological appearance from pit patterns, the algorithm effectively differentiates between genuine dyed tissues and inter-class confounders, thereby improving the accuracy of DLPs and DRMs classification.

The training and testing loss rate of the suggested model with different epochs is shown in Figure 9. The analysis shows that the loss decreases as the epoch increases. When the epoch value is 100, the training set achieves a minimum loss of 0.02% and the testing set achieves a minimum loss of 0.03%, and when the epoch is 20, the training set achieves a maximum loss of 0.4% and the testing set achieves a

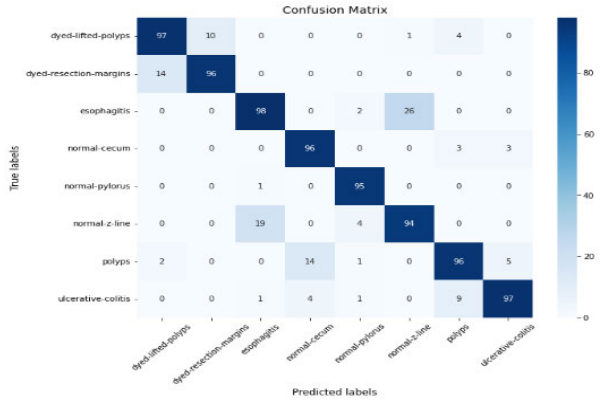

FIGURE 7. Confusion Matrix of the proposed model.

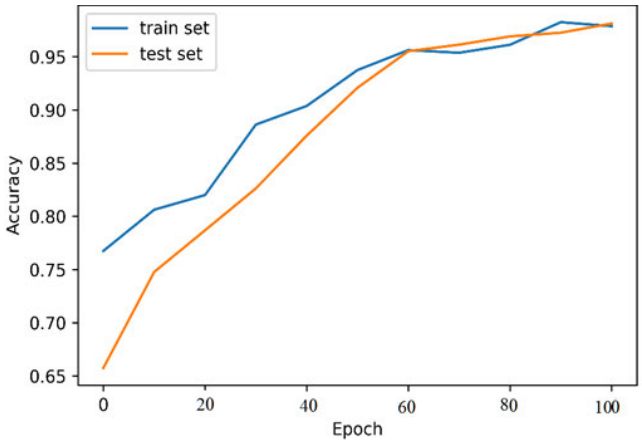

FIGURE 8. Accuracy of the proposed system.

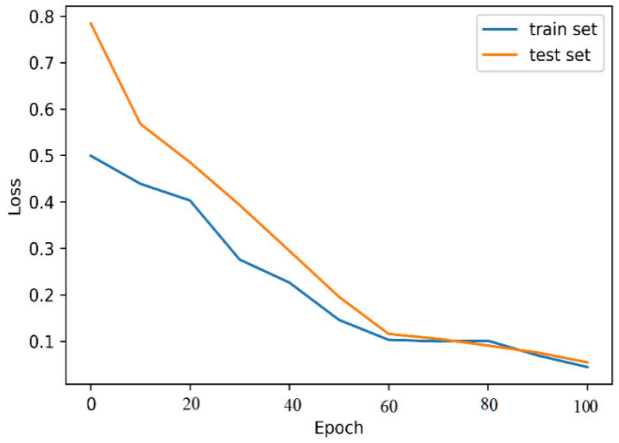

FIGURE 9. Loss of the proposed system.

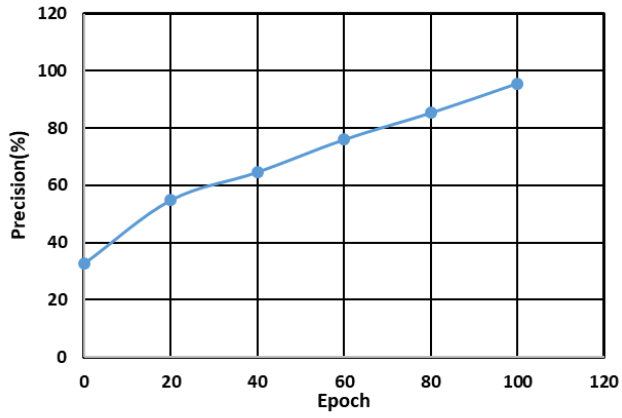

FIGURE 10. Precision of the proposed system.

maximum loss of 0.5%. The PitTree Fusion Algorithm helps in accurately localizing and distinguishing between normal tissue and dyed regions, which improves the model's capacity to distinguish between classes and lowers the loss rate.

Figure 10 illustrates the performances of the precision of the proposed model with varied epoch. When the epoch value is 20 it achieves the minimum precision value of 36% and while the epoch value is increased to 100, it achieves the maximum precision value of 95.6%. Hierarchical Spatio Pyramid TranfoNet aids the model to concentrate on relevant informative regions and extract features at various levels of granularity. This hierarchical representation of spatial context permits for a more inclusive understanding of the features present in endoscopic images, leading to improved precision.

The suggested model's recall performances are shown in Figure 11 with varied epoch. It obtains a minimum recall value of 64.5% when the epoch value is 20 and a maximum recall value of 97.4% when the epoch is 100. By effectively learning complex patterns crucial for disease classification, Efficient-CondConv SwishNet model reduces false negatives, thus improving recall performance.

The suggested system's sensitivity with different epochs is displayed in Figure 12. By increasing the epoch to 100, the suggested system's sensitivity reaches a maximum value of 94.9%, while at epoch 20, it reaches a low value of 52%. The Grad-CAM technique, employed post-classification, generates heatmaps to visualize areas contributing most to the model's decisions, resulting in more accurate heatmaps and improved sensitivity.

The specificity performances of the suggested model are shown in Figure 13 with varied epoch. It obtains the maximum specificity of 99.5% when the epoch is 100, and obtains the minimum specificity of 79%, when the epoch is 20. By effectively differentiating between genuine dyed tissues and inter-class confounders, the PitTree Fusion Algorithm reduces false positives and enhances specificity.

Figure 14 displays the proposed model's F1 score performances throughout a range of epochs. When the epoch value is 100, it achieves the maximum F1 score of 94.4% and the minimum F1 score of 42%, when the epoch is 20. By learning complex and fine patterns in endoscopic image data, the Efficient-CondConv SwishNet model improves its ability to accurately classify GI tract diseases, leading to an increase in F1 score.

The suggested model's AUC graph is shown in Figure 15 with varied epoch. The AUC increases as the epoch increases,

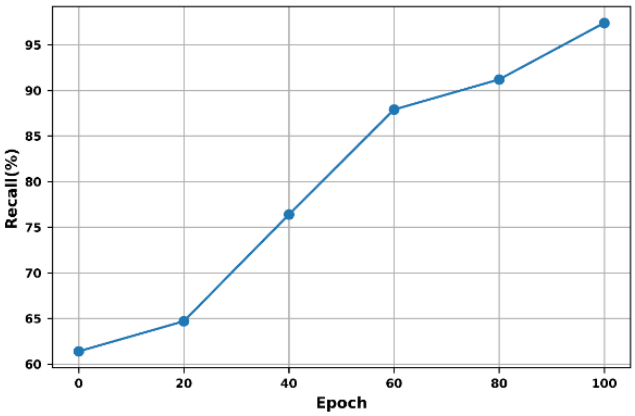

FIGURE 11. Recall of the proposed model.

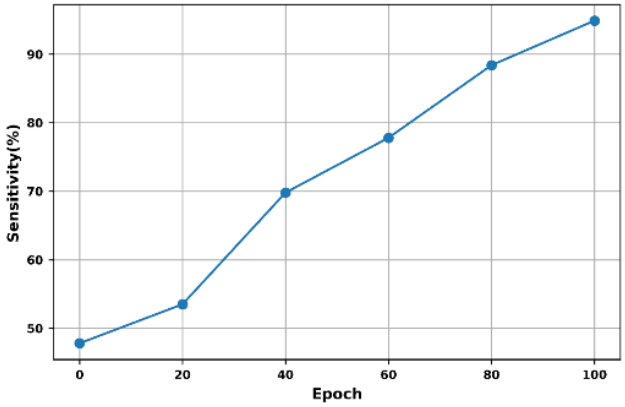

FIGURE 12. Sensitivity of the proposed system.

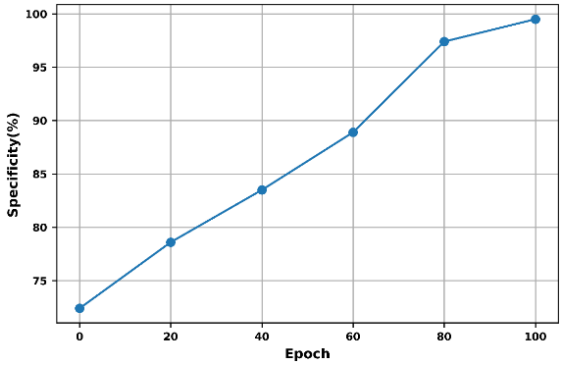

FIGURE 13. Specificity of the proposed model.

which results in better performance of the system. It reaches a maximum AUC of 99.99% when the epoch is 100 and a lowest AUC of 76% when the epoch value is 20. By capturing features at various levels of granularity, Hierarchical Spatio Pyramid TranfoNet model enhances its ability to discriminate between different classes of GI diseases, leading to an increase in AUC over time as the network learns more discriminative features.

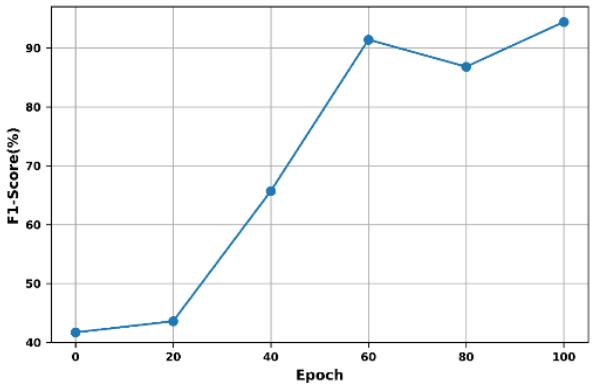

FIGURE 14. F1 Score of the proposed model.

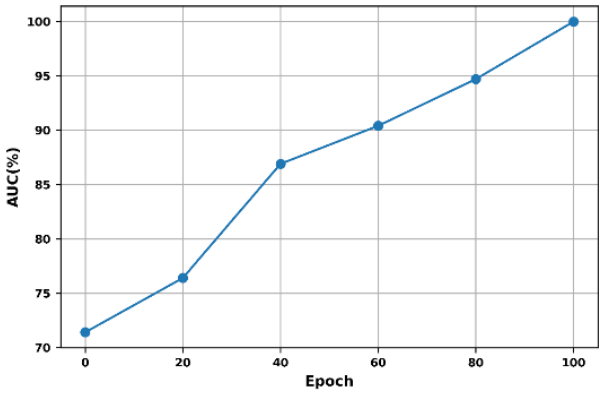

FIGURE 15. Area Under the Curve (AUC) of the proposed model.

The MCC of the suggested system with varying epochs is displayed in Figure 16. The MCC of the suggested system reaches a minimum of 65% when the period is 20 and a maximum of 92.7% when the epoch is increased to 100. By integrating MST analysis with Kudo's pit pattern analysis, the PitTree Fusion Algorithm enhances the model's ability to accurately localize and distinguish between normal tissues and dyed regions; this reduces misclassifications and ultimately contributing to higher MCC values.

The PPV performances of the suggested model are shown in Figure 17 with varied epoch. It obtains the maximum PPV of 99.7% when the epoch is 100, and obtains the minimum PPV of 58%, when the epoch is 20. The Efficient-CondConv SwishNet architecture enhances feature extraction from endoscopic images by leveraging the EfficientNet-CondConv architecture with Swish activation, which enables the model to capture more informative and discriminative features, leading to higher PPV.

The FPR rate of the suggested model with different epochs is shown in Figure 18. The analysis shows that the FPR decreases as the epoch increases. When the epoch value is 100, it achieves a minimum FPR rate of 0.002% and a maximum FPR rate of 0.010%, when the epoch is 20. Efficient-CondConv SwishNet utilizes EfficientNet-CondConv architecture with Swish activation to capture informative and discriminative features from

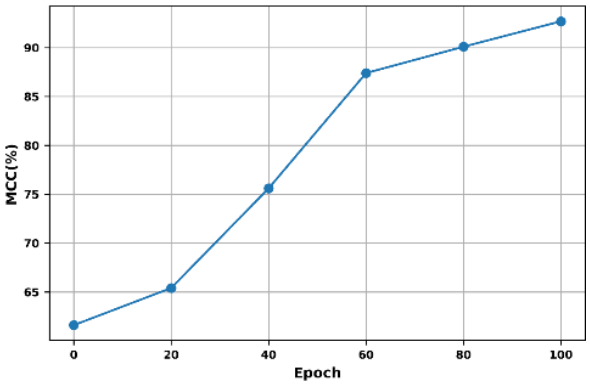

FIGURE 16. Matthews Correlation Coefficient (MCC) of the proposed system.

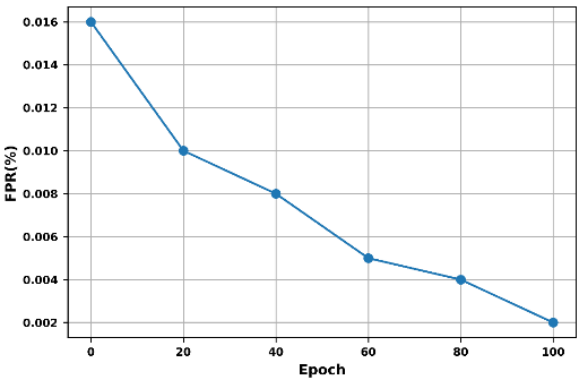

FIGURE 18. False Positive Rate (FPR) of the proposed model.

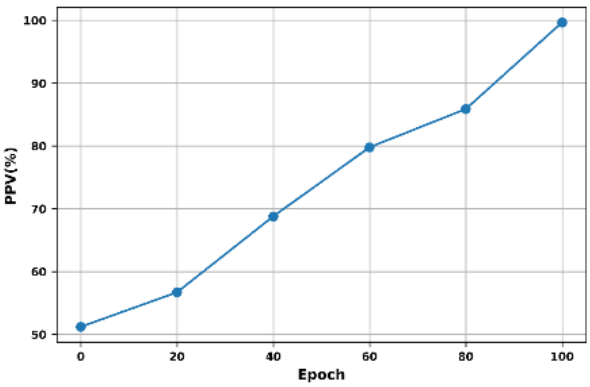

FIGURE 17. Positive Predictive Value (PPV) of the proposed model.

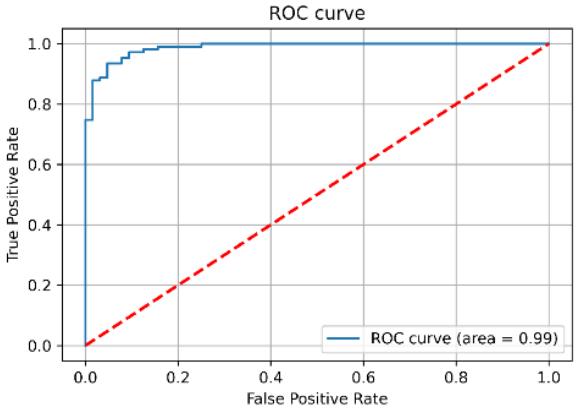

FIGURE 19. Receiver Operating Characteristic (ROC) of the proposed model.

endoscopic images, thus enabling the model to learn complex patterns crucial for accurate disease classification and reducing false positives.

A ROC curve is depicted in Figure 19, with the True Positive Rate on the y-axis and the False Positive Rate on the x-axis. A high true positive rate for low false positive rates is indicated by the graph's high performance curve. The legend mentions that the area under the ROC curve is 0.99 approximately, which signifies excellent performance of the classifier being evaluated.

E. COMPARISON OF PROPOSED MODEL WITH PREVIOUS MODELS

This section presents a metric-based comparison of the effectiveness of the proposed model with the outcomes of the current approaches. AUC, MCC, PPV, loss, FPR, sensitivity, specificity, recall, accuracy, precision, and F1 score are all compared. There are comparisons with the current methods, including AlexNet, GoogleNet, and ResNet 50.

The suggested model's accuracy is compared to that of current models. The existing models such as AlexNet, GoogleNet, and ResNet 50 achieves an accuracy value of 91.35%, 91.70%, and 93.01% respectively. Compared with existing models the proposed model achieves a high accuracy of 98.2%. The precision values of the current models, including AlexNet, GoogleNet, and ResNet 50, are 91.67%, 91.71%, and 93.0%, respectively. In comparison to current models, the suggested model attains a maximum precision rate of 95.6%.The recall of the suggested model is compared to that of the current models, recall values of 83.9%, 78.91%, and 90.33% are attained by the current models, including AlexNet, ResNet 50, and GoogleNet. The suggested model obtains a recall of 97.4% when compared to current models. The existing models such as AlexNet, GoogleNet, and ResNet 50 achieves a sensitivity value of 91.35%, 91.70%, and 93.0% respectively. Compared with existing models the proposed model achieves a high sensitivity of 94.9%.The specificity values of the current models, including AlexNet, GoogleNet, and ResNet 50, are 98.76%, 98.81%, and 99.0%, respectively. In comparison to current models, the suggested model attains a high specificity rate of 99.5%. The F1 score values of the current models such as, AlexNet, GoogleNet, and ResNet 50, are 91.80%, 91.68%, and 93.0%, respectively. The suggested model receives a maximum F1 score of 94.4% when compared to the current models. The existing models such as AlexNet, GoogleNet, and ResNet 50 achieves a AUC value of 99.98%, 99.984%, and 99.69%, respectively. Compared with existing models the proposed model achieves a high AUC of 99.99%.The MCC values of the current models,

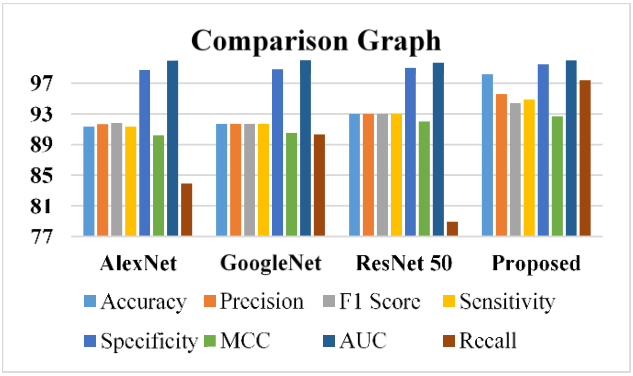

FIGURE 20. Overall comparison of the proposed model with the existing methods.

TABLE 1. Comparison analysis of the proposed model.

Parameters	AlexNet	ResNet 50	GoogleNet	Proposed
Accuracy %	91.35	93.01	91.70	**98.2**
Precision %	91.67	93.0	91.71	**95.6**
Recall %	83.9	78.91	90.33	**97.4**
F1-score %	91.80	93.0	91.68	**94.4**
Sensitivity %	91.35	93.0	91.70	**94.9**
Specificity %	98.76	99.0	98.81	**99.5**
MCC %	90.20	92.0	90.51	**92.7**
AUC %	99.98	99.69	99.984	**99.99**
Loss %	0.4	0.5	0.6	**0.03**

including AlexNet, GoogleNet, and ResNet 50, are 90.20%, 90.51%, and 92.0%, respectively. In comparison to current models, the suggested model attains a high MCC rate of 92.7%. The loss rate of the proposed model with that of the existing models. The existing models such as, ResNet-50, AlexNet, and GoogleNet, obtains a loss rate of 0.5%, 0.4%, and 0.6%, respectively. The proposed model achieves 0.03% loss rate compared with the existing models.

The overall contrast between the suggested model and the current techniques is shown in Figure 20. The suggested model outperforms the current models, including AlexNet, GoogleNet, and ResNet 50, in terms of several metrics, including accuracy, precision, recall, F1 Score, Loss, MCC, AUC, sensitivity, and specificity.

The proposed model's comparative analysis with the current models is shown in Table 1. The existing models such as AlexNet, ResNet 50, and GoogleNet have accuracy of 91.35%, 93.01%, and 91.70%; precision of 91.67%, 93.0%, and 91.71%; recall of 89.9 %, 78.91%, and 90.33%; F1 score of 91.80%, 93.0%, and 91.68%; sensitivity of 91.35%, 93.0%, and 91.70%; specificity of 98.76%, 99.0%, and 98.81%; MCC of 90.20%, 92.0%, and 90.51%; AUC of 99.98%, 99.69%, and 99.984%, and loss of 0.4%, 0.5%, and 0.6% respectively. Compared to the existing methods, the proposed model attains high accuracy of 98.2%, precision of 95.6%, recall of 97.4%, F1 score of 94.4%, sensitivity of 94.9%, specificity of 99.5%, MCC of 92.7%, AUC of 99.99%, and low loss rate of 0.03%.

TABLE 2. Comparison of training time.

Models	Training Time (sec)
AlexNet	33600
GoogleNet	36501
ResNet 50	41605
Proposed Model	**30000**

TABLE 3. Ablation study.

Model	Metrics			
	Precision	Loss	PPV	FPR
Hierarchical Spatio Pyramid TranfoNet	81	0.23	85	0.027
PitTreeFusion Algorithm	89	0.15	91	0.016
Efficient-CondConv SwishNet	92.4	0.08	95.5	0.007
Hierarchical Spatio Pyramid TranfoNet + PitTreeFusion Algorithm	88	0.20	88.7	0.012
PitTreeFusion Algorithm+Efficient-CondConv SwishNet	91.5	0.11	93	0.009
Hierarchical Spatio Pyramid TranfoNet + Efficient-CondConv SwishNet	87.3	0.17	90.1	0.016
Entire proposed model	**95.6**	**0.03**	**99.7**	**0.002**

Table 2 displays the training time comparison between the suggested model and the existing models. The training time of AlexNet, GoogleNet and ResNet is 33600s, 36501s, and 41605s. In contrast to the current approaches, the suggested model involves less training time of 30000s.

Overall, in the results section, the proposed model is compared to existing models and the performance is explained using graphs. This demonstrates that the novel Hierarchical Spatio Pyramid TranfoNet with PitTree Fusion and Efficient-CondConv SwishNet model has high accuracy of 98.2%, precision of 95.6%, recall of 97.4%, sensitivity of 94.9%, specificity of 99.5%, F1 score of 94.4%, AUC of 99.99%, MCC of 92.7%, PPV of 99.7%, and low loss rate of 0.03%, FPR of 0.002%, and training time of 30000s when compared to the previous models.

F. ABLATION STUDY OF THE PROPOSED MODEL

In an ablation study, each component of the proposed model is progressively removed or modified in order to evaluate how each one affects the model's performance on its own. This aids in appreciating the significance of every element and directs additional model enhancements or simplifications.

TABLE 4. Performance comparison with other researcher's work.

Method	Accuracy (%)
M.A.Khan et al,2020	98%
Saban Ozturk et al,2021	98.05%
Melaku Bitew Haile et al, 2022	98.03%
Ramamurthy et al,2022	97.99%
Salman Hosain et al, 2022	95.63%
Marwa Obayya et al,2023	94.25
Proposed	**98.2%**

Table 3 represents the ablation study of the proposed methods. The performance parameters like precision, loss, PPV, and FPR are taken and compared with novel methods. By comparing the Hierarchical Spatio Pyramid TranfoNet method first, then PitTree Fusion Algorithm second, third Efficient-CondConv SwishNet, then the combinations of the methods, and finally the entire proposed method values.

The classification studies that were conducted utilising the Kvasir dataset in the literature are shown in Table 4. When comparing the obtained performance performances, the suggested strategy performs better than alternative approaches. Marwa Obayya et al. had the lowest classification performance (94.25% accuracy rate) for the Kvasir dataset in the literature. With a 98.05% accuracy rate, Saban Ozturk et al. fare best in categorization. Table 4 illustrates that, with an accuracy rate of 98.2%, the suggested approach performed best. In the next section V conclude the paper with our proposed values.

V. CONCLUSION AND FUTURE WORK

The proposed Advanced GI Disease Classification model offered a sophisticated approach to improve the identification accuracy of gastrointestinal diseases particularly esophagitis, ulcerative colitis, dyed lifted polyps, and dyed resection margins in endoscopy images. Through the integration of a novel Hierarchical Spatio Pyramid TranfoNet, PitTree Fusion Algorithm, and Efficient-CondConv SwishNet, the model demonstrated enhanced discriminative power and accuracy in classifying overlapping visual characteristics and distinguishing between normal tissue and dyed regions. By leveraging spatial context, structural organization, and morphological appearance, the model effectively addressed the challenges associated with disease classification and localization within complex anatomical sites. The utilization of Grad-CAM further enhanced interpretability by visualizing the regions of the input image crucial for classification decisions, providing valuable insights into the features influencing the model's outcomes. Thus, the results of the simulation demonstrated that the proposed model has high accuracy of 98.2%, precision of 95.6%, recall of 97.4%, sensitivity of 94.9%, specificity of 99.5%, F1 score of 94.4%, AUC of 99.99%, MCC of 92.7%, PPV of 99.7%, and low loss rate of 0.03%, and FPR of 0.002% when compared to the previous models. Overall, this comprehensive approach holds promise for advancing the diagnosis and treatment of gastrointestinal diseases through improved endoscopic imaging analysis.

A promising future work direction would be to investigate methods for fusing multi-modal data sources, beyond just endoscopic images, to enhance the accuracy and reliability of GI disease classification. In a clinical setting, healthcare professionals often have access to various types of data, such as patient medical history, laboratory test results, and potentially even genomic or molecular biomarker information. By combining these diverse data modalities with the visual and spatial features extracted from endoscopic images, the model may be able to capture a more inclusive and holistic illustration of the underlying disease features.

ACKNOWLEDGMENT

The authors would like to thank Vellore Institute of Technology, Chennai, India, for their support.

REFERENCES

[1] R. K. Dey, M. E. Rana, and V. A. Hameed, "Analysing wireless capsule endoscopy images using deep learning frameworks to classify different GI tract diseases," in *Proc. 17th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM)*, Jan. 2023, pp. 1–7.

[2] S. Aliyi, K. Dese, and H. Raj, "Detection of gastrointestinal tract disorders using deep learning methods from colonoscopy images and videos," *Scientific Afr.*, vol. 20, Jul. 2023, Art. no. e01628.

[3] E. Klang, Y. Barash, R. Y. Margalit, S. Soffer, O. Shimon, A. Albshesh, S. Ben-Horin, M. M. Amitai, R. Eliakim, and U. Kopylov, "Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy," *Gastrointestinal Endoscopy*, vol. 91, no. 3, pp. 606–613.e2, Mar. 2020.

[4] S. Soffer, E. Klang, O. Shimon, N. Nachmias, R. Eliakim, S. Ben-Horin, U. Kopylov, and Y. Barash, "Deep learning for wireless capsule endoscopy: A systematic review and meta-analysis," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 831–839.e8, Oct. 2020.

[5] Z. Ding, H. Shi, H. Zhang, L. Meng, M. Fan, C. Han, K. Zhang, F. Ming, X. Xie, H. Liu, J. Liu, R. Lin, and X. Hou, "Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model," *Gastroenterology*, vol. 157, no. 4, pp. 1044–1054, Oct. 2019.

[6] E. Mousavi, A. H. Keshteli, M. Sehhati, A. Vaez, and P. Adibi, "Re-investigation of functional gastrointestinal disorders utilizing a machine learning approach," *BMC Med. Informat. Decis. Making*, vol. 23, no. 1, p. 167, Aug. 2023.

[7] A. Krenzer, S. Heil, D. Fitting, S. Matti, W. G. Zoller, A. Hann, and F. Puppe, "Automated classification of polyps using deep learning architectures and few-shot learning," *BMC Med. Imag.*, vol. 23, no. 1, p. 59, Apr. 2023.

[8] H. Zhuang, J. Zhang, and F. Liao, "A systematic review on application of deep learning in digestive system image processing," *Vis. Comput.*, vol. 39, no. 6, pp. 2207–2222, Jun. 2023.

[9] H. Bolhasani, S. J. Jassbi, and A. Sharifi, "DLA-E: A deep learning accelerator for endoscopic images classification," *J. Big Data*, vol. 10, no. 1, p. 76, May 2023.

[10] O. M. Mirza, A. Alsobhi, T. Hasanin, M. K. Ishak, F. K. Karim, and S. M. Mostafa, "Computer aided diagnosis for gastrointestinal cancer classification using hybrid rice optimization with deep learning," *IEEE Access*, vol. 11, pp. 76321–76329, 2023.

[11] M. S. Hossain, M. M. Rahman, M. M. Syeed, M. F. Uddin, M. Hasan, M. A. Hossain, A. Ksibi, M. M. Jamjoom, Z. Ullah, and M. A. Samad, "DeepPoly: Deep learning-based polyps segmentation and classification for autonomous colonoscopy examination," *IEEE Access*, vol. 11, pp. 95889–95902, 2023.

[12] M. Vania, B. A. Tama, H. Maulahela, and S. Lim, "Recent advances in applying machine learning and deep learning to detect upper gastrointestinal tract lesions," *IEEE Access*, vol. 11, pp. 66544–66567, 2023.
[13] M. W. Scheppach et al., "Detection of duodenal villous atrophy on endoscopic images using a deep learning algorithm," *Gastrointestinal Endoscopy*, vol. 97, no. 5, pp. 911–916, May 2023.
[14] M. Amirthalingam and R. Ponnusamy, "Improved water strider optimization with deep learning based image classification for wireless capsule endoscopy," in *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy (ICAIS)*, Feb. 2023, pp. 851–857.
[15] D. Jha, V. Sharma, N. Dasu, N. K. Tomar, S. Hicks, M. K. Bhuyan, P. K. Das, M. A. Riegler, P. Halvorsen, U. Bagci, and T. de Lange, "GastroVision: A multi-class endoscopy image dataset for computer aided gastrointestinal disease detection," in *Proc. Workshop Mach. Learn. Multimodal Healthcare Data*. Cham, Switzerland: Springer, Jul. 2023, pp. 125–140.
[16] A. K. M. S. Hosain, M. Islam, M. H. K. Mehedi, I. E. Kabir, and Z. T. Khan, "Gastrointestinal disorder detection with a transformer based approach," in *Proc. IEEE 13th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2022, pp. 0280–0285.
[17] M. A. Khan, S. Kadry, M. Alhaisoni, Y. Nam, Y. Zhang, V. Rajinikanth, and M. S. Sarfraz, "Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: A framework of best features selection," *IEEE Access*, vol. 8, pp. 132850–132859, 2020.
[18] M. B. Haile, A. O. Salau, B. Enyew, and A. J. Belay, "Detection and classification of gastrointestinal disease using convolutional neural network and SVM," *Cogent Eng.*, vol. 9, no. 1, Dec. 2022, Art. no. 2084878.
[19] M. Obayya, F. N. Al-Wesabi, M. Maashi, A. Mohamed, M. A. Hamza, S. Drar, I. Yaseen, and M. I. Alsaid, "Modified salp swarm algorithm with deep learning based gastrointestinal tract disease classification on endoscopic images," *IEEE Access*, vol. 11, pp. 25959–25967, 2023.
[20] S. Mohapatra, G. K. Pati, M. Mishra, and T. Swarnkar, "Gastrointestinal abnormality detection and classification using empirical wavelet transform and deep convolutional neural network from endoscopic images," *Ain Shams Eng. J.*, vol. 14, no. 4, Apr. 2023, Art. no. 101942.
[21] K. Ramamurthy, T. T. George, Y. Shah, and P. Sasidhar, "A novel multi-feature fusion method for classification of gastrointestinal diseases using endoscopy images," *Diagnostics*, vol. 12, no. 10, p. 2316, Sep. 2022.
[22] K. J. Nass, L. W. Zwager, M. van der Vlugt, E. Dekker, P. M. M. Bossuyt, S. Ravindran, S. Thomas-Gibson, and P. Fockens, "Novel classification for adverse events in GI endoscopy: The AGREE classification," *Gastrointestinal Endoscopy*, vol. 95, no. 6, pp. 1078–1085.e8, Jun. 2022.
[23] S. M. Fati, E. M. Senan, and A. T. Azar, "Hybrid and deep learning approach for early diagnosis of lower gastrointestinal diseases," *Sensors*, vol. 22, no. 11, p. 4079, May 2022.
[24] I. Iqbal, K. Walayat, M. U. Kakar, and J. Ma, "Automated identification of human gastrointestinal tract abnormalities based on deep convolutional neural network with endoscopic images," *Intell. Syst. Appl.*, vol. 16, Nov. 2022, Art. no. 200149.
[25] N. Sharma, S. Gupta, D. Koundal, S. Alyami, H. Alshahrani, Y. Asiri, and A. Shaikh, "U-net model with transfer learning model as a backbone for segmentation of gastrointestinal tract," *Bioengineering*, vol. 10, no. 1, p. 119, Jan. 2023.
[26] *Kvasir V2*. Accessed: 2017. [Online]. Available: https://www.kaggle.com/datasets/plhalvorsen/kvasir-v2-a-gastrointestinal-tract-dataset/data
[27] Ö. Öztürk and U. Özkaya, "Residual LSTM layered CNN for classification of gastrointestinal tract diseases," *J. Biomed. Informat.*, vol. 113, Jan. 2021, Art. no. 103638.
[28] Ö. Öztürk and U. Özkaya, "Gastrointestinal tract classification using improved LSTM based CNN," *Multimedia Tools Appl.*, vol. 79, nos. 39–40, pp. 28825–28840, Oct. 2020.
[29] M. Hmoud Al-Adhaileh, E. M. Senan, F. W. Alsaade, T. H. H. Aldhyani, N. Alsharif, A. A. Alqarni, M. I. Uddin, M. Y. Alzahrani, E. D. Alzain, and M. E. Jadhav, "Deep learning algorithms for detection and classification of gastrointestinal diseases," *Complexity*, vol. 2021, pp. 1–12, Oct. 2021.
[30] M. Sharif, M. A. Khan, M. Rashid, M. Yasmin, F. Afza, and U. J. Tanik, "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *J. Experim. Theor. Artif. Intell.*, vol. 33, no. 4, pp. 577–599, Jul. 2021.

V. SHARMILA received the B.Tech. degree in information technology from the Adhiyamaan College of Engineering (Affiliated to Anna University), Chennai, India, in 2009, and the M.E. degree in computer science and engineering from the Velammal Engineering College (Affiliated to Anna University), Chennai, in 2012. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai. She has published papers in reputed international conferences. Her research interests include medical image processing, computer vision, deep learning, and machine learning.

S. GEETHA (Senior Member, IEEE) received the B.E. degree in computer science and engineering from Madurai Kamaraj University, India, in 2000, and the M.E. degree in computer science and engineering and the Ph.D. degree from Anna University, Chennai, India, in 2004 and 2011, respectively. She has more than 18 years of rich teaching and research experience. She is currently a Professor and the Associate Dean of the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, India. She has published more than 80 papers in reputed international conferences and refereed journals. She has published four books. Her research interests include steganography, steganalysis, multimedia security, intrusion detection systems, machine learning paradigms, and information forensics. She was a recipient of the University Rank and Academic Topper Award from the B.E. and M.E. degrees, in 2000 and 2004, respectively. She was also a proud recipient of the ASDF Best Academic Researcher Award 2013, the ASDF Best Professor Award 2014, the Research Award in 2016, and the High Performer. She joins the Review Committee and the Editorial Advisory Board of journals, such as IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE TRANSACTIONS ON IMAGE PROCESSING, *Multimedia Tools and Applications* (Springer), and *Information Sciences* (Elsevier). She has given many expert lectures, keynote addresses at international and national conferences. She has organized many workshops, conferences, and FDPs.