Oncologie
DOI: 10.32604/oncologie.2020.013870
Article

Tech Science Press

Two Stage Classification with CNN for Colorectal Cancer Detection

Pallabi Sharma[1,*], Kangkana Bora[2], Kunio Kasugai[3] and Bunil Kumar Balabantaray[1]

[1]Department of Computer Science and Engineering, National Institute of Technology Meghalaya, Shillong, 793003, India
[2]Computer Science and Information Technology, Cotton University, Guwahati, 781001, India
[3]Department of Gastroenterology, Aichi Medical University, Nagakute, 480-1195, Japan
[*]Corresponding Author: Pallabi Sharma. Email: pallabishrma@nitm.ac.in

Abstract: In this paper, we address a current problem in medical image processing, the detection of colorectal cancer from colonoscopy videos. According to worldwide cancer statistics, colorectal cancer is one of the most common cancers. The process of screening and the removal of pre-cancerous cells from the large intestine is a crucial task to date. The traditional manual process is dependent on the expertise of the medical practitioner. In this paper, a two-stage classification is proposed to detect colorectal cancer. In the first stage, frames of colonoscopy video are extracted and are rated as significant if it contains a polyp, and these results are then aggregated in a second stage to come to an overall decision concerning the final classification of that frame to be neoplastic and non-neoplastic. In doing so, a comparative study is being made by considering the applicability of deep learning to perform this two-stage classification. The CNN models namely VGG16, VGG19, Inception V3, Xception, GoogLeNet, ResNet50, ResNet100, DenseNet, NASNetMobile, MobilenetV2, InceptionResNetV2 and fine-tuned version of each model is evaluated. It is observed that the VGG19 model is the best deep learning method for colonoscopy image diagnosis.

Keywords: Colon cancer; deep learning; polyps detection; CNN; colonoscopy

1 Introduction

The medical image processing application introduced in this paper is the automated supervision of colorectal cancer from colonoscopy videos. According to the World Health Organization (WHO), colorectal cancer is the third most common cancer after lung and breast cancer and the second most common cause of death from cancer worldwide [1]. Colorectal cancer is the development of malignant polyps (small chunks of an abnormally grown cell) in the large intestine, including the colon and rectum. So, it is necessary to identify and remove polyps early before it turns cancerous. Advances in early detection techniques help to reduce the vulnerability of cancer significantly. Colonoscopy is commonly used in the early detection of colorectal cancer. A colonic cancer patient should undergo a colonoscopy test every three years in compliance with the WHO [2,3]. During a colonoscopy examination, a long, flexible tube (colonoscope) is inserted into the rectum. A small camera at the tip of the colonoscope allows the doctor to inspect the inside of the colon on a screen and examine a polyp. If no abnormality is observed, then the polyp is surgically removed. On the other hand, if any abnormality is found, the further procedure requires the collection of biopsy tissue samples. This manual identification of anomaly is not an ideal approach because the process entirely depends on the proficiency or expertise of medical practitioners and technicians; moreover, it is time-consuming and involves observer biases. Researchers are working to bridge the gap between expert decision making and manual decision making. So, the successful removal of pre-cancerous

polyps with the help of automated computer-assisted techniques can provide an edge in reducing the mortality rate due to colorectal cancer.

To achieve this ultimate goal of colorectal cancer detection, a two-stage classification technique is proposed in this paper. In the first stage of classification, significant frames are extracted from all frames obtained from colonoscopy video. Significant frames indicate those frames which contain polyp information. The result of the first stage is aggregated in a second stage to reach an overall decision concerning the final classification of that frame to be neoplastic or non-neoplastic. A comparative assessment of 11 different Convolutional Neural Network (CNN) models have been performed to achieve this two-stage classification. The experiments have been performed on a generated dataset. The CNN models namely VGG16, VGG19, Inception V3, Xception, GoogLeNet, ResNet50, ResNet100, DenseNet, NASNetMobile, MobilenetV2, ResNet50V2 and fine-tuned version of each model is evaluated. Prior processing is also performed to further improve the results by removing camera information, text information, and specularity removal. This extensive study proves that the fine-tuned version of VGG19 outperforms all other models and can be applied to achieve automatic detection. This is an application-based study where an extensive in-depth analysis of different CNN models has been performed, which is by far the first work in this direction. The paper's key contribution can be summarized in the following points.

Evaluation of CNN as it applies in automatic significant frame detection: The camera present on the colonoscope sends the inside view of the colon as a video, and each video contains lots of uninformative frames that do not contain any polyp. Screening all frames individually in a video clip by the doctors to identify a polyp is tedious and meticulous. So this automated technique can assist doctors in the identification of significant frames.

Categorization of polyp into neoplastic or non-neoplastic: Non-neoplastic polyps are normal polyps, but if not treated, it can be cancerous over time. So, early removal of these polyps is highly essential. This automated classifier based on recent advanced CNN techniques can help in the early detection of malignancy in a polyp.

The paper is structured as follows–Section 2 provides some insight into existing works in this area. Section 3 presents an explanation of the proposed research that also provides details on methodologies. The findings and discussions are explained in Section 4. Finally, Section 5 includes the summary along with the future research perspective related to this domain.

2 Prior Work

Researchers are working on colonoscopy images and videos for different objectives to solve the problem of colon cancer detection and therapy planning for a long time. This paper mainly focuses on significant frame selection and polyp classification. The literature survey of this work mainly focuses on the specific works done related to our identified objectives. The literature survey is performed on two levels. In the first level, the study related to automated significant frame selection is included. The second level study focuses on the automated classification of polyp into neoplastic and non-neoplastic. The summary of the first level study is listed in a tabular form in Tab. 1. Brief detail on the literature related to the second-level study is listed in Tab. 2.

Table 1: Related work to detect significant frames for analysis of colorectal cancer

Author, Year	Methods used	Dataset	Publicly Available	Data description	Findings
Akbari et al. [4], 2018	CNN with binarized weight	Asu Mayo test clinic dataset	No	Twelve thousand eight hundred seventy-two frames were extracted from 14 colonoscopy videos for training and 4702 frames from 4 colonoscopy videos for testing.	Achieved accuracy-90.28% and binarization of weight and kernel reduce the size of the network leads to easier implementation in hardware.

Wang et al. [5], 2018	Deep learning algorithm.	developed image and video data set for training, and CVC-Clinical DB for validation	No	5,545 colonoscopy images, acquired using "Olympus EVIS LUCERA CV260 (SL)/CV290 (SL)" and "Fujinon 4400/ 4450 HD" at Tokyo, Japan	It achieved sensitivity of 94.38% (95% confidence interval (CI):93.80%, 94.96%) for the developed dataset and 88.24% (95% CI: 85.76%, 90.72%) for CVC Clinical DB dataset respectively.
Urban et al. [6], 2018	VGG19, VGG16, ResNet50 with data augmentation	Own generated dataset	No	8641 colonoscopy images contained 4,088 images of unique polyps of all sizes and morphologies and 4553 images without polyps stored at a resolution of 640 × 480 pixels	CNN initialized on VGG19 pre-trained on Imagenet dataset achieved the highest accuracy, AUC Sensitivity at 5% FNR, Sensitivity at 1% FNR 96.4 ± 0.3% 0.991 ± 0.001 96.9% 88.1%, respectively.
Sundaram et al. [7], 2019	Wiener filter→K-Means→ SGLDM (Spatial Gray-level dependency matrices)→SVM2	-	No	Wireless Capsule Endoscopy (WCE) images	This method achieved 96% sensitivity, 95.4% specificity and 95.7% accuracy in malignancy detection.
Patel et al. [8], 2020	VGG, ResNet, DenseNet, MnasNet, SENet	MICCAI 2017 Dataset, CVC ColonDB, ISIT-UMR Colonoscopy Dataset	Yes	Colonoscopy video	VGG19 achieved highest accuracy-79.78% and 83.52 F1-score.
Hasan et al. [9], 2020	contourlet transformation and CNN as a featues extractor and SVM as a Classifier	2015 MICCIA sub-challenge, Colon Video dataset	2015 MICCIA sub-challenge is public, Video dataset is not public	Endoscopy video	VGG19 outperforms ResNet50, VGG16 and AlexNet with an accuracy of 0.9619 and F1-score of 0.9609

Table 2: Related work to detect neoplastic and non-neoplastic polyps detection for colorectal cancer

Author, Year	Methods Used	Dataset	Publicly Available	Data description	Findings
Sundaram et al. [7], 2019	Wiener filter→K-Means→ SGLDM→SVM2	-	No	Wireless Capsule Endoscopy (WCE) images	This method gives 96% sensitivity, 95.4% specificity, and 95.7% accuracy in malignancy detection.
Barrientos et al. [10], 2020	CNN composed 4 convolutional layers followed by a max-pooling and VGG16 fine-tuned with RMS -prop optimizer.		No	600 images having 5 different resolution	VGG16 Fine-tuned gives accuracy-0.83, Precision-0.81, Recall-0.86 and F1-score-0.83.
Ghesu et al. [11], 2017	Behavior learning using deep reinforcement learning and multiscale image analysis	-	No	-	Extract anatomical structure, and robust against outliers. It can be used for multiple object detection.
Jakob Nikolas et al. [12], 2019	VGG19, AlexNet, ResNet50, GoogLeNet	-	No	-	CNN is good at assessing micro-environment human tumor and can directly predict prognosis from histopathology images and gives nine- class accuracy of >94%
Simon Grahm et al. [13], 2019	CNN with MIL (Minimal Information Loss) and residual units, and Atrous spatial pyramid pooling (ASPP) combined with the output of Deep Neural Network	GlaS, CRAG dataset	Yes	16 H&E stained histological WSIs, scanned with a Zeiss MI−RAX MIDI Slideb Scanner with a pixel resolution of 0.465 m/pixel	MILD Net gives MILD- Net gives F1-score 0.825, Object level (obj.) Dice 0.875 and obj. Hausdorff 160.14, which are higher than DCNN, DeepLab, Seg-Net, U−Net, FCN−8

From the literature, it is observed that although conventional machine learning techniques play a vital role in the automatic detection of colorectal cancer, it is still unable to meet the level of an expert endoscopist decision. Recently, deep learning techniques achieved satisfactory performance in many medical image processing applications, namely breast cancer analysis [14–18], lung cancer [19–21], skin cancer [22], etc. Very few works (presented in Tab. 1 and Tab. 2) have been performed using deep learning in the domain of colorectal cancer detection [4–13], and there are a lot more scopes available for improvements in that direction. For example, Khan et al. [23] presented ResNet101 based classification and mask recurrent CNN based segmentation scheme, where it fails to detect polyps in the bleeding region. There are some more challenges related to this study; such as high specularity, smaller polyps, flat polyps, and polyps in the left colon may be missed, polyps having less clear boundary may be missed, if an algorithm depends on the full appearance of a polyp, then the polyps behind a fold of the colon may not be detected.

3 Methodology

The entire process of detection and removal of polyps is a pipeline of different image processing tasks. Before applying deep learning for classification, a series of image processing task is performed to achieve the goal of polyps detection. The generalized workflow for this work is shown in Fig. 1.

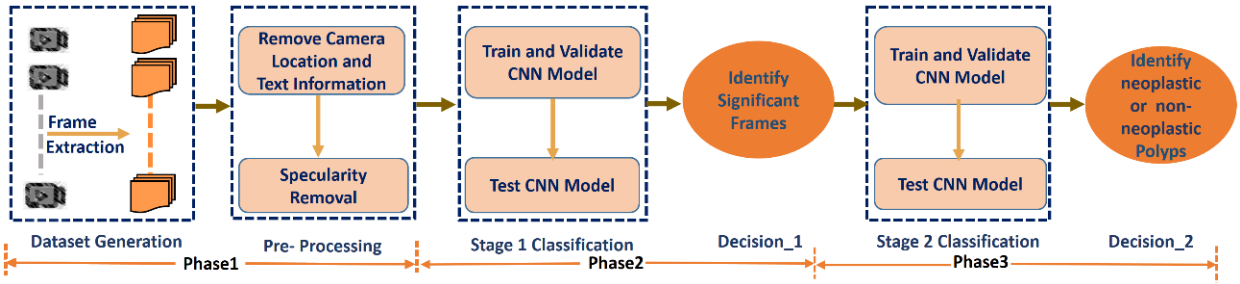


Figure 1: Generalized methodology followed in this work

As displayed in Fig. 1, the work has been completed in three phases. In Phase 1, preprocessing is performed to remove camera information, text information, and specularity removal. In Phase 2, the first stage of classification is performed to identify the significant frames. The final classification is performed in Phase 3. The different methodologies involved is explained below.

3.1 Data Acquisition

The biggest challenge in medical image processing is data acquisition. This is due to the reasons for requirements of ethical clearance, unable to receive patient consent, the problem in database standardization, and ground truth preparation. The involvement of medical experts is another basic need of the study to understand the requirements for database design. To avoid this phase, most researchers end up their study by using a publicly available database. Nevertheless, database generation is very much important to understand the need for any automated analysis. That is why this work has given an immense focus on database generation. But, for the comparison of the proposed work with existing algorithms, the training and testing on the benchmark database are highly essential to check the consistency of the performance. Keeping all these points into consideration, this work has been performed on both generated and publicly available benchmark databases. Following are the database details used for all the experiments.

Database 1 (DB1): DB1 is the database generated under the supervision of Kunio Kasugai at the Department of Gastroenterology, Aichi Medical University, Nagakute, Japan. The dataset consists of colonoscopy videos recorded with Narrowband Imaging (NBI) and White Light (WL). The expert has selected the frames which contain visible polyps. For Stage 1 classification, 900 WL images are available where 400 images contain polyps, and the rest of the images does not contain any polyps. For Stage 2

classification, 400 NBI images with 275 non-neoplastic images and 190 WL images, where 125 are neoplastic, and 107 are non-neoplastic, are available.

Database 2 (DB2): DB2 used in this study is a public database and is available in (http://www.depeca.uah.es/colonoscopydataset/). For each of the colonoscopy images, both WL and NBI frames are available. Twenty-one non- neoplastic images and 55 neoplastic images are considered from this dataset.

Database 3 (DB3): DB3 is also a public database available in https://datasets.simula.no/kvasir/. There is a set of 4,000 images in Kvasir Dataset v1, which contains eight different classes and 500 images for each class. For this work, we considered images from 2 classes, one for images that contain polyps and another one for images without polyps.

All the datasets are balanced by applying data-augmentation techniques such as shearing, rotation, skewing, zooming, and inverting to the images. After augmentation, the dataset contains 2400 sample images with 1200 samples from each group for classification.

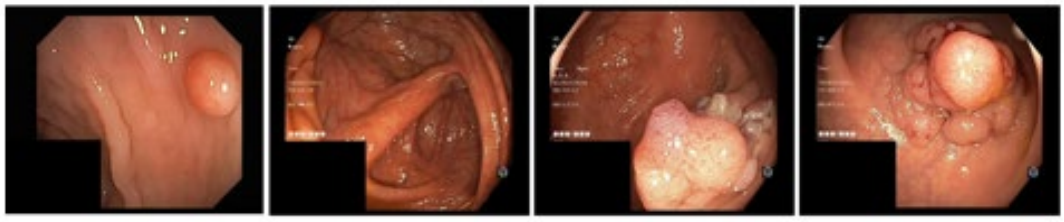
3.2 Pre-Processing

It is observed that unwanted information, such as camera location and text information is present in the extracted frames. Thus, camera location information is removed using a mask, and the resulting images can be seen in Fig. 2b.

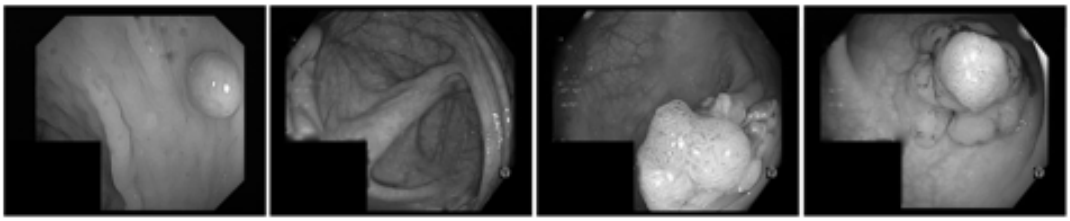
The spectral energy distribution of the reflected light from an object creates a specular effect when a video is captured [24,25]. Specularity causes negative impacts on the computer vision task, such as classification, segmentation, object detection [25], etc. Specularity can be viewed as strongly highlighted pixels in an image. The specular highlighted region may contain important information, e.g., information on shape, color, or texture, which may be significant. So, it is necessary to remove the specularity before feeding the images into the deep learning model.

Existing specularity removal techniques in the literature include threshold-based methods [26] to cluster highlighted and non- highlighted pixels, Support vector machine (SVM) based detection [26], color distribution characteristic based detection methods [27–29], filter-based detection using the information of neighbor pixel [30,31], etc. These methods are time-consuming and inefficient for real-time detection, and applying these methods leads to loss of information behind the highlighted spot and adversely affects the outcome. Principal component analysis (PCA) based techniques are the most promising techniques for removing the specularity [32,33]. Robust PCA (RPCA), a modified PCA based algorithm, is used in this work as a preprocessing technique for removing specularity. Fig. 2c shows the resulted image after applying RPCA. RPCA considers the pixels highlighted to be noise. RPCA decomposes the image into a low-rank and sparse matrix. The hidden information represented by the low-rank matrix is recovered by removing the specularly reflected pixels represented by the sparse matrix.

(a) Input Frames extracted from the video for pre-processing task



(b) Images after applying the mask



(c) Images after removal of specularity by applying Robust PCA

Figure 2: Results of preprocessing

3.3 Training CNN for Classification

CNN is a deep learning approach that can learn the features itself, and the user need not worry and spend too much time selecting important features. CNN can process 2D and 3D images, which is an advantage in real-time utilization of CNN as computer vision techniques for medical image processing [34]. CNN's comprise a profound feed-forward family where intermediate layers receive the features extracted by the previous layer as input and pass their results into the subsequent layers. This network is firmly rooted in learning hierarchical layers of conceptual representation that lead to different abstract levels. For an image, the lower layers of the network model represent the various points and edges on the image; the mid-layers define parts of an object while the higher layers relate to larger parts of the object and sometimes even to the object. Various deep learning models used in this study are VGG16 [35], VGG19 [35], GoogLeNet [36], ResNet50 [37], ResNet100 [37], Inception V3 [38], InceptionResNetV2 [37], MobilenetV2 [39], Xception [40], DenseNet [41], and NASNetMobile [42].

The models are fine-tuned by flattening the last convolutional layer to convert the extracted 2D feature matrix into a 1D vector. The original softmax layer of these models was replaced with a new 3-layer fully connected neural network in the fine-tuned model to differentiate the two classes. These three layers consist of two hidden layers (256 neurons→128 neurons) with the ReLu activation function and a 2-neuron output layer with softmax activation. Stochastic Gradient Descent (SGD) optimizer optimizes the fine-tuned model.

A dropout rate of 0.5 is used to the input of fully-connected layers to regulate over-fitting. Models are pre-trained on the ImageNet [43] dataset to avail the advantages of transfer learning to overcome the lack of availability of colonoscopy frames. All the extracted frames are re-scaled to 224 × 224 before feeding it to each model for classification.

The dataset is split into training, validation, and test dataset based on the train-test strategy. Two thousand sample images are used for training and validation purposes, and 400 unique images are used to test each of the models. Based on early-stopping criteria, the validation loss and accuracy are examined after each epoch, and training is stopped if the validation loss increases after a specific epoch. This approach helps us to avoid over-fitting.

3.4 Experimental Setup

Experiments are carried out in intel optimized Python 3 with Keras (Intel optimized Transorflow backend) using Google-colab that allows using of Tesla K80 GPU with 12 GB of GDDR5 RAM, Intel Xeon processors with two cores @2.2 GHz, and a total of 13 GB of ram.

3.5 Model Assessment Measure

The CNN models' assessments are performed based on four statistical measures, such as accuracy, precision, recall, and F1-score. The objective is to achieve a promising classification report.

In the present work, to evaluate the first decision, i.e., identification of significant frames, if an image that contains a polyp is classified as significant, then it is true-positive (TP) if an image that does not contain any polyps is classified as insignificant then it is true-negative (TN) and if the classification result for an image is opposite to the above-mentioned criteria then it false-negative (FN) and false-positive (FP) accordingly.

For the evaluation of the second decision, i.e., detection of neoplastic and non- neoplastic frames, if an image that contains a cancerous polyp is classified as non-neoplastic, then it is true-positive; if an image that does not contain any cancerous polyps is classified as neoplastic, then it is true-negative and if the classification result for an image is opposite to the criteria as mentioned above then it false-negative and false-positive accordingly. The evaluation criteria for the assessment measures used in this work are explained in this subsection.

The accuracy of classification shows the number of correct predictions. Thus, the higher the accuracy means the accurate prediction is; therefore, the classification process is better. Accuracy can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision is the ratio of correctly predicted positive observations of the total predicted positive observations.

$$Precision = \frac{TP}{TP + FP}$$

The recall is the rate of correctly predicted positive observation out of the total actual positive. So, higher recall indicates a low rate of misclassification.

$$Recall = \frac{TP}{TP + FN}$$

When the precision is high and the recall low or vise-versa, the direct comparison between the two models is difficult. Therefore, F1-score is considered to compare two models as it is the harmonic mean of precision and recall.

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Based on these criteria, a comparison is drawn for the best model for colon-cancer analysis.

4 Result and Discussion

4.1 Evaluation of Stage 1 Classification for Significant Frame Detection

In Fig. 4, the graphical representation of 'training loss vs. validation loss' and 'training accuracy *vs.* validation accuracy' for all the CNN architecture used in this paper to detect significant frames are shown. A better model means small loss value and high accuracy. The result of all the assessment measures for the CNN models used for this classification is listed in Tab. 3.

Table 3: Stage 1 classification results for the detection of significant frames

Methods	Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Testing Loss	Testing Accuracy	Precision	Recall	F1-Score
VGG16	50	0.35	84.75	0.24	79.75	0.30	87.50	92.57	81.40	86.63
VGG16 Fine-Tuned	40	0.05	98.44	0.11	95.00	0.11	95.00	97.35	92.46	94.84
VGG19	50	0.09	98.87	0.11	95.00	0.11	95.00	97.35	92.46	94.84
VGG19 Fine-Tuned	50	**0.01**	**99.94**	**0.11**	**95.75**	**0.10**	**95.75**	96.90	94.47	**95.67**
ResNet50	50	0.65	69.38	0.44	80.25	0.44	80.25	76.54	81.41	78.90
ResNet50 Fine-Tuned	50	0.18	91.87	0.28	87.75	0.28	87.75	94.11	80.40	86.72
ResNet101	15	0.11	96.06	0.23	91.00	0.23	91.00	97.10	84.42	90.32
ResNet101 Fine-Tuned	15	0.09	96.94	0.16	93.00	0.16	93.00	94.76	90.95	92.82
GoogLeNet	50	0.45	89.94	0.38	83.59	0.35	83.50	78.96	91.54	84.79
GoogLeNet Fine-Tuned	50	0.22	94.56	0.32	89.32	0.29	89.50	89.16	90.00	89.95
InceptionV3	50	0.44	79.87	0.47	78.75	0.47	78.75	83.92	70.85	76.83
InceptionV3 Fine-Tuned	50	0.31	85.75	0.34	84.75	0.34	84.75	84.84	84.42	84.63
DenseNet121	20	0.19	93.25	0.24	90.00	0.24	90.00	88.40	91.95	90.14
DenseNet121 Fine-Tuned	20	0.08	97.12	0.17	95.00	0.17	95.00	98.90	90.95	94.77
Xception	50	0.62	68.00	0.63	64.00	0.63	64.00	70.37	47.73	56.88
Xception Fine-Tuned	50	0.44	77.19	0.34	74.25	0.34	74.25	75.26	71.18	73.52
NASNetMobile	50	0.55	70.31	0.56	68.50	0.56	68.50	65.02	79.39	71.39
NASNetMobile Fine-Tuned	50	0.50	73.50	0.87	72.00	0.87	72.00	90.17	50.75	64.95
MobileNetV2	50	0.98	54.37	0.71	59.00	0.71	59.00	100.00	17.58	29.91
MobileNetV2 Fine-Tuned	49	0.52	72.94	0.87	73.25	0.87	75.00	96.26	51.75	67.32
InceptionResNetV2	35	0.47	74.75	0.51	71.75	0.51	71.50	71.50	71.85	71.67
InceptionResNetV2 Fine-Tuned	50	0.44	79.87	0.47	78.75	0.47	78.75	88.75	87.53	88.09

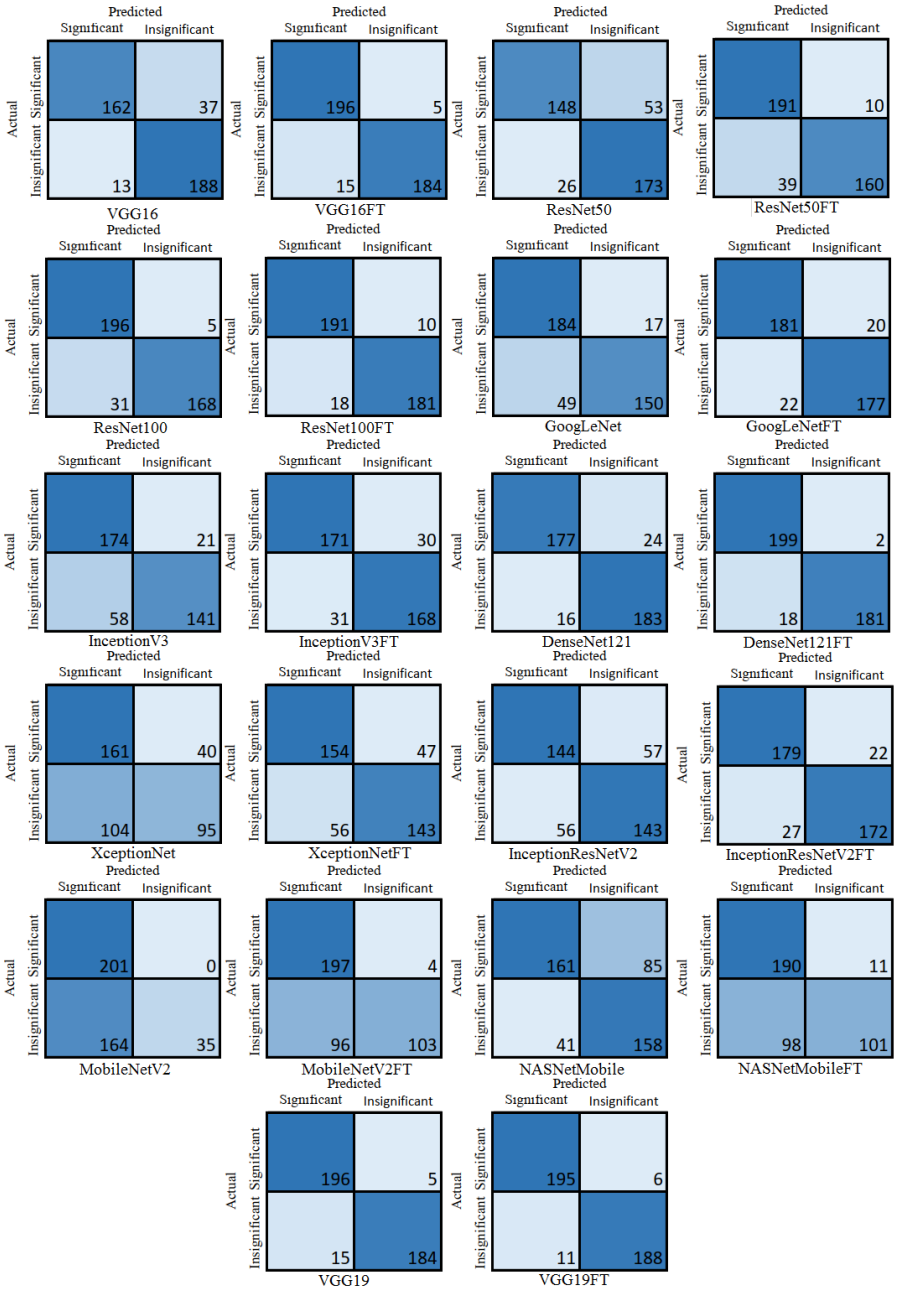


Figure 3: The confusion matrix for the task of classifying significant and insignificant frames

Fig. 3 shows the confusion matrix of significant and insignificant frame classification tasks. Comparing all the models present in this work based on results shown in Tab. 3 and Fig. 3, and Fig. 4, the fine-tuned VGG19 models display a highly satisfactory performance with a drop in loss rate and an increase in accuracy at each stage. It obtained a 95.75% test accuracy, F1-score of 95.67% with a precision of 96.90% and 94.47% recall value, and a 0.11 lowest validity loss in 50 epoch. The good performance of VGG 19 can be clearly observed from the graphs in Fig. 4.

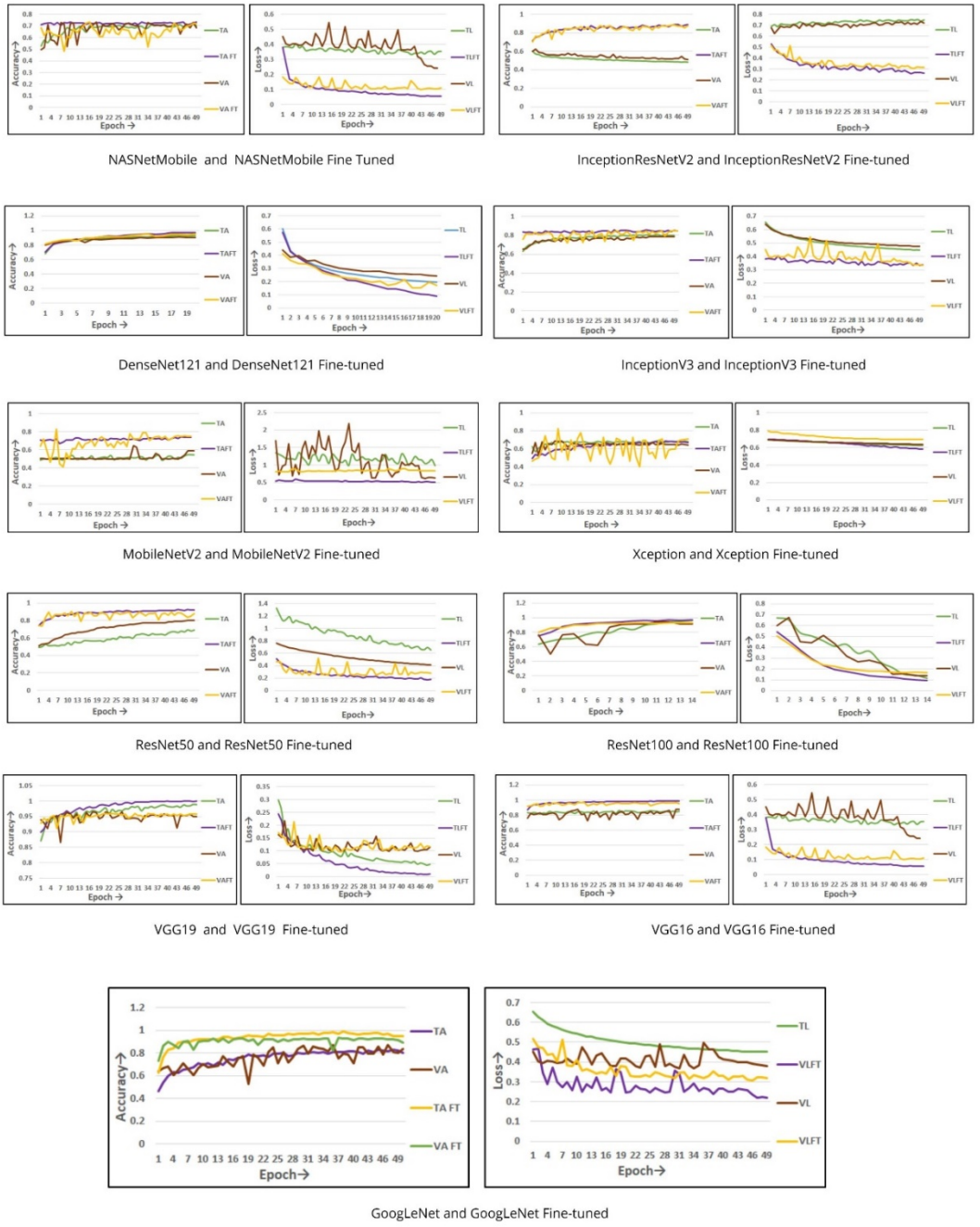


Figure 4: 'Training-accuracy *vs.* validation-accuracy' and 'training-loss *vs.* validation-loss' for the task of classifying significant and insignificant frames

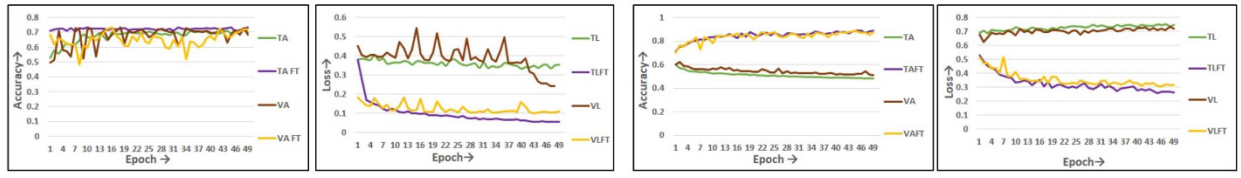
4.2 Evaluation of Stage 2 Classification for Categorization of Neoplastic and Non-Neoplastic Polyps

After successful identification of the significant frames, doctors need to check for the presence of an abnormality in the polyp. The result evaluated based on model assessment measures for each model is shown in Tab. 4.

Table 4: Stage 2 classification results to distinguish between neoplastic and non-neoplastic frames

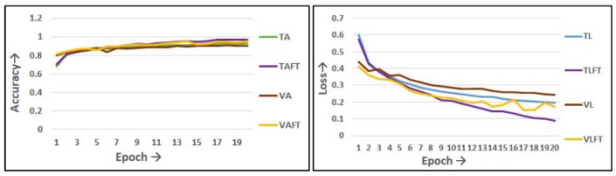
Methods	Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Testing Loss	Testing Accuracy	Precision	Recall	F1-Score
VGG16	50	0.35	84.75	0.24	79.75	0.30	87.50	92.57	81.40	86.63
VGG16 Fine-Tuned	40	0.05	98.44	0.11	95.00	0.11	95.00	97.35	92.46	94.84
VGG19	50	0.09	98.87	0.11	95.00	0.11	95.00	97.35	92.46	94.84
VGG19 Fine-Tuned	50	0.01	99.94	0.11	95.75	0.10	95.75	96.90	94.47	95.67
ResNet50	50	0.65	69.38	0.44	80.25	0.44	80.25	76.54	81.41	78.90
ResNet50 Fine-Tuned	50	0.18	91.87	0.28	87.75	0.28	87.75	94.11	80.40	86.72
ResNet101	15	0.11	96.06	0.23	91.00	0.23	91.00	97.10	84.42	90.32
ResNet101 Fine-Tuned	15	0.09	96.94	0.16	93.00	0.16	93.00	94.76	90.95	92.82
GoogLeNet	50	0.45	89.94	0.38	83.59	0.35	83.50	78.96	91.54	84.79
GoogLeNet Fine-Tuned	50	0.22	94.56	0.32	89.32	0.29	89.50	89.16	90.00	89.95
InceptionV3	50	0.44	79.87	0.47	78.75	0.47	78.75	83.92	70.85	76.83
InceptionV3 Fine-Tuned	50	0.31	85.75	0.34	84.75	0.34	84.75	84.84	84.42	84.63
DenseNet121	20	0.19	93.25	0.24	90.00	0.24	90.00	88.40	91.95	90.14
DenseNet121 Fine-Tuned	20	0.08	97.12	0.17	95.00	0.17	95.00	98.90	90.95	94.77
Xception	50	0.62	68.00	0.63	64.00	0.63	64.00	70.37	47.73	56.88
Xception Fine-Tuned	50	0.44	77.19	0.34	74.25	0.34	74.25	75.26	71.18	73.52
NASNetMobile	50	0.55	70.31	0.56	68.50	0.56	68.50	65.02	79.39	71.39
NASNetMobile Fine-Tuned	50	0.50	73.50	0.87	72.00	0.87	72.00	90.17	50.75	64.95
MobileNetV2	50	0.98	54.37	0.71	59.00	0.71	59.00	100.00	17.58	29.91
MobileNetV2 Fine-Tuned	49	0.52	72.94	0.87	73.25	0.87	75.00	96.26	51.75	67.32
InceptionResNetV2	35	0.47	74.75	0.51	71.75	0.51	71.50	71.50	71.85	71.67
InceptionResNetV2 Fine-Tuned	50	0.44	79.87	0.47	78.75	0.47	78.75	88.75	87.53	88.09

Fig. 5 shows the resulted graph for 'training accuracy vs. validation accuracy' and 'training loss *vs.* validation loss', and the confusion matrix is shown in Fig. 6 for all the CNN architecture used in this paper to classify neoplastic and non-neoplastic frames. For Stage 2 classification also, the fine-tuned version of the VGG19 model works best among all the models considered for this study.

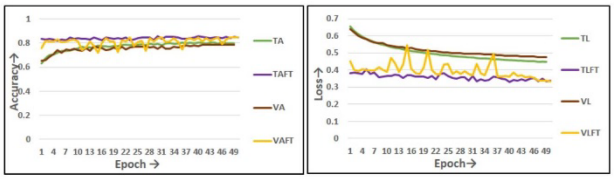


NASNetMobile and NASNetMobile Fine Tuned

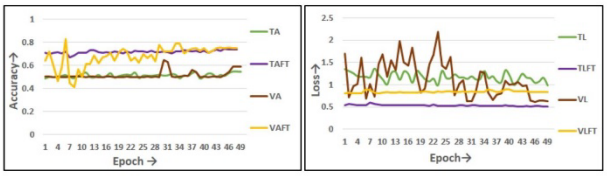
InceptionResNetV2 and InceptionResNetV2 Fine-tuned



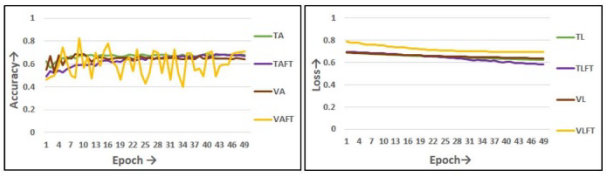
DenseNet121 and DenseNet121 Fine-tuned



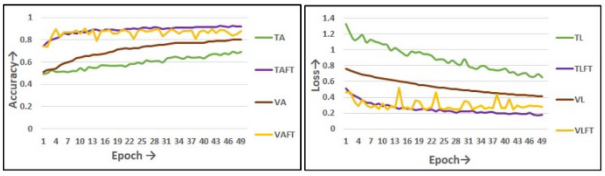
InceptionV3 and InceptionV3 Fine-tuned



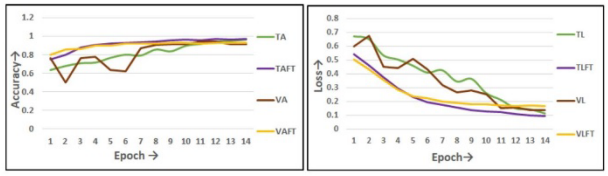
MobileNetV2 and MobileNetV2 Fine-tuned



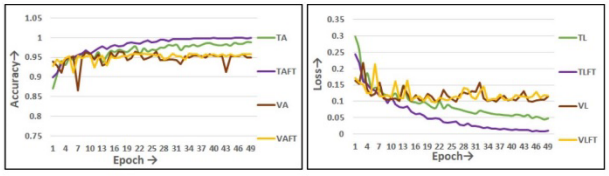
Xception and Xception Fine-tuned



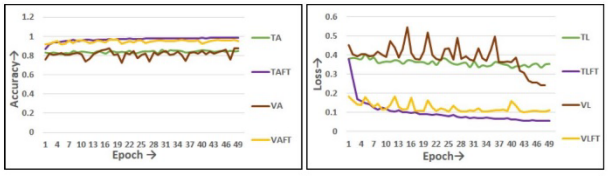
ResNet50 and ResNet50 Fine-tuned



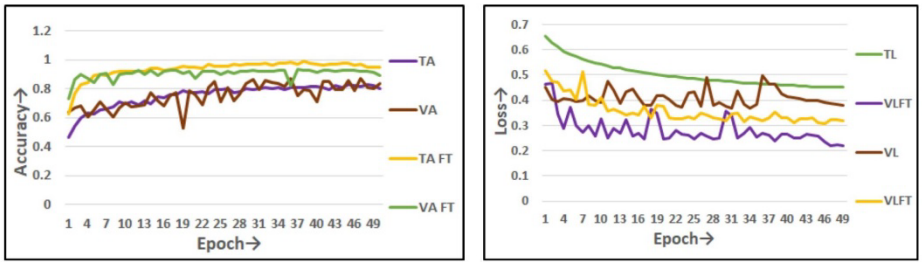
ResNet100 and ResNet100 Fine-tuned



VGG19 and VGG19 Fine-tuned



VGG16 and VGG16 Fine-tuned



GoogLeNet and GoogLeNet Fine-tuned

Figure 5: Training accuracy *vs.* validation accuracy and training loss *vs.* validation loss for the task of Neoplastic and non-neoplastic polyps detection

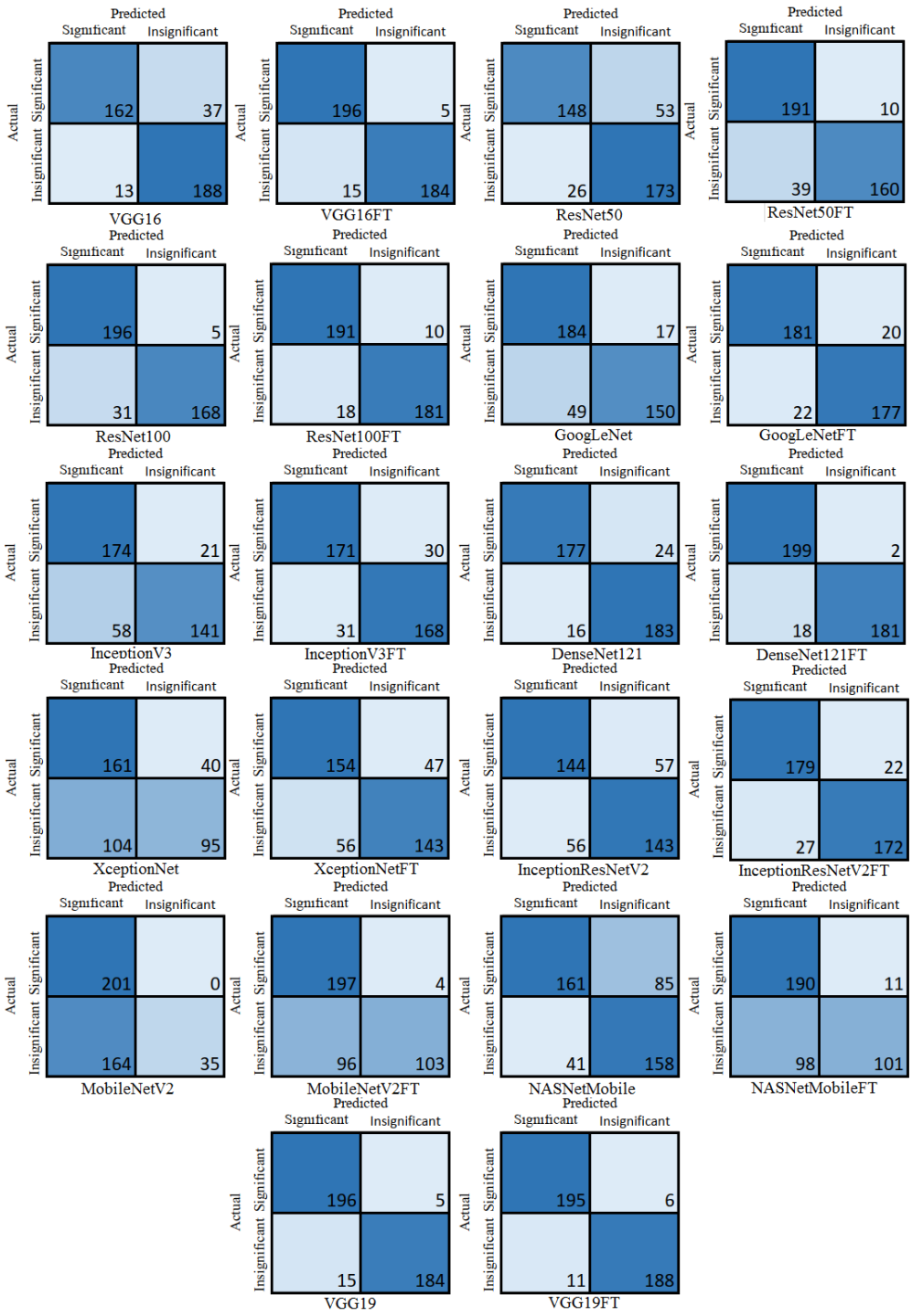


Figure 6: The confusion matrix for the classification of neoplastic and non-neoplastic polyps

Tab. 5 shows the classification results on different databases. It is clearly seen that performance is consistent across all the databases establishing the claim made. The fine-tuned version of the CNN model VGG19 can be effectively used in polyp classification compared to all the methods mentioned in Tab. 6.

Table 5: Classification result in different databases

Dataset used	Methods	Loss	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
DB1	VGG16 Fine Tuned	0.11	95.00	97.35	92.46	94.84
DB1	VGG19 Fine Tuned	0.10	95.74	96.90	94.47	95.67
DB2	VGG16 Fine Tuned	0.11	95.00	96.64	93.59	95.09
DB2	VGG19 Fine Tuned	0.10	95.75	96.03	95.56	95.80
DB3	VGG16 Fine Tuned	0.11	95.00	97.35	92.46	94.84
DB3	VGG19 Fine Tuned	0.10	95.75	96.90	94.47	95.67

4.3 Comparison with Existing Literature

The proposed work is compared with four existing literature's [4, 5, 6, 8, 44]. The database DB1 is used for comparison purposes. Tab. 6 displays the result of all the methods for Stage 1.

Table 6: Comparison with existing literature

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG16 Fine Tuned	95.00	**97.35**	92.46	94.84
VGG19 Fine Tuned	**95.75**	96.90	**94.47**	**95.67**
Akbari et al. [4]	90.28	74.34	68.32	71.20
Wang et al. [5]	90.50	-	94.38	-
NPI-CNN1 [6]	91.90	-	88.1	-
Patel et al. [8]	79.78	83.35	-	83.52
Shin et al. [44]	91.26	92.71	90.82	-

It is observed that the fine-tuned version of VGG19 outperforms all the methods. It proves that the proposed work is successful in establishing the claim made. The fine-tuned version of the CNN model VGG19 can be effectively used in polyp classification compared to all the methods mentioned in Tab. 6.

5 Conclusion

In this paper, a two-stage classification is presented to detect colorectal cancer from colonoscopy videos. In the first stage, frames of colonoscopy video are extracted and are rated as significant if it contains a polyp, and these results are then aggregated in a second stage to come to an overall decision concerning the final classification of that frame to be neoplastic and non-neoplastic. We investigated the applicability of deep learning to perform this two-stage classification and the CNN models, namely VGG16, VGG19, Inception V3, Xception, GoogLeNet, ResNet50, ResNet100, DenseNet, NASNetMobile, MobilenetV2, InceptionResNetV2 and fine-tuned version of each model is evaluated. It is observed that the two fine-tuned version of four models: VGG16, VGG19, ResNet100, and DenseNet121, achieved more than 90% of accuracy in both the stages, and the best result was achieved by fine-tuned VGG19 with a test accuracy of 95.75%. Transfer learning from the ImageNet dataset is one of the reasons that VGG19 outperforms some of the previous results mentioned in the literature where training CNN is done on raw data. Thus, we can expect performance gain if the transfer learning is made from the same domain dataset. After the categorization of neo-plastic and non-neoplastic polyps, the automated system will be more useful to the doctors if it can provide a precise 3D location. We believe that this work helps the research community in gaining acquaintance with the automatic detection process of colorectal cancer.

Acknowledgment: Authors of this article would like to thank Professor Manas Kamal Bhuyan, department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, for their help in preparaing the generated database (DB1) used in this study.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. WHO. *Cancer*. (2018). https://www.who.int/news-room/fact-sheets/detail/cancer.
2. Sonnenberg, A., Delcò, F. and Inadomi, J. M. (2000). Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Annals of Internal Medicine, 133(8),* 573–584.
3. Eaden, J. A., Ward, B. A., Mayberry, J. F. (2000). How gastroenterologists screen for colonic cancer in ulcerative colitis: an analysis of performance. *Gastrointestinal Endoscopy, 51(2),* 123–128.
4. Akbari, M., Mohrekesh, M., Rafiei, S., Soroushmehr, S. R., Karimi, N. et al. (2018) . Classification of informative frames in colonoscopy videos using convolutional neural networks with binarized weights. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 65–68.
5. Wang, P., Xiao, X., Brown, J. R. G., Berzin, T. M., Tu, M. et al. (2018). Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nature Biomedical Engineering, 2(10),* 741–748.
6. Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F. et al. (2018). Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology, 155(4),* 1069–1078.
7. Sundaram, P. S., Santhiyakumari, N. (2019). An enhancement of computer aided approach for colon cancer detection in WCE images using ROI based color histogram and SVM2. *Journal of Medical Systems, 43(2),* 29.
8. Patel, K., Li, K., Tao, K., Wang, Q., Bansal, A. et al. (2020) A comparative study on polyp classification using convolutional neural networks. *PLoS One, 15(7),* e0236452.
9. Hasan, M. M., Islam, N., Rahman, M. M. (2020). Gastrointestinal polyp detection through a fusion of contourlet transform and Neural features. *Journal of King Saud University–Computer and Information Sciences*. DOI 10.1016/j.jksuci.2019.12.013.
10. Patino-Barrientos, S., Sierra-Sosa, D., Garcia-Zapirain, B., Castillo-Olea, C., Elmaghraby, A. (2020). Kudo's classification for colon polyps assessment using a deep learning approach. *Applied Sciences, 10(2),* 501.
11. Ghesu, F. C., Georgescu, B., Zheng, Y., Grbic, S., Maier, A. et al. (2017). Multiscale deep reinforcement learning for real-time 3D-landmark detection in CT scans. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1),* 176–189.
12. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E. et al. (2019). Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Medicine, 16(1),* e1002730.
13. Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P. A. et al. (2019). MILD-net: minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis, 52,* 199–211.
14. Li, C., Wang, X., Liu, W., Latecki, L. J., Wang, B. et al. (2019). Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Medical Image Analysis, 53,* 165–178.
15. Saikia, A. R., Bora, K., Mahanta, L. B., Das, A. K. (2019). Comparative assessment of CNN architectures for classification of breast FNAC images. *Tissue and Cell, 57,* 8–14.
16. Qi, X., Zhang, L., Chen, Y., Pi, Y., Chen, Y. et al. (2019). Automated diagnosis of breast ultrasonography images using deep neural networks. *Medical Image Analysis, 52,* 185–198.
17. He, T., Puppala, M., Ezeana, C. F., Huang, Y. S., Chou P. H. et al. (2019). A deep learning–based decision support tool for precision risk assessment of breast cancer. *JCO Clinical Cancer Informatics, 3(2019),* 1–12.
18. Sun, W., Tseng, T. L. B., Zhang, J., Qian, W. (2017). Enhancing deep convolutional neural network scheme for

breast cancer diagnosis with unlabeled data. *Computerized Medical Imaging and Graphics, 57,* 4–9.

19. Książek, W., Abdar, M., Acharya, U. R., Pławiak, P. (2019). A novel machine learning approach for early detection of hepatocellular carcinoma patients. *Cognitive Systems Research, 54,* 116–127.
20. Jiang, J., Hu, Y. C., Liu, C. J., Halpenny, D., Hellmann, M. D. et al. (2018). Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. *IEEE Transactions on Medical Imaging, 38(1),* 134–144.
21. Zhang, S., Han, F., Liang, Z., Tan, J., Cao, W. et al. (2019). An investigation of CNN models for differentiating malignant from benign lesions using small pathologically proven datasets. *Computerized Medical Imaging and Graphics, 77,* 101645.
22. Nasr-Esfahani, E., Rafiei, S., Jafari, M. H., Karimi, N., Wrobel, J. S. et al. (2019). Dense pooling layers in fully convolutional network for skin lesion segmentation. *Computerized Medical Imaging and Graphics, 78,* 101658.
23. Khan, M., Khan, M. A., Ahmed, F., Mittal, M., Goyal, L. M. et al. (2019). Gastrointestinal diseases segmentation and classification based on duo-deep architectures. pattern recognition letters. https://doi.org/10.1016/j.patrec.2019.12.024.
24. Xia, W., Chen, E. C., Pautler, S. E. and Peters, T. M. (2019). A global optimization method for specular highlight removal from a single image. *IEEE Access, 7,* 125976–125990.
25. Yang, Q., Tang, J., Ahuja, N. (2014). Efficient and robust specular highlight removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(6),* 1304–1311.
26. Yu, D., Han, J., Jin, X., Han, J. (2014). Efficient highlight removal of metal surfaces. *Signal Processing, 103,* 367–379.
27. Morgand, A., Tamaazousti, M. (2014). Generic and real-time detection of specular reflections in images. *2014 International Conference on Computer Vision Theory and Applications, 1,* 274–282.
28. Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J. et al. (2007). Informative frame classification for endoscopy video. *Medical Image Analysis, 11(2),* 110–127.
29. Gao, Y., Yang, J., Ma, S., Ai, D., Lin, T. et al. (2017). Dynamic searching and classification for highlight removal on endoscopic image. *Procedia Computer Science, 107,* 762–767.
30. Yang, Q., Tang, J., Ahuja, N. (2014). Efficient and robust specular highlight removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(6),* 1304–1311.
31. Yang, Q., Wang, S., Ahuja, N. (2010). Real-time specular highlight removal using bilateral filtering. *European Conference on Computer Vision,* 87–100. Springer, Berlin, Heidelberg.
32. da Silva Queiroz, F., Ren, T. I. (2014). Automatic segmentation of specular reflections for endoscopic images based on sparse and low-rank decomposition. *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images,* 282–289.
33. Tan, K., Wang, B., Gao, Y. (2017). Automatic specular reflections removal for endoscopic images. *Ninth International Conference on Digital Image Processing*.
34. Ker, J., Wang, L., Rao, J., Lim, T. (2017). Deep learning applications in medical image analysis. *IEEE Access, 6,* 9375–9389.
35. Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
36. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. et al. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 1–9.
37. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 770–778.
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2818–2826.
39. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W. et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
40. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 1251–1258.
41. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks.

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708.

42. Zoph, B., Vasudevan, V., Shlens, J., Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition,* 8697–8710.
43. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. et al. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition,* 248–255.
44. Shin, Y., Balasingham, I. (2017). Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 3277–3280.