Multi-Class Classification on Capsule Endoscopy Dataset Using an Ensemble of Machine Learning and Deep Learning

Madhura Prakash M[1], Dr. Krishnamurthy G N[2]
[1]Research Scholar, Dept. of CSE, BNM Institute of Technology, Bangalore
[2]Principal, BNM Institute of Technology, Bangalore
(E-mail: madhuraprakash18@gmail.com)

Abstract**—** Endoscopy is a procedure that is used to examine the inner parts of the human organs. The examination is assessed via a video output produced from the procedure. A special type of endoscopy procedure is a wireless capsule endoscopy (WCE) procedure in which the patient swallows a pill-based camera. This pill traverses through the digestive system and captures all the details. The average duration of the pill traversal in the entire system is about 7 to 8 hours and this results in about 80,000 frames. The physician will have to monitor these frames looking for any abnormalities or any lesions. Computer vision algorithms help in processing the frames and extracting out the relevant portions of the frame. The artificial intelligent solutions backed by computer vision algorithms can be used to develop solutions to automate the process of extracting relevant frames of interest and also to classify these frames to denote the probable class of abnormality occurrence or to classify them as normal. In this work, publicly available curated WCE colon disease dataset has been used to perform multi-class classification. This work progresses in two stages. The first stage is based on computer vision techniques for feature extraction followed by classification using machine learning algorithms resulting in binary classification of normal versus abnormal endoscopy frames. The second stage is based on the deep learning models to perform a multi-class classification. The proposed system has successfully classified WCE frames as normal versus abnormal with an accuracy of 92% using merged feature vectors from the feature extraction algorithms applied to classifier models of SVM, Random Forest, and Logistic Regression. The proposed methodology has obtained an accuracy of 94.26% using a standard deep convolutional neural network architecture namley SqueezeNet for the multi-class classification. The work analyzes the performance of classification models on the indiviual and merged feature vectors. The work also compares the performance of customary deep convolution neural networks for multi-class classification on capsule endoscopy frames.

Keywords— *Capsule endoscopy, SIFT, GLCM, Classifier models, Convolution neural network*

I. INTRODUCTION

The advantages of non-invasive or minimally invasive procedures are gaining importance in the medical diagnosis and therapy because of the various advantages it provides like less preparation, no anaesthesia, quick recovery. Endoscopy is a procedure that is used in both diagnosis and in therapy in the various organs of the human body ranging from abdomen, joints, lungs and so on. A special type of endoscopy procedure that is used in the diagnosis of the entire digestive system is the wireless capsule endoscopy (WCE). In this procedure the patient swallows a camera-based capsule that traverses the entire path covering the large and small intestine along with the colon region. The capsule constantly captures the video frames that are transmitted to the physician. The whole process takes about 7 to 8 hours to complete and produces an average number of 80,000 frames. The physician examines these frames looking for any abnormalities like bleeding, polyps, tumours or any other lesions. Artificial intelligent based solutions backed by deep learning techniques can aid the doctors by helping out to pick frames of relevant interest so as to keep the doctor's focus on critical regions of interest. [1]

In this work the widely available curated WCE colon disease dataset from public repository has been considered to perform a multi-class classification on the WCE images. The work progresses in two stages. In the first stage feature extraction algorithms namely scale-invariant feature transform (SIFT) and gray -level co-occurrence matrix (GLCM) are used to extract relevant features from the images. The standard classifier models like Support Vector Machine (SVM), Random Forest (RF) and Logistic Regression are run on the extracted features of SIFT and GLCM and the combination of these 2 feature vectors. An analysis of the performance of these classifier models on the individual feature vectors of SIFT and GLCM independently and in their combination has been done in the phase one. The result of phase one is a binary classification indicating a normal WCE frame or an abnormal WCE frame. In the phase two, the abnormal WCE frame are further classified into specific category belonging to either of ulcerative colitis, polyps or esophagitis. For this multiclass classification a region of interest in the frame is computed using computer vision technique and further standard deep learning architecture like SqueezeNet, ResNet, MobileNet and VGG16 have been used via the transfer learning of model weights procedure to train these models on the considered data and analyse their classification performance.

This work is this a combination of traditional image processing techniques for feature extraction followed by standard machine learning based classifier models and computer-vision based techniques followed by the deep learning model-based architecture. The work also presents a detailed analysis of these feature extraction techniques, the machine learning classification algorithms, and the deep learning architecture for their performance individually and in combination for the considered dataset. The analysis of these results has been presented in detail in the further sections.

II. EXPERIMENTAL METHODS

A. Materials

The WCE image dataset representing the colon disease is curated and is publicly available [2,3,4]. The dataset is a balanced version for all representative images. The dataset has been divided into training and testing folders. Each of the normal, ulcerative colitis, polyps and esophagitis have 800 images in the training folder and 200 each in the testing folder. The dataset mainly consists of the capsule endoscopy frames belonging to the classes depicting colon disease. The frames are collection from the KVASIR publicly available dataset.

FIGURE 1. SAMPLE IMAGES FROM THE WCE COLON DISEASE CURATED DISEASE DATASET

B. Methodology

The goals of the proposed methodology are twofold: (i) classify any given WCE image into normal or abnormal (ii) classify the colon disease present in the image if it is classified as abnormal in the first step. The first includes the extraction of SIFT and GLCM features from the images in the dataset. A brief description of these two feature extraction techniques are as follows:

B.1 Gray Level Co-occurrence Matrix (GLCM)

GLCM is a feature extraction method that determines the texture of the image by estimating the spatial relationship between the pixels. The texture of the given image is computed in GLCM by assessing the frequency of pairs of pixels with specific values. The corelation between the pairs of pixels occurance is analyzed and statistical measures from the resultant matrix is extracted to compute GLCM [5]. The GLCM can be assessed statiscally for extracting the tecture information from an image. The statiscal measures on GLCM values that are computed by this algorithm are: 1. Contrast – that measures the local changes corrsponding to the intensity contrast between a pixel and its neighbor, computed over the entire image. 2. Correlation – computaion of linear dependency of pairs of pixels. 3. Energy- Angular's moment that provides the sum of squares in GLCM is computed. 4. Homogeneity – Estimation of homogenous pixel values that helps in determining the texture feature of the image.

B.2 Scale Invariant Feature Transform (SIFT)

SIFT is a computer vision algorithm for feature extraction. This algorithm extracts feature points on the given image in such a way that the algorithm is invariant to the image’s scaling and rotation which generally occurs as a part of the data augmentation technique. The algorithm proceeds in 4 stages to compute the feature points [6]:

i. Detection of peak points in the scale space

Any considered image in the real-world is subjected to a multi-scale nature. A paticular image can be clearly visible and assessed at a particular scale. This step of the algorithm computes the different scale points of the image by using a Gaussian blurring on the image. The scale space of an image is computed using a function $L(x,y,\sigma)$. This is done by applying convolution of a Gaussian kernel (Blurring) at different scales with the input image. The scale space is defined by the function given in equation 1:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

Here * is the convolution operator, G is a Gaussian function and I is an input image.

There are multiple techniques for determining steady key points in the scale space. One of these approaches is Difference of Gaussian, which derives the scale-space peak pints D (x, y, σ) by comparing the two images, one of which has a scale k times that of the other. Equation 2 gives the following expression for D (x, y, σ):

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y) \quad (2)$$

Each point should be compared to its eight counterparts on the identical scale, as addition as its nine neighbors on the above and below scales, in effort to determine the local maxima and minima. The present value is the local minima if it has the smallest value among all of the adjacent points, or the local maxima if it has the highest value.

ii. Keypoint Localization:

The primary purpose of this stage is to eradicate unnecessary keypoints from the sample by identifying those that have inferior presentation around edges or less contrast. Equation 3 expresses the peak point's location, z, as follows:

$$Z = 〚(\partial^2 D)/(\partial x^2)〛^{(-1)} (\partial D)/\partial x \quad (3)$$

If the resulting value at z is less than a threshold value then this point is not considered. This eliminates the extreme values having low contrast. To remove extreme values based on poor localization, the difference of Gaussian function is used.

iii. Assignment of Orientation:

The principal objective of this stage is to emphasize on delivering the keypoints a consistent orientation based on the local features of the image. In order to achieve invariance to image rotation, the keypoint identifier can then be represented in proportion to this orientation.

The key points scale is used to select the Gaussian smoothed image. Orientation histogram is structured from the gradient orientations of represntattive key points. The highest peak in the histogram is computed. Using a combination of this peak along with another local peak of height within 0.8 times of this peak, a keypoint of that direction is created. Few of the key points may have multiple orientation value. To resolve this, a parabola is fitted to the 3 histogram values closest to each peak to interpolate the position of the peak.

iv. Keypoint Descriptor:

Keypoint descriptors can also be created using local gradient data. The gradient information is turned around to have the same direction as that of the keypoint and then weighted by a Gaussian with a variance value of 1.5 times the keypoint scale. Based on the gathered data over a frame centered on the keypoint, a collection of histograms are produced. Keypoint descriptors typically employ a collection of 16 histograms that are arranged in a 4 by 4 grid with 8 orientation bins on each grid, one for each of the main compass directions and another for each of their midpoints. This facilitates the creation of a feature vector with 128 elements.

The features obtained from the two feature extraction techniques namely GLCM and SIFT are combined together to form the resultant vector. This vector is considered as an input to the machine learning based classifier models in the training and testing phases. The features computed are used as input to the classifier models individually and in combination to assess the performance of the classifier models. The curated dataset considered is divided into training and testing portions in the ratio of 80:20 as 80%. Further, a set of machine learning based classifiers are implemented on the extracted GLCM and SIFT features. The classifier models that are used to obtain the stage 1 binary classification results and for the comparative analysis are as follows:

B.3 Support vector Machine (SVM)

Support vector machines is a supervised machine learning algorithm that is applicable to both classification and regression problems. The SVM computes a hyperplane that divides the data points considered. This is achieved iteratively in two phases by the algorithm.

- The algorithm continuously constructs hyperplanes that identify the classes amongst the datapoints.
- The hyperplane that pertinently separates the classes between the datapoints with the least number outliers is selected.

A kernel employed by SVM translates an input data vector into a higher 22-dimensional vector. There are two varieties of kernels: i. Linear Kernel: Used when data can be separated into classes using straight lines, or when the data is linearly separable. ii. Radial Basis Function (RBF) Kernel: This method converts the input space into higher-dimensional spaces, such as quadratic, cubic, etc. The SVM algorithm's gamma values can be manually adjusted. A gamma value of 0.1 is generally preferred [7].

B.4 Logistic Regression (LR):

The logistic regression algorithm can be used for both binary and multi-class classification tasks. The logistic regression algorithm receives as input the feature vectors. The probabilities of occurrence of the data points are computed either using the sigmoid function or the softmax function. The sigmoid value computation takes place in the binary classification problems and the softmax value computation takes place in the multi-classification problems. These computed probability values serve as decision points for assessing the class of each of the data point. The class with a high probability value of occurrence of a datapoint is considered as the final target class [8].

B.5 Random Forest (RF)

Random forest is a supervised machine learning algorithm that is applicable for both classification and regression problems. It is an ensemble of several decision trees and it is mostly used for classification problems.

The random forest algorithm results in a numerous set of data decision trees and then gets the predictions from all of these trees. The final classification result from the random forest algorithm is chosen by a majority voting technique. The major advantage of the Random Forest is that it helps in reducing the problem of over-fitting, which is one of the lacunas in the decision tree algorithm-based models. [9].

The algorithm starts with an initial selection of random samples from the dataset. A decision tree is constructed for each sample set and the classification result from each decision tree is computed. A voting is computed for every classification result. The class with the majority voting that has resulted from all of the intermediary decision tree is declared as the final class.

B.6 K-Nearest Neighbor (KNN)

KNN is yet another supervised machine learning algorithm used for classification purposes. This belongs to a category of lazy learning algorithms in machine learning because the algorithm begins its action only when the test instance is given as input. The feature vector of all the given data points / images are considered to compute the similarity or the distance between the test image and the original images in the database. K is chosen as a random value. The assesment of class of a test instance is done by computing theK nearest neighbouring points for the test image [10]. In the first stage of this algorithm the dataset is divided into training and testing portions. In the second step, the K value is determined as number of neighbors to be chosen for each test instance. In the last stage for each of the data point ‘i’ algorithm carries out the following steps:

- The distance between the test instance and all other instances in the dataset is computed via the Euclidean distance measure
- The distance computed are arranged in ascending order.
- The test instance classification result is based on the classification label of the K nearest neighbour and the maximum occuring class value is chosen.

B.7 AdaBoost

This term stands for Adaptive Boosting which is an ensemble modeling method to build a strong classifier by repeatedly increasing the weights of the incorrect classifications of the weak classifier. This boosting technique was initially developed for binary classification problems and the algorithm follows the following steps [11]:

1. Once we get the dataset, we assign equal weights to each data point.
2. The model is trained on these training examples and misclassified points are identified.
3. Weights of misclassified points are increased.
4. if required results are obtained Go to step 5 else Go to step 2
5. End

B.8 XGBoost classifier algorithm

XGBoost classifier is an extreme gradient boosting algorithm which comes under ensemble-based models. This model works by sequentially building decision trees on the features and optimizing the loss/residual function from the previous decision tree. Model work prpgrsses in 7 steps as follows:

1. Constructing the base model by finding the probability (Pr) and assigning the probabilty of any class.
2. Calculate similarity score using residual score and initialize it at the start of the algorithm.
3. Construct the base decision tree using these similarity score and residual score.
4. Calculate the Information Gain (I.G) of each node
5. Best IG score of the node will be used for the next decision tree root selection.
6. Update log (Pr) = log (Pr/1-Pr) and base model prediction = activation function (log (Pr))

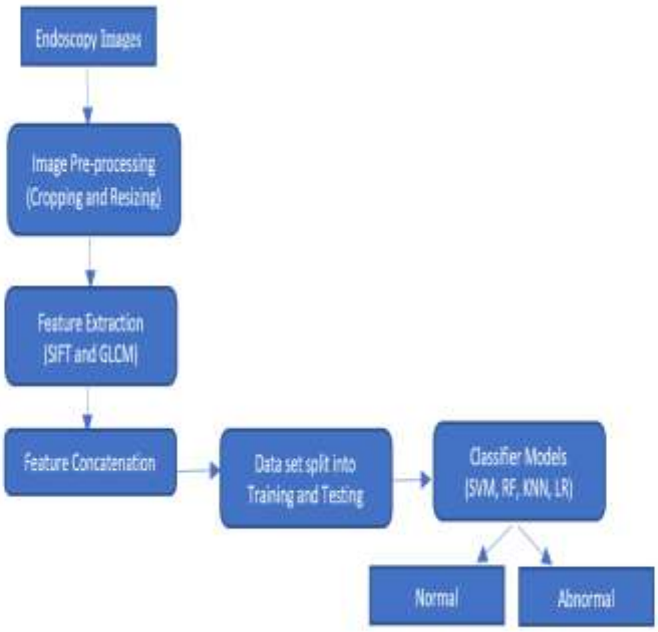

FIGURE 2. FLOW OF THE PHASE 1 DEPICTING THE BINARY CLASSIFICATION PROCESS.

The flow of binary classification is as depicted in the figure 2. The images in the dataset are manually cropped to extract required regions of interest as the dataset. The two algorithms used to extract the texture features are GLCM & SIFT. The feature vectors are then concatenated to form the combined feature vector. The data is divided in the ratio of 80% for training and 20% for testing SVM, Logistic Regression, KNN, Random Forest are used to segregate the result into normal or abnormal WCE frames. The process of multiclass classification is performed using transfer learning and the following pre-trained models have been used to perform multiclass classification on the pre-processed images.

B.9 Mobile Net

This convolution neural network architecture was proposed by Andrew G. Howard [13]. This streamlined architecture was specially designed for mobile and embedded devices with the capability of giving results in a lesser amount of time using less computation power. The architecture follows the principle of depth wise separable convolutions. The main idea was to separate the filter's depth and spatial dimensions which is then followed by point-wise 26 convolution. The overall architecture of MobileNet consists of 28 layers. It can easily be distinguished from a standard CNN since each 3x3 depth separable layer is followed by batch normalization and ReLU activation function layer. This is again repeated for the point-wise separable layer. This model takes an input of size 224x224x3.

B.10. VGG-16

This convolution neural network was initially proposed for the purpose of classification and detection. This model gives an accuracy of 92.7% on the famous imageNet dataset. The motivation behind developing this model was to improve the existing accuracy of the Alexnet model by changing the large kernels into multiple 3x3 kernel-sized filters which are placed in series with each other. The model accepts an input of size 224x224 in RGB format. All the convolution layers have a filter size of 3x3 and all the max pooling layers have pool size of 2x2[14]. In the output layer the model uses softmax activation function. As the name suggests it consists of 16 layers in total and the other variants can have 19 layers with corresponding weights.

B.11 SqueezeNet

This convolution neural network model was specially designed by considering the fact that an indistinguishable level of accuracy can be obtained by reducing the size and number of parameters that the model uses during the training phase. The idea here is to reduce the number of parameters and maintain the same level of accuracy. The Squeezenet model uses 15 times less parameters. and also the size of the model is less than 0.5 MB [15].

Use of such a compressed and squeezed model helps to run the model on various distributed servers and also in the environment where there is a limitation of the memory available. It replaces the conventional 3x3 filters with 1x1 point wise filters which required processing power 9 times lesser than the former, these reduced filters are termed as fire modules. It consists of 18 layers and uses an input size of 227x227 in the input layer.

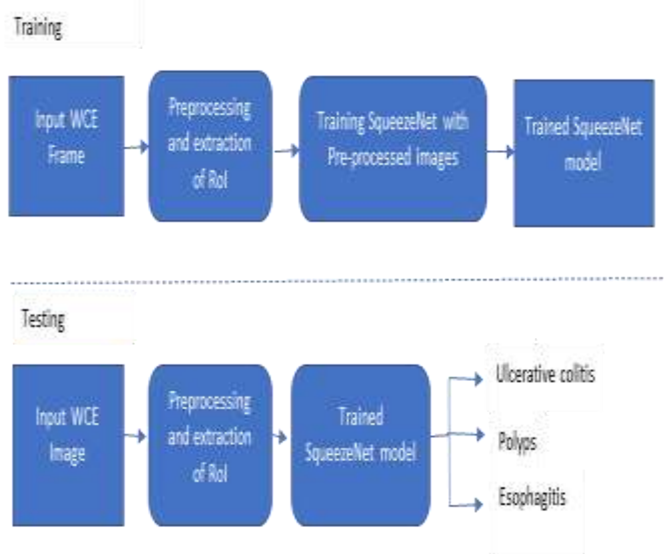

FIGURE 3. HIGH LEVEL ARCHITECTURE OF PHASE 2 DEPICTING THE MULTI-CLASS CLASSIFICATION.

The phase 2 for multi- class classification follows the steps as shown in figure 3. Initially the input images are resized into 224 *224 dimension. Next the images are converted into HSV(Hue-Saturation-Value) format. The Region of Interest (ROI) is computed by applying a mask to the image. This is follwed by image thresholding. A Hough Circle technique is applied in order to get the ROI. The images are thus pre-processed. The pre-trained models namely MobileNet, VGG-16 and SqueezeNet are considered. The models are trained for 30 epochs on the considered data to fine tune the weights as to achieve higher classification accuracy. The performance of all of the classifier models and the deep learning architectures are analyzed and the results are presented in the next section.

III. RESULTS AND DISCUSSION

The accuracy of multiple feature extraction methodologies applied as classification parameters to machine learning models trained with various algorithms, such as SVM-Linear, SVM-Radial basis function, Random Forest, Logistic regression, and KNN, has been attempted to be evaluated in this work. The tables below can be used to compare and contrast the accuracy studies. The accuracy of several classification methods using SIFT features is shown in the table below, with the SVM algorithm and logistic regression having the highest classification accuracy at 92%. An accuracy of 84% was reached with the random forest method. As compared to the other classification models, the KNN algorithm has achieved a least accuracy of 68% for the binary classification of the endoscopy frames into normal and abnormal.

TABLE I. A COMPARTIVE ANALYSIS OF THE PERFORMANCE OF CLASSIFIER MODELS ON EXTRACTED SIFT FEATURES

Classification Algorithm	Accuracy
SVM-LINEAR	92%
SVM-RBF	92%
Random Forest	84%
Logistic Regression	92%
KNN	68%

The performance of the various classification algorithms on the extracted GLCM features are assesed next. The table below describes the accuracy of various classification algorithms with the use of various GLCM features. Logistic Regression algorithm has achieved the highest classification accuracy of 75%. The SVM-Linear kernel algorithm achieved accuracy of 71% while SVM-RBF achieved the lowest accuracy i.e. 46%. Random forest algorithm achieved an accuracy of 42%. The KNN algorithm has achieved an accuracy of 58% for the binary classification of the endoscopy frames into normal and abnormal.

.

TABLE II.A COMPARTIVE ANALYSIS OF THE PERFORMANCE OF CLASSIFIER MODELS ON EXTRACTED GLCM FEATURES

Classification Algorithm	Accuracy
SVM-LINEAR	71%
SVM-RBF	46%
Random Forest	42%
Logistic Regression	75%
KNN	58%

A combination of GLCM and SIFT feature vectors are next applied on various classifier algorithms. From the two feature sets described above that were employed for classification, it is evident that the algorithms performed more effectively when the SIFT features were added. Additionally, the GLCM features have also contributed significantly to accuracy, which introduces the notion of combining the SIFT and GLCM features into a new feature set. In an additional effort to boost accuracy, a combination of the features from SIFT and GLCM are considered, which are 2303 and 5 respectively, to create a final vector set with 2308 features.

TABLE III. A COMPARTIVE ANALYSIS OF THE PERFORMANCE OF CLASSIFIER MODELS WITH MERGED SIFT AND GLCM FEATURES.

Classifier	Accuracy
SVM-LINEAR	92%
SVM-RBF	92%
Random Forest	91%
Logistic Regression	90%
KNN	56%

The accuracy results from several classification algorithms employing the combined feature set of GLCM and SIFT features are shown in the above table. SVM has the best accuracy of 92%. KNN algorithm has a least accuracy of 56% for the binary classification of the endoscopy frames into normal and abnormal.

For multi-class classification for the detection of the type of colon disease abnormality in the WCE images the accuracy of deep learning models such as VGG 16, MobileNet and SqueezeNet are examined and analyzed. The comparative study of the accuracies can be analyzed as given in the table below.

TABLE IV. ACCURACY RATES OF DEEP LEARNING MODELS ON MULTI-CLASS CLASSIFICATION

Sl. No.	Model	Validation Accuracy
1.	VGG 16	96.88
2.	MobileNet	90.62
3.	SqueezeNet	97.66

The table shown above shows the accuracies obtained by applying various Deep Learning Models namely VGG16, MobileNet and SqueezeNet. The highest accuracy obtained is 97.66% by SqueezeNet. The MobileNet model resulted in the lowest accuracy i.e., 90.62% and VGG16 achieved accuracy of 96.88%.

IV. CONCLUSION

In this work an ensemble of machine learning algorithms and deep-learning based solution has been provided for the task of multi-class classification on the WCE image. This method provides a detail analysis and comparison of the performance of individual classifier models like SVM, RF, KNN, LR on the standard image features that have been extracted. The performance all of the classifier models have been analyzed. Further taking into account the current research trends in image classification based on deep learning models, the standard state-of-the art architectures namely, ResNet, VGG16, MobileNet V1 and SqueezeNet have been explored for multi-class classification. The concept of transfer learning has been explored to start the model training on the considered dataset from a relevant reference point. The performance of these deep learning architectures on the multi-class classification technique has also been analyzed and presented. This work signifies that a combination of hand-crafted features and off-the shelf features from deep learning can help in achieving better accuracy and precision in the classification problems. Further, any expanse of initial pre-processing done on the images considered before passing them to training on the deep learning architecture will alleviate the load and help in optimization.

REFERENCES

[1] Iddan, G., Meron, G., Glukhovsky, A., & Swain, P. (2000). Wireless capsule endoscopy. Nature, 405(6785), 417-417.

[2] 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. https://doi.org/10.1145/3083187.3083212.

[3] J. Silva, A. Histace, O. Romain, X. Dray and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer", International Journal of Computer Assisted Radiology and Surgery, vol. 9, no. 2, pp. 283-293, 2013. DOI:10.1007/s11548-013-0926-3.

[4] F. J. P. Montalbo, "Diagnosing Gastrointestinal Diseases from Endoscopy Images through a Multi-Fused CNN with Auxiliary Layers, Alpha Dropouts, and a Fusion Residual Block," Biomedical Signal Processing and Control (BSPC), vol. 76, July, 2022, doi: 10.1016/j.bspc.2022.103683.

[5] "Texture Analysis Using the Gray-Level Co-Occurrence Matrix (GLCM)" - MATLAB & Simulink.MathWorksIndia.[Online].Available:https://in.mathworks.com/help/images/textureanalysis using-the-gray-level-co-occurrence-matrixglcm.html.

[6] Lindeberg, Tony. (2012). Scale Invariant Feature Transform. 10.4249/scholarpedia.10491.

[7] Cristianini, N., Ricci, E. (2008). Support Vector Machines. In: Kao, MY. (eds) Encyclopedia of Algorithms. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30162-4_415

[8] Wright, R. E. (1995). Logistic regression.

[9] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324.

[10] Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). KNN Model-Based Approach in Classification.

[11] Chengsheng, Tu & Huacheng, Liu & Bing, Xu. (2017). AdaBoost typical Algorithm and its application research. MATEC Web of Conferences. 139. 00222. 10.1051/matecconf/201713900222.

[12] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[13] Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

[14] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[15] Iandola, Forrest & Han, Song & Moskewicz, Matthew & Dally, William & Keutzer, Kurt. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size.

[16] "Computer Vision and Pattern Recognition"[Online] Available: https://arxiv.org/abs/1602.07360

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.