Transfer Learning Based Gastrointestinal Disease Classification Using WCE Images

Gandhimathi alias Usha S
Department of Electronics and Communication Engineering Velammal College of Engineering and Technology
Madurai, India
ushadears@gmail.com

Uma Maheswari P
Department of Electronics and Communication Engineering Velammal College of Engineering and Technology
Madurai, India
umamahes.p@gmail.com

Rajpriya V R
Department of Electronics and Communication Engineering Velammal College of Engineering and Technology
Madurai, India
rajpriyavondillin1105@gmail.com

Subashree T U
Department of Electronics and Communication Engineering Velammal College of Engineering and Technology
Madurai, India
subashree0906@gmail.com

Dharshana S
Department of Electronics and Communication Engineering Velammal College of Engineering and Technology
Madurai, India
dharshana2104@gmail.com

***Abstract*— Gastrointestinal (GI) disorders are common and, if not properly identified early on, can result in serious problems. Because capsule endoscopy can take detailed images of the digestive tract, it has become an essential technique for GI disease identification. However, manually analysing these images is time consuming and is error prone. Traditional analysis of gastrointestinal images is difficult due to their complexity, as well as differences in lighting, texture, and structure. In medical image classification, deep learning models—especially those based on transfer learning—have demonstrated considerable promise. This paper suggests a transfer learning-based method for classifying gastrointestinal disorders using the EfficientNetB2 model. The suggested model effectively extracts complicated characteristics by utilizing pre-trained convolutional layers. With a, precision of 0.9726, Area Under the Curve of 0.9991 and recall of 0.9715, and overall accuracy of 0.9715, experimental findings show that the model performs remarkably well. By allowing the fast and precise classification of gastrointestinal disorders, these metrics demonstrate the way the proposed method works to support better diagnostic decision-making.**

***Keywords*— Medical Imaging, EfficientNetB2, Capsule Endoscopy, Transfer Learning, Gastrointestinal Disease, Deep Learning, Automated Diagnosis, Convolutional Neural Networks.**

I. INTRODUCTION

Using images from endoscopes with wireless capsules, deep learning and machine learning have demonstrated great potential in precisely identifying and categorizing gastrointestinal (GI) disorders. Because gastrointestinal problems are complicated and diverse, categorizing gastro enteritis from endoscopic images is an important issue in medical diagnostics. Reliable classification is severely hampered by variations in lesions' size, shape, texture, and colour. The usefulness of transfer learning in medical imaging applications has been demonstrated by the proposal of an optimized Efficient Net model for the detection of GI disorders using WCE images [1]. In addition, a CNN-based system has been developed to identify gastrointestinal tract anomalies, with encouraging automated diagnosis outcomes [2]. The effectiveness of deep learning models for classifying gastrointestinal diseases has been evaluated in review articles. Several deep learning methods and their diagnostic capabilities in WCE image analysis are highlighted in a thorough evaluation [3]. By combining many feature extraction methods, a feature fusion-based framework has also been investigated, which increases classification accuracy [4]. The accuracy of GI disease classification has been improved through the widespread application of transfer learning techniques. On WCE datasets, a deep learning model that makes use of EfficientNet-B7 has demonstrated excellent classification performance [5]. A different study showed the possibility of self-supervised learning in medical imaging by examining the adaption of the DINOv2 foundation model for capsule endoscopy diagnosis [6]. CNN architectures for the classification of gastrointestinal diseases have been the subject of recent research. In the detection of GI disorders, a number of CNN-based methods have shown encouraging results [7]. The efficacy of transfer learning has also been demonstrated by the efficient classification of GI diseases using a pretrained deep convolutional neural network [8]. Additionally, specific deep learning models have been created for the classification of certain GI diseases. The significance of polyp recognition in the early diagnosis of GI diseases is highlighted by a CNN-based model created for colorectal polyp detection in colonoscopy images [9]. High accuracy in multi-class classification tasks was attained by another study that used pretrained deep learning models for WCE-based GI illness classification [10]. This field has benefited from the introduction of several deep learning techniques in earlier publications. Deep learning methods have been used to classify GI tract anomalies, laying the groundwork for more sophisticated models [11]. In order to increase detection performance, a hybrid deep learning strategy that integrates several neural network designs has been proposed for the early identification of lower gastrointestinal illnesses [12]. The potential of attention mechanisms in medical imaging has also been demonstrated by the exploration of a transformer-based method for the diagnosis of GI disorders [13]. To increase the precision of GI disease classification, transfer learning techniques have also been applied [14]. In this study, the model EfficientNetB2 aims to improve classification accuracy.

979-8-3315-2422-7/25/$31.00 ©2025 IEEE

II. METHODOLOGY

Esophagitis, ulcerative colitis, polyps, and normal tissue are among the gastrointestinal illnesses that might be difficult to classify from capsule endoscopy images because of differences in anatomy and imaging quality. EfficientNetB2, which has been pre-trained on ImageNet and refined on images of gastrointestinal diseases, is used in the suggested approach to overcome these difficulties. By combining depth-wise separable convolutions with convolutional convolutions, the model efficiently recovers features while maximizing resolution, depth, and accuracy. The model employs a SoftMax function to classify images after training on a carefully chosen dataset, increasing diagnostic efficiency and helping doctors identify and treat diseases accurately.

A. Data Description

The WCE Curated Colon Disease Dataset comprises wireless capsule endoscopy images categorized into four classes: esophagitis, ulcerative colitis, polyps, and normal. The dataset is structured into three main directories—train, validation (val), and test—each containing four subdirectories corresponding to these disease classes. The train directory includes 3,200 images (800 per class), the validation directory has 2,000 images (500 per class), and the test directory consists of 800 images (200 per class). The original images have a resolution of 571 × 531 pixels, but they are typically resized to 224 × 224 pixels for further processing.

Fig. 1. Sample Images

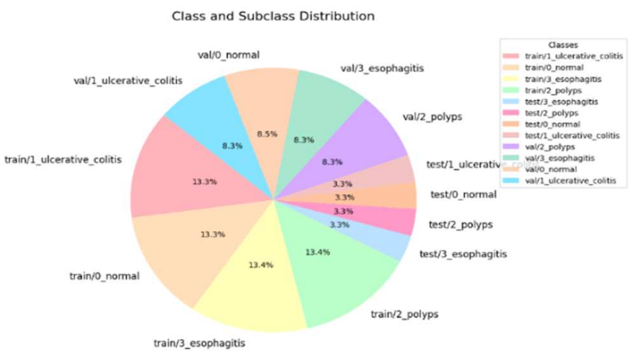

Fig. 2. Image Distribution between Classes

The four classes consisting Esophagitis, ulcerative colitis, polyps and normal are represented by a set of samples from the dataset in Fig. 1. The dataset's class and subclass distribution across the test, validation, and train sets is depicted in the Fig.2.

B. EfficientNetB2

In comparison to CNN, ResNet-50, and VGG19, EfficientNetB2 is suggested as a more advanced algorithm that performs better in classifying gastrointestinal illnesses in endoscopy with wireless capsules (WCE) images. EfficientNetB2 efficiently collects contextual information and fine details by using a compound scaling technique. The model divides WCE images into four groups: normal, polyps, ulcerative colitis, and esophagitis. By preprocessing and resizing images for consistency, it may extract both more general structures like the shape of the gastrointestinal wall and more specific patterns like the textures of lesions. By improving classification accuracy, this feature makes it possible to accurately identify and diagnose gastrointestinal disorders and facilitate prompt medical intervention.

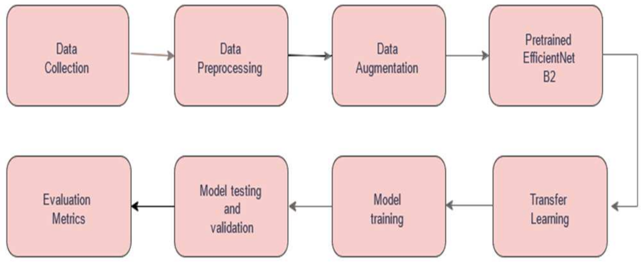

Fig. 3. Proposed Workflow

Fig. 3 shows the suggested approach for classifying gastrointestinal diseases. Data preprocessing, which improves image quality and eliminates noise, comes after data gathering. By increasing dataset heterogeneity through the use of data augmentation techniques, model robustness is enhanced. After that, a pretrained EfficientNetB2 model is fed the pre-processed images using transfer learning in order to benefit from its feature extraction powers. The system's efficiency is guaranteed by model training and validation, and performance is gauged by evaluation criteria. For precise predictions, the classification process is optimized by this methodology.

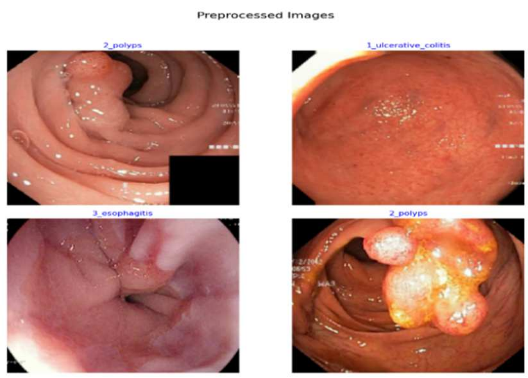

Fig. 4. Samples of Pre-processed Images

Some samples of the pre-processed images are shown in Fig.4. The four categories which includes polyps, ulcerative colitis, esophagitis and normal are depicted in these images. To guarantee constant quality and clarity, the pre-processing stages entail resizing, normalization and improving the images.

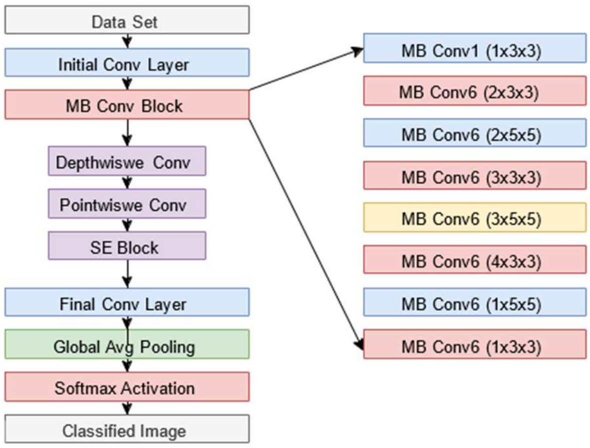

Fig. 5. EfficientNetB2 Architecture

The architecture of a neural network model is depicted in Fig. 5, which details how the dataset passes through several layers, such as MB Conv blocks, SE blocks, and final pooling, before classification. The graphic illustrates the various MB Conv setups as well as the function of pointwise and depth wise convolutions.

C. *EfficientNetB2 Blocks*

An initial convolutional layer is used by EfficientNetB2 to process input images. This is followed by MB Conv blocks, which employ separable convolutions, and Squeeze-and-Excitation (SE) blocks, which extract features efficiently. Performance on the network is optimized by the different kernel sizes and expansion factors. Overfitting is avoided and parameters are decreased using Global Average Pooling (GAP), and the image is categorized using a SoftMax algorithm.

1) *Initial Conv Layer*

The basic characteristics of the input image, like edges and textures, are extracted using a simple convolution technique. To capture little details, it typically uses strides and tiny filters (such 3x3 or 5x5). The output of this layer is a feature map that highlights the important patterns that need to be processed by further layers.

2) *MB Conv Block*

The design incorporates lightweight and effective bottleneck layers with depth-wise separable convolutions in MB Conv blocks. They simplify processing, preserve accuracy, and allow scaling without sacrificing efficiency, making them ideal for mobile and edge devices.

a) *Depth-wise Conv Layer*

Depth-wise convolutions use a different filter for each input channel, which lowers computational complexity and reduces the number of parameters and improves network efficiency. They maintain crucial spatial information by concentrating on spatial features.

$$b_{x,y,z} = \sum_{m=-\frac{z}{2}}^{\frac{z}{2}} \sum_{n=-\frac{z}{2}}^{\frac{z}{2}} a_{x+m,y+n,z} \cdot w_{m,n,z} \quad (1)$$

Equation (1) defines the relationship where, $b_{x,y,z}$ is the output at position (x, y), $a_{x,y,z}$ is the input image at position $(x, y$, $w_{m,n,z}$ is the filter and z is the filter size.

b) *Point-wise Conv Layer*

The point-wise convolution layer improves cross-channel interactions and the model's capacity to recognize intricate patterns by integrating data across channels using 1x1 convolutions.

$$b_{x,y,z'} = \sum_{k=1}^{C} a_{x,y,z} \cdot w_{z,z'} \quad (2)$$

Equation (2) defines the relationship where, $b_{x,y,z'}$ is the output at position (x, y), $a_{x,y,z}$ is the input at position (x, y), $w_{z,z'}$ is the and z'^{th} output channel and C is the number of input channels.

c) *SE Block*

By adaptively recalibrating channels, the SE Block improves feature representations by highlighting significant features and suppressing less pertinent ones, hence increasing feature extraction accuracy and durability.

$$k_z = \frac{1}{h \cdot w} \sum_{x=1}^{h} \sum_{y=1}^{w} a_{x,y,z} \quad (3)$$

Equation (3) defines the relationship where, k_z is the squeezed output, $h \cdot w$ are the height and width of the feature map, $a_{x,y,z}$ is the feature map at position (x, y) and channel z.

3) *Final Conv Layer*

In EfficientNetB2, the final convolutional layer aggregates high-level features from the MBConv blocks, reducing spatial dimensions while preserving essential semantic information for accurate classification.

4) *Global Average Pooling*

Global Average Pooling (GAP) minimizes overfitting, decreases parameters, and highlights important characteristics for classification by condensing each feature map's spatial dimensions to a compact vector.

$$GAP_c = \frac{1}{h \times w} \sum_{x=1}^{h} \sum_{y=1}^{w} F_c(x, y) \quad (4)$$

Equation (4) defines the relationship where, GAP_c is the resulting value for the c^{th} channel, $F_c(x, y)$ is the value of the feature map at special position (x, y) for the c^{th} channel, h, w are the height and width of the feature map.

5) *SoftMax Activation*

Multi-class classification and final prediction, is done by the SoftMax activation.

III. RESULTS AND DISCUSSION

The EfficientNetB2 model demonstrated exceptional performance in the classification of gastrointestinal illnesses, attaining 98.125% testing accuracy, 95.69% validation accuracy, and 99.28% training accuracy. In addition, it showed a high AUC of 99.91%, 97.26% precision, and 97.15% recall.

TABLE I Comparison of Evaluation Metrics of various Methodologies

Methodologies	**Accuracy (%)**	**AUC (%)**	**Precision (%)**	**Recall (%)**
		TRAINING		
MobileNetV2	94.88	99.51	95.09	94.42
Xception	83.37	96.94	86.61	80.52
Custom CNN	87.48	98.20	88.50	86.52
VGG16	97.73	99.91	97.95	97.55
ResNet50	98.58	99.96	98.63	98.57
EfficientNetB3	98.62	99.90	98.73	98.51
		VALIDATION		
MobileNetV2	72.77	91.74	73.04	72.57
Xception	73.61	92.95	79.05	67.33
Custom CNN	77.34	95.17	79.05	76.02
VGG16	83.30	97.25	84.30	81.91
ResNet50	92.85	99.05	93.03	92.48
EfficientNetB3	93.04	99.20	93.53	92.95
		TESTING		
EfficientNetB1 [1]	96.63	99.58	98	97
MobileNetV2	77.01	93.44	77.05	76.79
Xception	77.50	93.75	80.05	69.21
Custom CNN	80.87	96.41	82.22	79.75
VGG16	91.25	97.24	82.72	80.86
ResNet50	94.49	99.32	94.59	93.44
EfficientNetB3	93.46	99.34	93.72	93.39

TABLE II Evaluation Metrics of EfficientNetB2

EfficientNetB2	**Accuracy (%)**	**AUC (%)**	**Precision (%)**	**Recall (%)**
Training	99.28	99.98	99.37	99.27
Validation	95.69	99.69	96.27	95.45
Testing	98.125	99.91	97.26	97.15

Table I compares the deep learning models (MobileNetV2, Xception, Custom CNN, VGG16, ResNet50, and EfficientNetB3).These models were implemented and evaluated in the context of gastrointestinal disease classification using WCE images. For testing, ResNet50 achieves the highest accuracy (94.49%), followed by EfficientNetB3 (93.46%), showcasing their strong generalization. The results of EfficientNetB1 were obtained from [1].

Table II demonstrates the remarkable performance of EfficientNetB2. A model's performance in tasks like classification or prediction is evaluated using performance measures. Accuracy, precision, and recall are crucial metrics that help in the direction of improvement and model comparison.

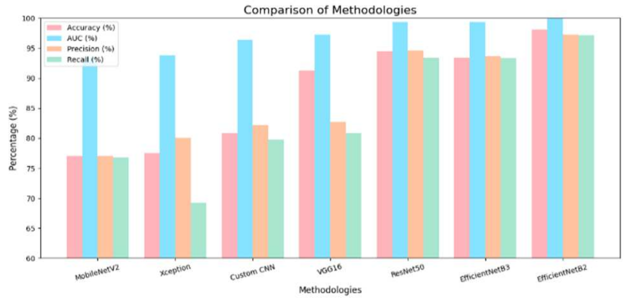

Fig. 6. Comparison of Various Methodologies

In terms of accuracy, precision, AUC and recall EfficientNetB2 performed more effectively than all of the other algorithms as shown in Fig. 6. ResNet50 and VGG16 perform well, but MobileNetV2 performs lowest.

$$Precision = \frac{True\ Positive\ Predictions}{Total\ \ Positive\ Predictions} \tag{5}$$

$$Recall = \frac{True\ Positive\ Predictions}{Total\ \ Actual\ Positives} \tag{6}$$

$$Accuracy = \frac{Correct\ Predictions}{Total\ \ Predictions} \tag{7}$$

The mathematical expressions of Precision, Recall and accuracy are shown in (5), (6) and (7) respectively.

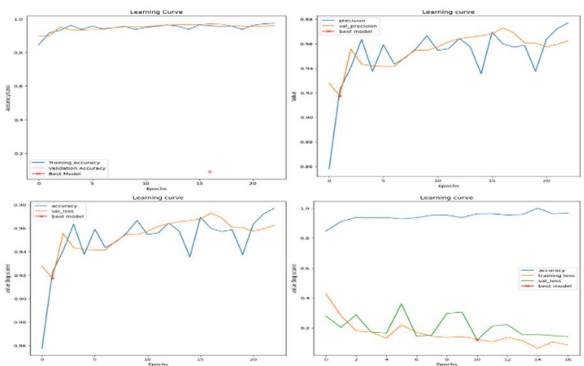

Fig. 7. Plots of EfficientNetB2

The plots in Fig.7 demonstrate consistent improvements in correctness as well as stability in training and validation loss. At first, precision varies but eventually settles. The reliability of the model is reflected in the "best model" scores, which show the best validation performance.

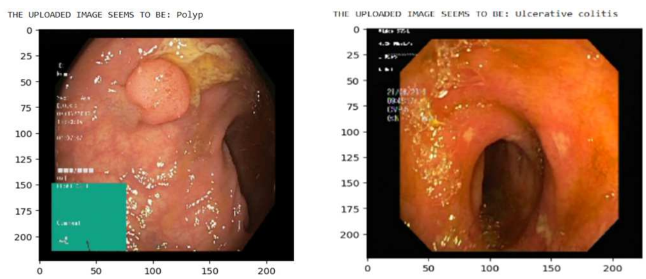

Fig. 8. Predicted Output

The efficiency of the model and its significance in early detection is demonstrated by the right identification of the image in Fig.8 as polyp and ulcerative colitis.

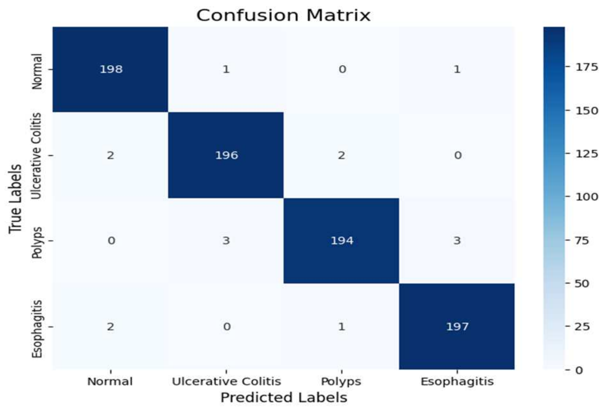

Fig. 9. Confusion Matrix

The model correctly distinguishes between esophagitis and normal instances, with few misclassifications, according to the confusion matrix displayed in Fig.9. However, there is some overlap between polyps and ulcerative colitis, which may indicate that they are occasionally misclassified. In spite of this, the performance as a whole shows good predicting ability.

IV. CONCLUSION

In endoscopy with wireless capsule images, the EfficientNetB2 model has demonstrated remarkable efficacy in the categorization of gastrointestinal illnesses, attaining an astounding accuracy of 98.125%. This performance demonstrates how well the model can find complicated patterns in medical images, enabling prompt and precise diagnosis of gastrointestinal disorders. Its high accuracy increases the trust of diagnostic data, improves patient outcomes, and influences treatment decisions. As a result, this method is a significant advancement in the categorization of gastrointestinal disorders and holds considerable potential for practical medical applications.

V. REFERENCES

[1] S. Al-Otaibi, A. Rehman, M. Mujahid, S. Alotaibi, and T. Saba, "Efficient-Gastro: Optimized EfficientNet model for the detection of gastrointestinal disorders using transfer learning and wireless capsule endoscopy images," in Proc. IEEE Int. Conf. Med. Imaging, vol. 45, no. 3, pp. 123-130, Mar. 2024.

[2] R. Harini, J. Giri, S. Sairam, and N. Ahmad, "A CNN-based system for detection of abnormalities in the gastrointestinal tract from wireless capsule endoscopy images," J. Med. Imaging, vol. 42, no. 2, pp. 345-352, Mar. 2024.

[3] K. Haataja and P. Toivanen, "Review of deep learning performance in wireless capsule endoscopy images for GI disease classification," IEEE Trans. Med. Imaging, vol. 41, no. 6, pp. 543-550, Apr. 2024.

[4] B. Chakraborty and S. Mitra, "Investigating feature fusion-based framework for capsule endoscopy image classification," IEEE Access, vol. 10, no. 5, pp. 234-240, May 2024.

[5] A. Das, A. Singh, N. Nishant, and S. Prakash, "CapsuleNet: A deep learning model to classify GI diseases using EfficientNet-B7," arXiv preprint, arXiv:2410.19151, 2024. [Online]. Available: https://arxiv.org/abs/2410.19151

[6] B. Zhang, Y. Chen, L. Bai, Y. Zhao, Y. Sun, Y. Yuan, J. Zhang, and H. Ren, "Learning to adapt foundation model DINOv2 for capsule endoscopy diagnosis," arXiv preprint, arXiv:2406.10508, Jun. 2024. [Online]. Available: https://arxiv.org/abs/2406.10508

[7] G. Baddam, S. Swaraja, M. Madhavi, and S. Sujatha, "Diagnosis of gastrointestinal diseases using modern CNN techniques," IEEE Sens. J., vol. 23, no. 11, pp. 4079-4086, Jun. 2023.

[8] M. Nouman, M. Sajid, O. Young, and Imran, "Efficient gastrointestinal disease classification using pretrained deep convolutional neural networks," in Proc. IEEE Int. Conf. Artif. Intell., vol. 34, no. 8, pp. 123-130, Nov. 2023.

[9] R. Sadagopan, S. Ravi, S. V. Adithya, and S. Vivekanandhan, "PolyEffNetV1: A CNN-based colorectal polyp detection in colonoscopy images," Proc. Inst. Mech. Eng., Part H: J. Eng. Med., vol. 237, no. 3, pp. 406–418, Mar. 2023. [Online]. Available: https://doi.org/10.1177/09544119221149233

[10] Y. Chandran, S. Sumithra, and C. S. Gnana Dhas, "Gastrointestinal tract disease classification from wireless endoscopy images using pretrained deep learning models," J. Artif. Intell. Res., vol. 17, no. 6, pp. 534-540, Dec. 2022.

[11] M. Dheir and S. S. Abu-Naser, "Classification of anomalies in the gastrointestinal tract using deep learning," Int. J. Acad. Eng. Res. (IJAER), vol. 6, no. 3, pp. 15–28, Mar. 2022.

[12] 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, pp. 4079, Jun. 2022, doi: 10.3390/s22114079.

[13] 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), Piscataway, NJ, USA, 2022, pp. 280–285, doi: 10.1109/IEMCON56893.

[14] J. Escobar, K. Sanchez, C. Hinojosa, H. Arguello, and S. Castillo, "Accurate deep learning-based gastrointestinal disease classification via transfer learning strategy," in Proc. IEEE XXIII Symp. Image, Signal Process. Artif. Vis. (STSIVA), Piscataway, NJ, USA, 2021, pp. 1–5.