IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.10, October 2022 E-ISSN: 2347-2693

A Transfer Learning-Based Efficient Deep-Learning Methodology for Multi-Class Classification of Endoscopy Frames

Madhura Prakash M.[1*], L. Krishnamurthy G.N.[2]

[1]Research Scholar, Department of Computer Science and Engineering, BNMIT, Bangalore, India
[2]Principal, BNMIT, Bangalore, India

[*]Corresponding Author: madhuraprakash18@gmail.com, Tel.: +91-9986330078

DOI: https://doi.org/10.26438/ijcse/v10i10.2328 | Available online at: www.ijcseonline.org

Received: 21/Sept/2022, Accepted: 06/Oct/2022, Published: 31/Oct/2022

Abstract— Computer-aided detection and analysis of anatomical structures and pathology with the support of Artificial Intelligence aids the medical experts by contributing to better utilization of the expert's focus time. The gastrointestinal (GI) tract can be assessed for the presence of several irregularities like ulcers, bleeding, inflammation, polyps, and tumor that can be present in several diseases ranging from precancerous lesions to cancer. These abnormalities differ in their appearance by having a different shape, size variation, color and texture differences, and they generally show up to be outwardly comparable to the regular regions in the GI tract. This presents a challenge in designing an efficient classifier that can handle intra-class variations. An endoscopy procedure is performed to detect and diagnose GI abnormalities and to observe the GI pathology. A sequence of video frames of the GI region is captured during the investigation of the tract. A flexible tube with a camera-fitted at the end is injected into the patient's body through the oral or rectum during the procedure. The frames captured can be analyzed for abnormality classification and lesion segmentation. The analysis is challenging because the frames may have low contrast, uniform background, color variations and indefinite lesion shapes. This makes the segmentation and classification on these frames a challenging task. In this effort, a transfer learning-based deep learning architecture has been employed for performing the multi-class classification of endoscopy frames. The proposed model has been trained and tested on the widely available KVASIR dataset and an average accuracy of 81% has been achieved.

***Keywords*—**Gastrointestinal tract, Transfer Learning, Deep Learning architecture, Classification

I. INTRODUCTION

Several diseases can have an impact on the digestive system of humans. The gastrointestinal (GI) tract accounts for locating commonly occurring cancers. Each year, there are roughly 2.8 million new cases of esophageal, stomach, and colorectal cancer and 1.8 million fatalities [1]. Examinations using endoscope are done for analyzing the GI tract to examine the pathology and detect any abnormalities that are present. Endoscopy examinations use digital, high-definition endoscopes to do live, video examinations of the GI tract. An endoscope is an illuminated optical device that is used to observe the GI tract. The videos thus captured are assessed by the doctors for assessing the presence of any abnormalities. This assessment is time-consuming and laborious for the expert medical staff. Also, manual assessments can be prone to errors. Disease classification is crucially essential because it can affect how treatments are chosen.

An automated disease identification system is therefore the need of the hour to aid the medical fraternity and also to better utilize the expert's time. Inequalities will be decreased, quality will be raised, and the few medical resources will be used more effectively with automatic detection, recognition, and assessment of abnormal results.Understanding the visual content in an image is very important in the case of disease detection and classification. This is one of the major undertaking of Computer vision-based algorithms. Since identifying abnormalities in the endoscopy frames is a very difficult work as, the frame quality may not be upright. This can result in the loss of crucial information. It is time-coonsuming for the expert to analyze this data manually and the analysis maybe prone to human error. In addition, various types of lesions can look quite alike to one another, and the same type of lesion might exhibit variations in color and shape. These factors make it difficult to distinguish between different types of lesions. To help with automatic analysis of the endoscopic frames, computer-aided diagnostic (CAD) systems have been developed.

The amount of images that the doctor must study can be significantly reduced by the spontaneous detection of abnormalities on the endoscopic frames, allowing the specialist to concentrate only on the pertinent data. There are four categories of GI lesions: lymphangiectasias, polyps, inflammatory lesions, and vascular lesions [2]. The likelihood of the patients surviving can be increased if the precancerous abrasions are discovered in an early stage before advancement. Deep learning architectures have been performing very well in recent years, especially in the classification and segmentation of images. These architectures have been used to analyze the medical images generated from different modalities to perform disease classification and segmentation.In this effort, a deep learning architecture based on transfer learning has been proposed to perform multi-class classification of the GI tract pathology and other abnormalities. The KVASIR dataset has been employed to train and test the model architecture.

The rest of the paper is organized as takes after : section 2 gives an outline of the later works that have utilized profound learning-based design for endoscopy image classification. Section 3 gives the details of the dataset and the execution outline. Section 4 presents the outcomes. The concluding comments are given in section 5.

II. RELATED WORK

The researchers in [3] have performed a multi-class classification of the Wireless Capsule Endoscopy (WCE) frames for the lesions found in the Gastrointestinal Tract. The 4 types of lesions considered here are “(a) Angiodisplasya, the typical vascular lesion, (b) Ulcer; (c) Lymphangiectasia, and (d) Polyp”. The dataset from the GIANA Endoscopic Vision Challenge was employed for classification of lesions. The images had been pre-processed using Adaptive Contrast Diffusion, Homographic filtering, and Multi-Scale Retinex with Colour Reestablishment. Different variants of DenseNet and ResNet were implemented and the performance of these models has been analyzed by the researchers. The proposed model had achieved precision and recall of 90 percent.

The authors in [4] implemented an image-saliency model by creating a bag-of-words. An architecture based on convolution neural network (CNN) to multi-scale feature extraction was used. The proposed method was based on weakly-supervised learning based on high-level semantic annotations. The methodology was evaluated on a widely available KID dataset.

The work in [5] focused on a mathematical approach for discriminating ulcer versus non-ulcer picture elements in different color spaces. For maximum efficiency, the extracted feature vector was used to compute the performance measures using SVM and a grid search approach. The performance in terms of accuracy, sensitivity, and specificity were 97.89%, 96.22%, and 95.09%, respectively.

The training in [6] adopted a deep network architecture and a model named “bleedy image recognizer” was trained to classify WCE images with blood. Here the MobileNet was integrated with a customized CNN model. The model was trained and tested on a dataset of 1650 WCE images, and its performance was assessed using the metrics of accuracy, precision, recall, F1 score, and Cohen's kappa. The outcomes showed accuracy, precision, recall, F1 score, and Cohen's kappa values were correspondingly 0.993, 1.000, 0.994, 0.997, and 0.995.

The authors in [7] proposed a fusion convolutional neural network for irregularity detection in endoscopy frames. Three concurrent convolutional neural networks, each of which has unique feature learning capabilities were used. While the second architecture used cosine-normalized convolution, the first network used depthwise separable convolution. In order to obtain relationships from the statistical data gathered from the features produced by the first and second phases, a meta-feature extraction method was included in the third phase. On the KID and KVASIR datasets, the network model was developed and evaluated. According to the model, categorization accuracy was 98%. For the classification of gastrointestinal disorders and the identification of ulcers, a deep learning-based technique was proposed in [8]. The segmentation of ulcers using a customized mask Recurrent Convolutional Neural Network (RCNN) was suggested here. A bounding box was generated around the ulcer detected area and a segmented mask region were produced by training the Mask RCNN model with ulcer annotated images. The ResNet101 pre-trained CNN model was improved by transfer learning to produce features during the classification phase. For the final classification, the features were fed into a Multi-class Support Vector Machine (MSVM) with a cubic kernel function. According to the cubic SVM, classification accuracy of 99.13% is attained.

In [9] a correlation coefficient-based feature selection approach was combined with principal component analysis (PCA) for feature selection. Support vector machine was used to classify the extracted features on multiple classes. The methodology was evaluated on images that depicted ulcer, bleeding versus healthy. The performance accuracy was 98.3%. In [8] Weighted k-Nearest-Neighbor (WKNN) classifier was applied on the features extracted from hue, saturation, lightness (HSV) color space. Accuracy of 98.8%, has been claimed by the authors.

In [10] feature extraction was performed based on Speeded Up Robust Features (SURF). The color information from images were obtained. The features extracted were then used to train and test a Support Vector Machine (SVM) so as to classify the images into normal or abnormal. An accuracy of 94.58% was achieved for binary classification and an accuracy of 82.91% for multi-class classification.

A pre-trained deep convolutional neural network VGG-16 model was used in [10] for image categorization into anatomical landmarks, abnormalities of pathology, and identification of endoscopic procedures. A transfer learning technique was implemented by the authors for generation of solution. Additionally, data augmentation was carried out to emphasize the significance of the requirement of large data for deep models. Prior to and during data augmentation, the accuracy rates were 96.9% and 98.8%, respectively.

The researchers in [11] used a pre-trained CNN namely AlexNet for endoscopy image classification. A denoising convolutional neural network (DnCNNs) was used as a advancement tool. Classification accuracy of 90.17% has been achieved on KVASIR medical images. In [12], conspicuous convolutional features were used to represent images. The convolutional kernels from the first layer of a pre-trained convolutional neural network (CNN) are examined based on their sensitivity to colors and textures, and organized into several separate groups. The dominant features in each cluster are combined using spatial maximal activator pooling (SMAP) into a feature map. This map preserves the layout information. The aggregated feature map was utilized to generate a progressing window-based organized pooling method to extract spatial layout features and global shape data to populate feature histograms.

The authors in [13] have used a distillation-based technique for training deep learning models efficiently by employing a student -teacher based architecture. An F1 -score of 4.7% was obtained for the endoscopy dataset from KVASIR and 3.2% for the colonoscopy dataset from HyperKvasir. Three cutting-edge architectures, “Inception-ResNet-v2, Inception-v4, and NASNet”, were trained on the Kvasir dataset for the study [14], and their classification performance was examined using validation data. 85% of the Kvasir dataset's photos are used for training per epoch, and the remaining 15% are held back for validation. “Inception-v4, Inception-ResNet-v2, and NASNet” all contributed to the accuracy results, which were 0.9845, 0.9848, and 0.9735, respectively.

The authors in [15] combined Residual Neural Network and Faster R-CNN for multi-class classification on the KVASIR dataset. The authors generated 4900 bounding boxes annotations emphasizing some classes from the KVASIR dataset. These annotations are publicly available. In [16] a convolution neural network called “HarDNet-MSEG” was used for polyp segmentation. The network consisted of a backbone and a decoder. The backbone was a low memory CNN called HarDNet68. The decoder part was a Cascaded Partial Decoder. A mean dice of 0.904 was achieved by the researchers. The researchers in [17] have used deep learning for classifying rh colorectal image from endoscopy. A ResNet50 architecture was used and a regularization technique called “DropBlock” was used to decrease overfitting and manage the presence of noise and other image artifacts.

III. MEHODOLOGY

A. PreprocessIng

The pre-processing techniques employed are

- Downsampling: The images have been downsampled to a size of 128*128 using the cv2.INTER_AREA interpolation. An interpolation function has been used to examine neighborhoods of pixels and use these neighborhoods to optically decrease the size of the image with introducing minimal distortions
- Normalization: The technique of envisaging picture data pixels (intensity) to a pre-defined range (often (0,1) or (-1, 1)) is carried out in normalization. It is sometimes known as data rescaling.[18]
- Data Augmentation: Without gathering new data, small adjustments have been made to the dataset's current data to boost its diversity. It is a strategy for expanding the dataset. Data augmentation methods like flipping data horizontally and vertically, rotating data, cropping data, shearing data are common in image dataset. Data augmentation helps in restricting a neural network from picking up unrelated features.
- Image standardization: Similar heights and widths have been achieved through scaling and preprocessing of the images. The information has been scaled down to have a mean of 0 and a standard deviation of 1 (unit variance). Information consistency and quality are upgraded through standardization.

B. Dataset

The “KVASIR Dataset” [1] consists of eight separate classes, each with 500 pictures, and contains 4000 endoscopic gastrointestinal disorders. Each category in the dataset has multiple sets of photos, including anatomical landmarks and pathological abnormalities. “Z-line, pylorus, and cecum” are anatomical markers, and “esophagitis, polyps, and ulcerative colitis” are pathological findings. The images depicting “dyed and lifted polyps and dyed resection margins”, are related to the obliteration of lesions. The collection contains images with various dimensions ranging from 720 576 to 1920 1072 pixels. Using an electromagnetic imaging technology, a green image in some of the given classes of images shows the location and constitution of the endoscope inside the intestine. A few of the representative frames from each of the classes in the “KVASIR dataset” have been depicted in figure 1 below.

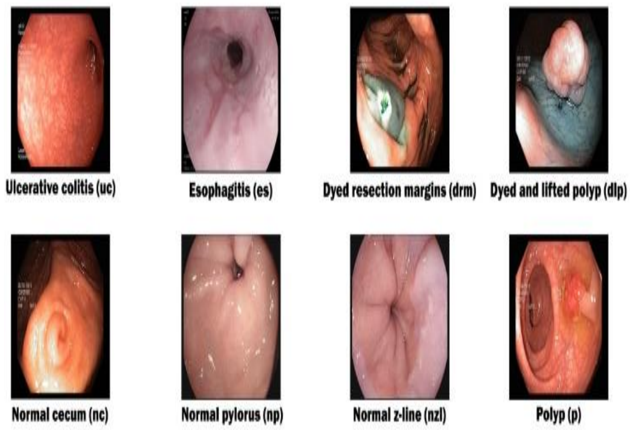


Figure 1. Representative frames from each of the classes in the “KVASIR dataset” [1]

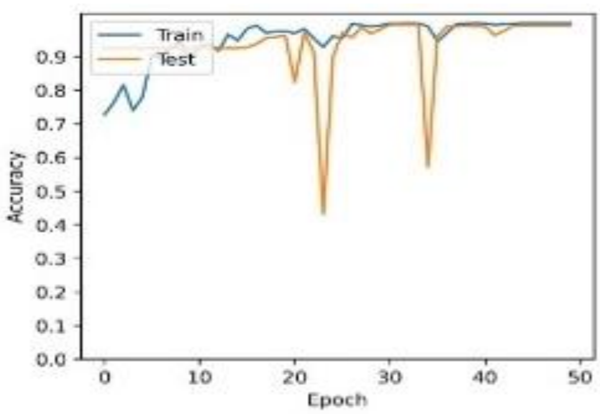
C. Model Architecture

A VGG19 pre-trained CNN architecture has been used by employing the transfer learning procedure to load the initial weights for the neural network architecture. The pre-trained VGG19 architecture [19] is combined with the sequential layer CNN architecture. Normalizing the output of the previous layers is delegated to the batch normalization layer. The intermediary layers and the dense layer have been incorporated with ReLu activation function. The output layer is a dense layer with a softmax activation to enable multi-class classification. The model summary is shown in figure 2.

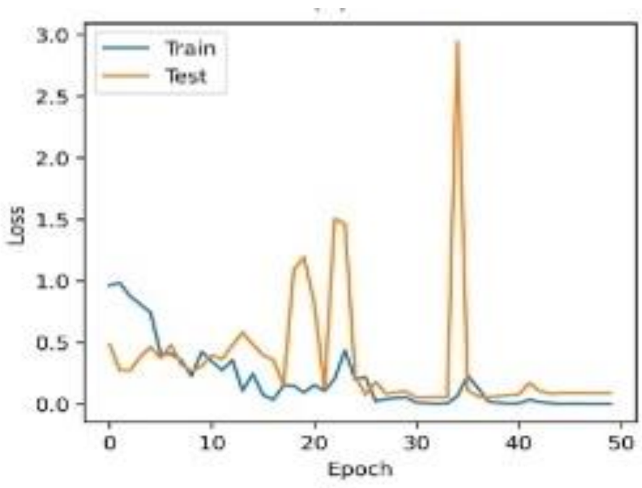
```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         (None, 128, 128, 3)       0
_________________________________________________________________
batch_normalization_1 (Batch (None, 128, 128, 3)       12
_________________________________________________________________
vgg19 (Model)                (None, 4, 4, 512)         20024384
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 4, 4, 128)         65664
_________________________________________________________________
flatten_1 (Flatten)          (None, 2048)              0
_________________________________________________________________
dense_1 (Dense)              (None, 512)               1049088
_________________________________________________________________
dense_2 (Dense)              (None, 8)                 4104
=================================================================
Total params: 21,143,252
Trainable params: 21,143,246
Non-trainable params: 6
_________________________________________________________________
```

Figure 2. Model summary

The model is compiled with the Adam optimizer with an original learning rate of 2e-5 and a decay of 1e-6. The model has been trained with the categorial cross-entropy loss function. The learning rate has been reduced on the plateau with a factor of 0.8 by monitoring the validation loss with a patience of 7. Adam optimiser aids in reaching a global minima while training out model. Early stopping has been included with patience of 12. Early Stopping helps in halting the training of the model early if there is no increase in the parameter. The validation accuracy in the parameter being monitored. The model has been trained on 50 epochs. The model's accuracy and loss during training are as depicted in the figure



(a)



(b)

Figure 3. Model accuracy and loss incurred during training

D. ImplementatIon DetaIls

The implementation has been done using TensorFlow in the backend. To improve the model performance and generalizability image augmentation technique has been used with the help of the image data generator module. Augmentation techniques have been done by using the shear range, zoom range, and rotation range. The horizontal flipping has been made active and the width and height shift range of 0.1 has been used.

Transfer learning has been employed to design the model architecture A standard model is utilized as the basis for a new model using the deep learning practice known as transfer learning. In this situation, the model is used to a uphold, similar work as an optimization to enable rapidly progress when modeling the second task [20]. Transfer learning enables better optimization as it allows the model to start training for a relevant point of network weights instead of starting from random allocation of weights.

The network training in deep learning generally begins with a random assignment of weights to all the edges in the network and gradually fine-tuning these weights to appropriate values based on the given data. The fine-tuning of weight is based on the loss calculation at each run and computing the contribution of loss from current weights. The weight upgradation depends on the several factors like the error gradient and learning rate of the architecture. The network training procedure can be better optimized my initializing the weights at a standard relevant point. The standard CNN models which can identify and classify general objects can used and their weights can be used as a starting relevant point. The schematic diagram of a transfer learning procedure is shown in figure 4.

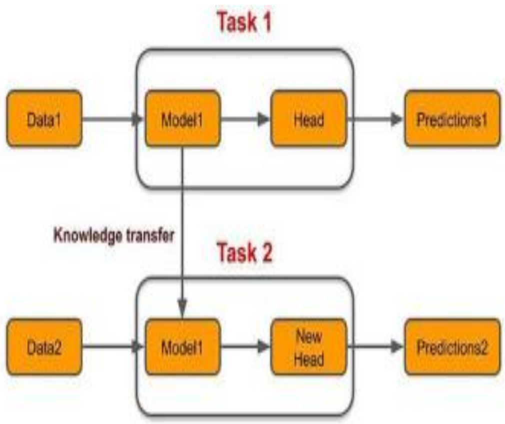


Figure 4. Schematic representation of a transfer learning procedure

IV. RESULTS AND DISCUSSION

E. EvaluatIon MetrIcs

The ratio of the number of true positives to the number of false positives, or tp/ (tp + fp), is utilized to quantify precision. Precision is the classifier's inherent potential to avoid classifying a negative sample as positive.

The ratio of the amount of true positives to the quantity of false negatives, tp/ (tp + fn), is termed as the recall. Recall is fundamentally the classifier's competence to discover all of the positive samples.

A weighted harmonic mean of the precision and recall can be employed to comprehend the F-beta score, which has a greatest value of 1 and a worst value of 0. The amount of instances of each class is the support [21].

F. Results

Twenty percentage of the considered data was used in validating the model and the model has been tested on the test set. The test set included original images from the data source and the variedly augmented images from the source. The model has achieved a classification performance of f1 score, and support of 0.97,0.98, 0.99, and 117 respectively. This result signifies a good performance of the model.

Classification Report

	precision	recall	f1-score	support
normal-pylorus	0.97	0.98	0.97	86
normal-z-line	0.58	0.99	0.73	89
dyed-resection-margins	0.89	0.67	0.76	99
ulcerative-colitis	0.83	0.82	0.82	87
dyed-lifted-polyps	0.76	0.93	0.84	115
normal-cecum	0.88	0.96	0.92	110
esophagitis	0.96	0.47	0.63	117
polyps	0.86	0.74	0.80	97
accuracy			0.81	800
macro avg	0.84	0.82	0.81	800
weighted avg	0.84	0.81	0.81	800

Fig 5. Model performance

V. CONCLUSION

In this work, a deep learning-based solution has been proposed for the multi-class classification of endoscopy images. The transfer learning procedure has been implemented using the pre-trained weights of the VGG19 model. The main advantage of using a transfer learning technique for deep learning training procedure is to have a standard starting point of the network weights. This procedure makes the learning process efficient, thereby optimizing the model training process. Other standard CNN models like ResNet, MobileNet, and EfficientNet can be explored for better performance in transfer learning.

REFERENCES

[1] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pål Halvorsen. 2017. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). Association for Computing Machinery, New York, NY, USA, **164–169, 2017.**

[2] Khan, U.A., Javed, A. & Ashraf, R. An effective hybrid framework for content based image retrieval (CBIR). Multimed Tools **Appl 80, 26911–26937, 2021.**

[3] Valério, Maria & Gomes, Sara & Salgado, Marta & Oliveira, Helder & Cunha, António. (2019). Lesions Multiclass Classification in Endoscopic Capsule Frames. Procedia Computer Science. **164. 637-645, 2019.**

[4] Vasilakakis, M.D., Diamantis, D., Spyrou, E. et al. Weakly supervised multilabel classification for semantic interpretation of endoscopy video frames. Evolving Systems **11, 409–421, 2020.**

[5] Suman, S.; Hussin, F.A.; Malik, A.S.; Ho, S.H.; Hilmi, I.; Leow, A.H.-R.; Goh, K.-L. Feature Selection and Classification of Ulcerated Lesions Using Statistical Analysis for WCE Images. Appl. Sci., **07,1097, 2017.**

[6] F. Rustam et al., "Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model," in IEEE Access, **Vol. 9, pp. 33675-33688, 2021.**

[7] Jain, S., Seal, A., & Ojha, A. (2022). A Hybrid Convolutional Neural Network with Meta Feature Learning for Abnormality Detection in Wireless Capsule Endoscopy Images. ArXiv, abs/**2207.09769, 2022.**

[8] Mehshan Ahmed Khan, Muhammad Attique Khan, Fawad Ahmed, Mamta Mittal, Lalit Mohan Goyal, D. Jude Hemanth, Suresh Chandra Satapathy, Gastrointestinal diseases segmentation and classification based on duo-deep architectures, Pattern Recognition Letters, ISSN 0167-8655. Vol.131, pp.193-204, 2020.

[9] Journal of Mechanics in Medicine and Biology, automated ulcer and bleeding classification from wce images using multiple features fusion and selectionamna liaqat, muhammad attique khan, jamal hussain shah, muhammad sharif, mussarat yasmin and steven lawrence fernandes, Vol. 18, no. 04.

[10] Mamun, Abdullah & Hossain, Md & Hossain, Motaher & Hasan, Galib. (2019). Discretion Way for Bleeding Detection in Wireless Capsule Endoscopy Images. 1-6. 10.1109/ICASERT.2019.8934589.

[11] Vats, Vanshika, et al. "SURF-SVM based identification and classification of gastrointestinal diseases in wireless capsule endoscopy." arXiv preprint arXiv:**2009.01179, 2020.**

[12] Benkessirat, Amina & Benblidia, Nadjia & Beghdadi, Azeddine. (2021). Gastrointestinal image classification based

on VGG16 and transfer learning. 1-5. 10.1109/ICISAT54145.2021.9678481.

[13] Ahmed, A. (2022). Classification of Gastrointestinal Images Based on Transfer Learning and Denoising Convolutional Neural Networks. In: Saraswat, M., Roy, S., Chowdhury, C., Gandomi, A.H. (eds) Proceedings of International Conference on Data Science and Applications. Lecture Notes in Networks and Systems, Springer, Singapore. **Vol 288, 2022.**

[14] Ahmad, J., Muhammad, K., Lee, M.Y. et al. Endoscopic Image Classification and Retrieval using Clustered Convolutional Features. J Med Syst **41, 196, 2017.**

[15] H. L. Gjestang, S. A. Hicks, V. Thambawita, P. Halvorsen and M. A. Riegler, "A self-learning teacher-student framework for gastrointestinal image classification," 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), **pp. 539-544, 2021.**

[16] Timothy Cogan, Maribeth Cogan, Lakshman Tamil, MAPGI: Accurate identification of anatomical landmarks and diseased tissue in the gastrointestinal tract using deep learning, Computers in Biology and Medicine, Volume 111,2019,103351, ISSN 0010-4825.

[17] Trung-Hieu Hoang, Hai-Dang Nguyen, Viet-Anh Nguyen, Thanh-An Nguyen, Vinh-Tiep Nguyen, and Minh-Triet Tran. 2019. Enhancing Endoscopic Image Classification with Symptom Localization and Data Augmentation. In Proceedings of the 27th ACM International Conference on Multimedia (MM '19). Association for Computing Machinery, New York, NY, USA, **2578–2582, 2019.**

[18] Huang, C. H., Wu, H. Y., & Lin, Y. L. (2021). Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. arXiv preprint arXiv:2101.07172.

[19] S. Poudel, Y. J. Kim, D. M. Vo and S. -W. Lee, "Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network," in IEEE Access, **vol. 8, pp. 99227-99238, 2020.**

[20] Li, Qiang & Yang, Yingjian & Guo, Yingwei & Li, Wei & Liu, Yang & Liu, Han & Kang, Yan. (2021). Performance Evaluation of Deep Learning Classification Network for Image Features. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3048956.

[21] Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI. **2, 420, 2021.**

AUTHORS PROFILE

Madhura Prakash M is working as a Technical Specialist with the Research and Development team at Forus Health Private Limited, Bangalore. She is also a Research Scholar at the Research center of Department of Computer Science and Engineering, BNM Institute of Technology, Bangalore. Her research interests include developing Artificial Intelligence backed Computer Vision techniques for extracting insights and scene understanding in medical videos and images.

Dr.Krishnamurthy G N is currently working as a Prinipal at BNM Institute of Technology. He has guided several Ph.D. scholars. He has served as Registrar (Evaluation) of Visvesvaraya Technological University, Belgaum, for a period of 2 Years on deputation from BNM Institute of Technology during the period from 22nd September 2010 to 21st September 2012 and again from 4th December 2012 to 2nd February 2013.

