Received: 9 January 2023 | Revised: 6 December 2023 | Accepted: 17 December 2023

DOI: 10.1002/ima.23017

RESEARCH ARTICLE

WILEY

A comprehensive study on automatic non-informative frame detection in colonoscopy videos

Rukiye Nur Kaçmaz[1] | **Refika Sultan Doğan**[2] | **Bülent Yılmaz**[3,4]

[1]Software Engineering Department, Erciyes University, Kayseri, Turkey

[2]Bioengineering Department, Abdullah Gül University, Kayseri, Turkey

[3]Electrical Engineering Department, Gulf University for Science and Technology, Mishref, Kuwait

[4]Electrical-Electronics Engineering Department, Abdullah Gül University, Kayseri, Turkey

Correspondence
Rukiye Nur Kacmaz, Software Engineering Department, Erciyes University, Kayseri, Turkey.
Email: rukiyenurkacmaz@gmail.com

Funding information
Turkish Higher Education Council

Abstract

Despite today's developing healthcare technology, conventional colonoscopy is still a gold-standard method to detect colon abnormalities. Due to the folded structure of the intestine and visual disturbances caused by artifacts, it can be hard for specialists to detect abnormalities during the procedure. Frames that include artifacts such as specular reflection, improper contrast levels from insufficient or excessive illumination gastric juice, bubbles, or residuals should be detected to increase an accurate diagnosis rate. In this work, both conventional machine learning and transfer learning methods have been used to detect non-informative frames in colonoscopy videos. The conventional machine learning part consists of 5 different types of texture features, which are gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), neighborhood gray-tone difference matrix (NGTDM), focus measure operators (FMOs), and first-order statistics. In addition to these methods, we utilized 8 different transfer learning models: AlexNet, SqueezeNet, GoogleNet, ShuffleNet, ResNet50, ResNet18, NasNetMobile, and MobileNet. The results showed that FMOs and decision tree combination gave the best accuracy and f-measure values with almost 89% and 0.79%, respectively, for the conventional machine learning part. When the transfer learning part is taken into account, AlexNet (99.85%) and SqueezeNet (98.80%) have the highest performance metric results. This study shows the potential of both transfer learning and conventional machine learning algorithms to provide fast and accurate non-informative frame detection to be used during a colonoscopy, which may be considered the initial step in identifying and classifying colon-related diseases automatically to help guide physicians.

KEYWORDS

colonoscopy, feature extraction, image processing, machine learning, transfer learning

1 | INTRODUCTION

Abnormality detection in the colon and rectum is challenging during conventional colonoscopy (CC) procedures. While moving the probe during CC, the physicians have to continuously check the monitor on which hundreds of frames are displayed in short periods.[1] In addition, the automatic labeling of polyps, bleeding, and various abnormalities on such videos is an ongoing effort by many researchers. These artifacts hamper the accurate diagnosis rate thus the automatic removal of such non-informative frames would be highly beneficial for physicians. Moreover, the execution of algorithms for subsequent automatic identification of abnormalities only on the informative frames would also speed up the process and yield more accurate results. A new method for detecting and classifying gastrointestinal diseases using wireless capsule endoscopy (WCE) images is proposed. The core of this framework is the combination of HSI active contour and a newly improved detection method using MAP.[2] Another study used deep learning method for gastrointestinal disease detection.[3] However, numerous non-informative frames arise during this procedure due to specular reflection from the light source, improper contrast levels from insufficient, or excessive illumination inside the colon, gastric juice, bubbles, or residuals.[4,5] In the last two decades, many studies aimed at non-informative frame detection for various causes of non-informativeness as well as feature extraction approaches.[6] Oh et al. proposed using Canny edge detection and thresholding[7] and separation of in-focus and out-of-focus frames.[8] Sun et al. suggested a method for removing non-informative frames from wireless colon endoscopy recordings, such as stomach liquid and bubbles. The feature extraction methods employed in that research were the local histogram, local binary pattern (LBP), and discrete cosine transform (DCT).[9] Another study by van Dongen et al. focused on the automatic recognition of informative frames for early diagnosis of esophageal cancer. The color histogram and DCT coefficients were used as features for classification.[10] Tajbakhsh et al. tiled the images and calculated two-dimensional (2D) DCT dominating coefficients for each tile. They then rebuilt the image and utilized a separate map to determine whether that particular frame was non-informative or not. Bubbles, motion blurring, and reflection artifacts were all identified as non-informative frames in that study.[11] Tong et al. claimed that the Harr wavelet transforms effectively detects a blurry image and even quantifies its blurriness level.[12] Arnold et al. employed the discrete wavelet transform (DWT) coefficients as features from the colonoscopy videos and simply looked at the brightness channel of the images.[13] Using discrete Fourier transform (DFT) and texture analysis, An et al. devised a method for detecting out-of-focus frames.[14] To identify bleeding, polypectomy, residue, or feces in colonoscopy videos, Cho et al. suggested employing a hierarchical SVM model.[15] In order to deblur videos, they proposed a Spatiotemporal Pyramid Network (SPN) and Spatiotemporal Pyramid Generative Adversarial Network (SPGAN).[16] The focus measure operators (FMOs) are a collection of feature extraction methods divided into six groups/families. The gradient operator, Laplacian, DWT, DCT, image statistics, and a category that comprised miscellaneous feature extraction methods are among these categories. FMOs can be used to calculate the focus level of an image for each pixel. FMOs have been presented as a method for determining image quality[17] or sharpness.[18] Furthermore, they were used on microscopic images.[17–19] In this study, we propose extracting features from colonoscopy frames using a single methodology or a combination of methodologies to identify non-informative frames with various artifact types, such as motion artifact, specular reflection, improper contrast levels, gastric juice, bubbles, and residuals. To the best of our knowledge, FMOs[20] as a whole, as well as texture analysis techniques like gray level run length matrix (GLRLM)[21] and neighborhood gray-tone difference matrix (NGTDM),[21] have never been employed to detect non-informative colonoscopy frames. Furthermore, our database contained six different types of artifacts, compared with one or two types used in several earlier studies. FMOs produced the best f-measure and accuracy values, which is why we investigated the performance and computation time of each FMO family separately. In recent years, deep learning (DL) models have been used to handle a variety of image processing challenges, which necessitate large datasets. Convolutional neural networks (CNN) are this field's most widely used deep learning technique. However, access to a large volume of data is not always possible for a given problem in the medical area. To overcome this difficulty, transfer learning was developed.[22] Transfer learning with different DL architectures is a relatively new approach for detecting non-informative frames, in addition, to feature extraction and classification applications. In a recent study, Yao et al. employed gray level co-occurrence matrix (GLCM) and CNN for feature extraction and classification on detecting non-informative frames only with blurring and specular reflection artifacts. Their database included 12 830 informative and 3829 non-informative frames.[1] In another study, Islam et al. used transfer learning on several well-known architectures such as AlexNet, GoogleNet, ResNet, and SimpleNet. Frames containing artifacts like blurring, water, and bubbles were included in their non-informative frames set.[23] In another study,

FIGURE 1 Informative (A) and non-informative (B–D) sample frames from our database: (a) clear, (b) bubbles, (c) specular reflection, (d) motion artifact.

Putten et al. tested ResNet and the Hidden Markov Model (HMM) approach on a dataset with 3883 frames.[24] The main goal of this study is to compare 8 different transfer learning methods with various depths and number of parameters, a total of 143 image textural and statistical features from different families of methods, and 3 different conventional machine learning-based classification approaches to detect 6 different types of non-informative frames automatically. To the best of our knowledge, no studies have covered such a wide range of methodologies in this topic of research.

2 | MATERIALS AND METHODS

2.1 | Data acquisition

The conventional colonoscopy (CC) videos used in this research were retrieved from the website in reference.[25] From 43 videos, a total of 11 491 frames were extracted. The videos included frames from both healthy and diseased colons (ulcerative colitis, Crohn's disease, cancer, and polyps). The frames were labeled as informative and non-informative by visual inspection. Any frame that included the artifacts such as motion artifact, specular reflection, inappropriate contrast levels, gastric juice and bubbles, and residuals were labeled as non-informative. Figure 1 depicts sample frames that are either clear (a) or that include bubbles (b), specular reflection (c), and motion artifact (d).

In order to have consistency between machine learning and transfer learning parts of the study, the frames were split into training (10 113 frames from 37 videos), validation (714 frames from 3 videos), and test (664 frames from 3 videos) sets. In the training, validation, and test sets 5064, 485, and 507 frames were non-informative, respectively. In order to make this part of our study more understandable, we added both the number of frames and all ratios of training, test, and validation parts to the manuscript. This information is shared in Table 1.

2.2 | Study overview

As shown in Figure 2, after the frame extraction from the videos and labeling each frame as informative or non-informative, we followed two pathways in our study. The first one included the use of conventional machine learning methods (feature extraction and classification) subsequent to several preprocessing steps. The second one involved investigating several transfer learning architectures to detect the informative frames. Finally, a comparison of the performances of these methodologies was performed. Pre-trained models function as a fundamental basis for the application of transfer learning. These models have acquired valuable features from extensive datasets, such as ImageNet, which can prove advantageous for a multitude of tasks. Fine-tuning pre-trained models requires fewer labeled data. Since they have learned general characteristics, they frequently generalize well to new tasks, even when task-specific data are limited. The models used in this study are so diverse because we aim to conduct a comprehensive study. We also conducted an ablation study to make our research deeper. Experimental research in disciplines such as machine learning and deep learning requires ablation studies. They involve meticulously removing or modifying specific components, features, or parts of a model or system to determine their individual contributions and performance effects. Ablation study is important for understanding model components. Since we wanted to examine the effect of the normalization layer as an ablation technique in this study, we wanted to examine the

TABLE 1 Total number and ratio of frames used in this study.

	Informative	Uninformative	Total	Ratio
Training	5049	5064	10 113	88%
Validation	229	485	714	6.2%
Test	157	507	664	5.8%

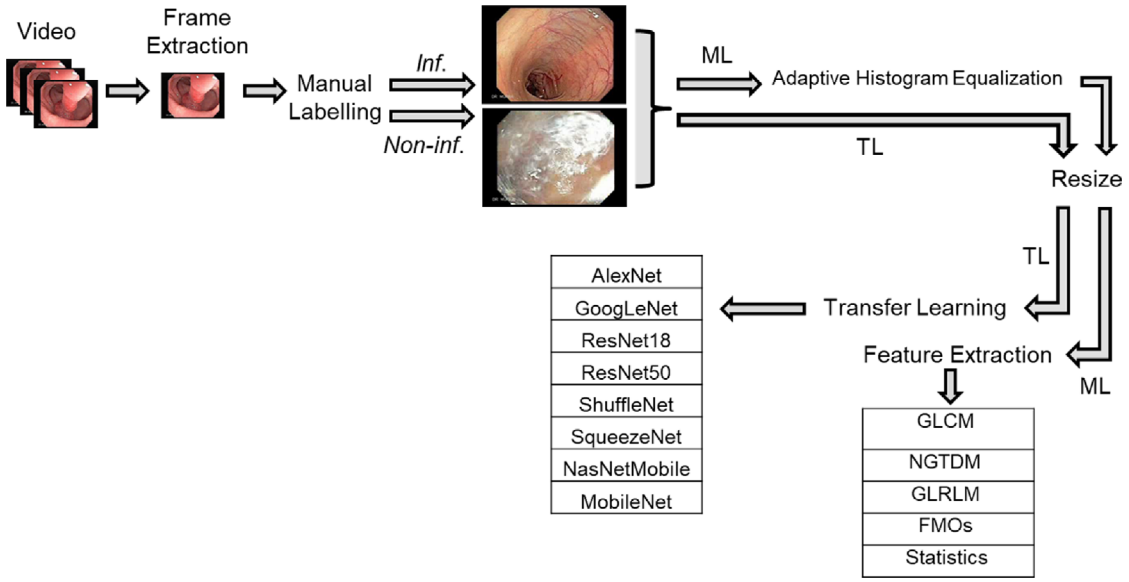


FIGURE 2 The overview of the scheme used in this study.

effect of removing one of the normalization layers before removing them all. In this study, we looked at the effect of normalization as an ablation technique. Since the depths of the pre-trained models, we use vary from each other, the number of normalization layers they contain is also different from each other. For this reason, we first removed one normalization layer from all models and then examined how it affected the result by removing all normalizations.

2.3 | Conventional machine learning

2.3.1 | Preprocessing

As the first preprocessing step, automatic cropping was performed on video frames to obtain a consistent/fixed size (176-by-156 pixels). This procedure helped to remove black regions from the colonoscopy images that contained the date and the patient's name at the periphery. Secondly, the adaptive histogram equalization (AdaptHistEq)[26] was employed to improve the contrast in the images. This method is different from the conventional histogram equalization (HE). In the AdaptHistEq method, histogram equalization is performed on small tiles (8-by-8-pixel squares) rather than the entire image as in the conventional HE. The texture features were extracted from the frames once the cropping and AdaptHistEq-based preprocessing steps were completed.

2.3.2 | Feature extraction

The features used in this study can be grouped into five categories: focus measure operators (FMOs),[20] first-order statistics[20] (kurtosis, standard deviation, and skewness), gray level co-occurrence matrix (GLCM),[27] gray level run length matrix (GLRLM),[27] and neighborhood gray-tone difference matrix (NGTDM).[27] A total of 143 features were extracted: 100 GLCM, 7 GLRLM, 5 NGTDM, 28 FMOs, and 3 first-order statistics. To be more specific, the FMOs included 6 gradient- (Gaussian derivative, energy of gradient, threshold gradient, squared gradient, tenengrad, and tenengrad variance), 4 Laplacian- (modified Laplacian, energy of Laplacian, variance of Laplacian, and diagonal of Laplacian), 3 wavelet- (sum of wavelet coefficients, variance of wavelet coefficients, and ratio of wavelet coefficients), 2 DCT- (energy ratio and reduced energy ratio), and 5 statistics-based (gray level variance, gray level local variance, normalized gray level, histogram entropy, and histogram range) features in addition to 8 miscellaneous features (absolute central

moment, Brenner's measure, image contrast, image curvature, Hemli's and Scherer's mean, steerable filters-based features, spatial frequency measure, and Vollath's autocorrelation). The mathematical details of the FMOs can be found in Reference [20] From each frame, 143 features were extracted which were then normalized using the min-max method, in which the minimum and maximum values for each feature were normalized to 0 and 1, respectively, and the others were linearly scaled. Previous studies have employed various features, with varying levels of accuracy. To conduct a comprehensive study, we have chosen to incorporate all these features, ensuring a holistic approach. This inclusion allows us to compare the results of both machine learning and transfer learning methods. Furthermore, we have conducted comparative analyses in this study.

2.3.3 | Classification

Only the training and test sets were used in this part of the study. We first employed all features to train and test the system, then focused on FMOs and used a decision tree classification approach to investigate the performance of several feature families listed under FMOs both as one category (gradient-based, Laplacian-based, etc.) and in combination with other categories. Three types of classifiers were investigated in terms of execution time and performance including random forests, support vector machines, and decision tree techniques, and the decision tree classifier was faster than the other classifiers while maintaining similar accuracies. Furthermore, decision tree classification has several advantages, including being comprehensive and user-friendly and having high specificity. A decision tree is a classification method that builds a model in the form of a tree structure with decision nodes and leaf nodes. The decision tree algorithm is created by breaking the dataset into smaller and smaller sections. One or more branches can be found in a decision node.[28] The number of the split was 7, and the cross-validation value was set to 10.

2.4 | Transfer learning

To detect non-informative frames, several well-known transfer learning architectures such as AlexNet, SqueezeNet, GoogleNet, ShuffleNet, ResNet50, ResNet18, NasNetMobile, and MobileNet that are pre-trained with the ImageNet were investigated. MATLAB platform was utilized to fine-tune the models/networks. As a matter of fact, the number of convolution layers, fully connected layers, pooling layers, and parameters differ between the models.[29] Transfer learning parameters are given in Table 2. The initial learning rate is selected as 0.0003. We did not choose the initial learning rate as too low or too high, as both have their drawbacks. If the initial learning rate is chosen too low, the training process will take too long, and if it is chosen too high, undesirable results may be obtained in the loss function. The weights were updated in each training iteration by calculating the gradient of the loss function, with "MiniBatchSize" set to 10. In this case, 6 epochs were used to train the model, with each epoch being the number of times the gradient descent algorithm was applied to the whole training set. The "MaxEpochs" value has been set to 6 in order to avoid the overfitting problem. The training dataset was shuffled in each epoch to obtain more accurate results. In Table 5, results were evaluated not only in terms of accuracy value but also f-measure. The formulas of variables are defined in equations 1–4.[30] The idea here was to compare the performances of different architectures with different depths and complexity for this specific problem. We resized images according to the network types. No image enhancement was performed.

TABLE 2 Transfer learning parameters and values.

Training parameter	Value
Initial learning rate	0.0003
MiniBatchSize	10
MaxEpochs	6
Shuffle	"every-epoch"

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

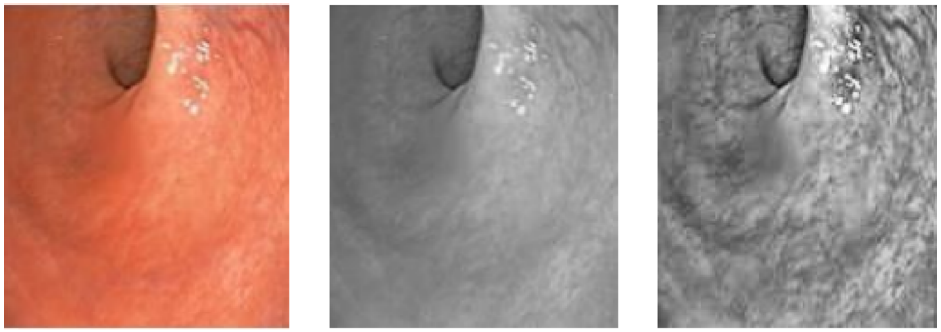
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{f} - \text{measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (4)$$

3 | RESULTS AND DISCUSSION

3.1 | Preprocessing

The cropping and adaptive histogram equalization (AdaptHistEq) used in the preprocessing step resulted in improved visibility of details (such as the vessels in the



(A) (B) (C)

FIGURE 3 (A) Original frame, (B) gray scale, (C) adaptive histogram equalization output.

Feature type	No. of features	Accuracy	f-measure	Precision	Recall
GLCM	100	0.8509	0.6551	0.7231	0.8987
NGTDM	5	0.7410	0.6126	0.4739	0.8662
GLRLM	7	0.6491	0.4466	0.3561	0.5987
FMOs	28	**0.8886**	**0.7874**	0.7173	0.8726
Statistics	3	0.6220	0.4990	0.3634	0.7962
All features	143	0.8419	0.7042	0.6313	0.7962

TABLE 3 Machine learning-based classification results for 5 different texture features using a decision tree algorithm.

colon tissue) as shown in Figure 3. AdaptHistEq method outperformed conventional HE because it includes a contrast-width restriction that prevents noise amplification and improves frame visibility.

3.2 | Conventional machine learning

This research mainly aimed at the comparison of the informative and non-informative frame discrimination performances of conventional machine learning and transfer learning-based approaches. In that regard, a machine learning-based technique was studied employing various features and various feature extraction methodologies. While Table 3 shows the results of machine learning-based categorization for each type of feature extraction approach, Table 4 shows the results of FMOs subcategories. When Table 3 is evaluated, the following interpretation can be made. Focus measure operators are a kind of machine learning model, commonly referred to as FMOs, and exhibit a remarkable accuracy level of 0.8886. This finding suggests that FMOs have the ability to effectively capture essential patterns and information, hence enabling precise classification of instances. FMOs have demonstrated outstanding performance and hence merit serious consideration for the purpose of feature selection.The GLCM method exhibits a notable level of accuracy, as evidenced by a value of 0.8509. This particular feature type demonstrates efficacy in differentiating between classes and may be considered a viable option when seeking a trade-off between accuracy and the quantity of features. This particular approach is highly effective for our classification work. The accuracy of NGTDM is comparatively lower than that of FMOs and GLCM, with a value of 0.7410. The efficacy of this particular feature type in capturing the unique attributes of our data may be limited, and it may be essential to engage in further feature engineering or selection in order to enhance its performance. GLRLM features have the lowest accuracy among the investigated feature types, with a value of 0.6491. This implies that the aforementioned features may not sufficiently capture the fundamental patterns within the dataset for the purpose of categorization. It may be necessary to investigate alternate feature types or preprocessing techniques. The feature type "Statistics" demonstrates the lowest accuracy, with a value of 0.6220. This suggests that placing exclusive reliance on statistical features may prove inadequate in attaining precise classification within the given context. It is advisable to contemplate the integration of additional types of features or the improvement of our feature extraction methodology. When we combine all the features together, results in a relatively high level of accuracy, indicating that the combination of features from various categories has the potential to capture further information. Nevertheless, it is crucial to take into account the trade-off between dimensionality and computing complexity that arises from utilizing all features. When we examine the categories of FMOs group by group, we understand the following results in Table 4.

TABLE 4 Machine learning results of FMO subcategories using decision tree classification.

FMO feature type	No. of features	Accuracy	f-measure	Precision	Recall
Laplacian-based	4	0.5858	0.5201	0.3582	0.9490
Wavelet-based	3	0.6084	0.4758	0.3481	0.7516
Statistics	5	0.7425	0.5799	0.4720	0.7516
Gradient-based	6	0.7575	0.5903	0.4915	0.7389
Miscellaneous	8	0.7963	0.6800	0.6120	0.7650
DCT-based	2	**0.8358**	**0.7212**	0.6326	0.8981

The Laplacian-based method is a technique that relies on the Laplacian operator. The feature type with the lowest accuracy among the several types is the Laplacian-based FMO features, with an accuracy score of 0.5858. This implies that they might not adequately capture the fundamental patterns in our data for the purpose of classification. It may be necessary to investigate alternate types of features or contemplate using more feature engineering techniques. The other feature employed in this study is wavelet-based. The wavelet-based FMO characteristics demonstrate a moderate level of accuracy, as indicated by the value of 0.6084. These characteristics have the potential to gather information in the frequency domain of the data. Additional analysis and feature engineering may be required in order to improve the performance of the classification task. Another feature type in our study is "Statistics." The statistical FMO features demonstrate a satisfactory level of accuracy, as indicated by the value of 0.7425. These features are capable of capturing information regarding the distribution and variation of data, which might prove to be valuable in the context of classification. The integration of these feature categories has the potential to yield enhanced outcomes. Gradient-based features based on gradients in the FMO features provide reasonable accuracy with 0.7575 accuracy. Gradients in the data are captured by these features, making them potentially valuable for finding edges and patterns. Feature extraction or combination with other types may improve their usefulness, but they are still a good option for classification. Miscellaneous groups include different types of 8 features. Miscellaneous features show a high degree of accuracy as well (0.7963). This feature type's classification efficacy can be attributed, in part, to the fact that it captures distinct and varied aspects of the data. It is an attractive possibility for our classification study. DCT-based FMO features have the best accuracy (0.8358) of any of the feature categories. This means that they successfully classify our data by capturing its most notable patterns and attributes. Whether we are selecting features, DCT-based features are a great option for this work. In addition to this, DCT-based features succeed in these results using only 2 features. It means DCT-based features can be used for this purpose effectively.

In conclusion, the selection of feature type is a crucial component of the machine learning pipeline. Focus measure operators (FMOs) and gray level co-occurrence matrix (GLCM) exhibit high levels of accuracy and hence should be regarded as the primary factors to be taken into account. The utilization of all attributes has been found to yield favorable outcomes; however, it is important to note that this approach may introduce heightened intricacy. Additional research and exploration are necessary to fully explore the capabilities of NGTDM. Furthermore, it is important to extract features in order to optimize its performance. On the contrary, GLRLM and Statistics may not be the most appropriate feature types for our particular classification task. It is important to consider the trade-offs among feature dimensionality, computational resources, and the task's unique requirements while making decisions for feature selection.

The results showed that FMOs had the best accuracy and f-measure values with almost 89% and 0.79%, respectively, compared with GLCM, GLRLM, NGTDM, and statistics-based feature extraction algorithms. According to machine learning results, FMOs features have the highest accuracy (0.8886) and f-measure (0.7874) values. It is easy to understand that with FMOs using less feature and obtaining similar results is possible. The accuracy and f-measure values obtained when all features were combined were not as high as when only FMOs were used as a complete set. Using a decision tree classifier, classification performances based on feature families of FMOs such as the DCT, gradient, Laplacian, wavelet, miscellaneous, and image statistics are shown in Table 2. Miscellaneous (MISC) features (only 8 features) and DCT-based features (only 2 features) produced classification accuracies of 79% and 83%, respectively. Using DCT-based features only, allows both to get high results and reduce the workload. In conclusion, the selection of the FMO feature type should be in accordance with the specific classification task's demands and the attributes of the data. The DCT-based and miscellaneous FMO features exhibit outstanding performance and so require

TABLE 5 Transfer learning results of informative and non-informative discrimination.

Model	Accuracy	f-measure
AlexNet	**0.9985**	**0.9968**
SqueezeNet	**0.9880**	**0.9748**
GoogLeNet	0.9849	0.9690
ShuffleNet	0.9639	0.9250
Resnet18	0.9623	0.9191
NasNetMobile	0.9428	0.8774
Resnet50	0.9367	0.8591
MobileNet	0.9051	0.8153

development. Gradient-based and statistical features have been found to exhibit high accuracy and can serve as essential parts of feature extraction. Additional revision may be necessary to enhance the classification performance of wavelet-based and Laplacian-based features.

3.3 | Transfer learning

In this study, the feasibility of 8 transfer learning methods has been investigated. Table 5 depicts that the best and the worst performances were attained using AlexNet (99.85%) and MobileNet (90.51%), respectively. However, even the lowest performance obtained using a transfer learning method was higher than the best result of the conventional machine learning approaches. AlexNet exhibits superior accuracy compared with other models, positioning itself as the best performer in effectively classifying cases. Both SqueezeNet and GoogLeNet exhibit a noteworthy degree of accuracy, with SqueezeNet demonstrating a slightly greater performance in comparison with GoogLeNet. However, fewer parameters and effective compression techniques help SqueezeNet reduce training and running times. ShuffleNet and ResNet18 demonstrate similar degrees of accuracy, both achieving satisfactory results in terms of accuracy. These solutions are suitable choices in cases when there is a need to achieve a balanced optimum between accuracy and processing efficiency. NasNetMobile and Resnet50 demonstrate relatively lower accuracy compared with the top-performing models. Nevertheless, they serve as acceptable alternatives, especially in situations where computational resources are readily available. Among the models provided, MobileNet demonstrates the least favorable level of accuracy. While sacrificing some accuracy, this method may be preferred in situations where the focus is on minimizing model size and processing resources rather than achieving the highest level of accuracy. In the final analysis, the selection of a model should be by the particular demands of the given work and the limitations imposed by the available resources. When prioritizing accuracy, it is advisable to consider employing advanced models like as AlexNet, SqueezeNet, and GoogLeNet. By contrast, in situations where there exist limitations on computational resources, models such as ShuffleNet and MobileNet demonstrate greater suitability for the assigned task. We also observed that the computational cost was more advantageous in transfer learning than the fastest machine learning-based classification approach. Also, pre-trained models save time. Table 6 contains the confusion matrices for all pre-trained models that were employed. By providing an exhaustive overview of the performance metrics of each pre-trained model, the confusion matrix enhances the comprehension of their respective capabilities and limitations. By including this supplementary data, we hope to augment the simplicity and comprehensiveness of our examination.

When we applied the ablation technique to our models, we observed varied performance outcomes. For instance, in AlexNet, removing a single normalization layer led to a slight decrease in classification accuracy, while removing all normalization layers resulted in a significant decrease. This could be attributed to the interaction between different normalization methods. Interestingly, in models such as Resnet18, Resnet50, ShuffleNet, and MobileNet, removing the normalization layers led to increased accuracy, regardless of the number of layers removed. We chose to ablate the normalization layer because it is a common element in most of the models we studied. It serves as an intersection layer, allowing us to observe its impact. It is worth noting that GoogLeNet and SqueezeNet do not include normalization layers, making the ablation of these layers an effective means to understand the behavior of each model. Since our models vary in depth, the effects of removing normalization layers differ from model to model, resulting in varying rates of increase or decrease in accuracy. The detection of non-informative frames from videos has been the focus of various studies and is a difficult problem to deal with, as it hinders or delays disease detection for traditional colonoscopy videos and wireless capsule endoscopy. When previous works are taken into account, some studies focused on feature extraction and machine learning methods, while others used conventional neural networks to detect non-informative frames. Feature extraction and machine learning-based studies can be summarized as follows: Fan et al. conducted research on WCE frames to choose informative images and they preferred using the histogram and mean shift features in 500 frames.[31] Using color (RGB) and texture (SURF)

TABLE 6 Confusion matrix of transfer learning results for informative and non-informative discrimination.

Model type			Predicted class	
			Informative	Non-informative
AlexNet	True Class	Informative	156	1
		Non-informative	0	507
SqueezeNet	True Class	Informative	155	2
		Non-informative	6	501
GoogleNet	True Class	Informative	156	1
		Non-informative	9	498
ShuffleNet	True Class	Informative	148	9
		Non-informative	15	492
Resnet-18	True Class	Informative	142	15
		Non-informative	10	497
NasNetMobile	True Class	Informative	136	21
		Non-informative	17	490
ResNet-50	True Class	Informative	128	29
		Non-informative	13	494
MobileNet	True Class	Informative	139	18
		Non-informative	45	462

attributes, WCE images were segmented in a different study as either clear, turbid, or bubble-free in 50 videos.[32] Another study's objective was to identify non-informative frames during the bronchoscopy procedure. This study included images with reflections, loss of focus, impurities, and motion blur. They used the DCT spectrum, MPEG-7 edge, zero cross-edge detection, and color transformation (HSV) in 768 frames.[33] Additionally, utilizing edge-based approaches in 387 frames, Rangseekajee and Phongsuphap proposed a method to categorize thoracoscopic frames as informative or non-informative.[34] By means of the discrete Fourier transform (DFT) and texture analysis (gray level co-occurrence matrix), An et al. investigated the feasibility of distinguishing informative and non-informative images on 5971 colonoscopy images.[35] In addition to these studies, edge detection, wavelet-based studies, and color-based studies have been conducted.[1–5] Apart from these kinds of studies, some transfer learning studies were also conducted. For instance, Yao et al. used CNN and the gray level co-occurrence matrix (GLCM) to extract features and classify non-informative frames that only had blurring and specular reflection abnormalities.[1] In a different study, Islam et al. preferred to use transfer learning methods to a number of well-known architectures, including SimpleNet, ResNet, GoogleNet, and AlexNet. They included blurry, watery, and bubble-filled frames in their collection of uninformative frames.[23] ResNet and Hidden Markov Model (HMM) are also used to detect non-informativeness in 3883 frames. Frames can be deemed non-informative for a variety of reasons, such as out-of-focus images, esophageal contractions, motion blur, a profusion of bubbles, and a sharp contrast from the lighting.[24] The studies carried out so far, the methods they used, and their results are listed in Table 7. It is obvious that there are a few traditional machine learning and transfer learning studies. In our study, however, we did not just use one kind of texture feature or transfer learning technique to obtain our results. When our study is examined in terms of machine learning and transfer learning parts, we included more methods than the previous studies did. However, it has been clearly understood that we are more advantageous than most of the other studies in terms of the number of frames and the number of videos collected from different patients in our study. Our non-informative frame detection using transfer learning results showed extremely promising performance and outperformed prior studies. When the previous studies are examined, there is no study that reached such high results with so many frames and methods. Our study has many advantages over the previous methods quantitatively and qualitatively. In addition to these outputs, our non-informative database contains 6 different types of artifacts that are motion artifact, specular reflection, improper contrast, gastric juice, bubbles, and residuals not just two or three. The focus measure operators performed the best among the feature extraction approaches, according to our

TABLE 7 Results of previous works on non-informative frame detection.

Reference number	Data type	No. of data	Method	Result	Performance metric types
1	Colonoscopy	16 659 frames	CNN GLCM	0.77	f-measure
6	Colonoscopy	2000 frames	Edge Detection	0.95	Accuracy
7	Colonoscopy	9 videos	Edge Detection	0.95	Precision
8	Endoscopy	923 frames	Edge Detection	0.95	Accuracy
9	WCE	3 videos	Histogram Color LBP DCT	0.99	Accuracy
11	Colonoscopy	2172 frames	DCT Color	0.97	Accuracy
12	Colonoscopy	2355 frames	Haar Wavelet	0.986	Accuracy
13	Colonoscopy	15 000 frames	2D DWT	0.923	Accuracy
23	Colonoscopy	6805 frames	GoogLeNet ResNet SimpleNet AlexNet	0.94	f-measure
24	Endoscopy	3883 frames	HMM ResNet11	0.91	f-measure
31	WCE	500 frames	Histogram Mean Shift	0.875	Specificity
32	WCE	50 videos	RGB Color Surf Feature	0.852	Accuracy
33	Bronchoscopy	768 frames	Edge Detection HSV Analysis MPEG-7edge Histogram DCT Spectrum	0.93	Accuracy
32	Endoscopy	387 frames	Edge Detection	0.951	Accuracy
33	Endoscopy	5971 frames	DFT GLCM	0.981	Precision

findings. We observed that FMOs may be used to eliminate non-informative frames. We employed five different feature extraction methods to complete the elimination, which resulted in a total of 143 different features. We may claim that all FMO subcategories would be used to detect non-informative frames. When subcategories were examined separately, the miscellaneous family, which had not previously been used in the literature, produced the second-best results for our dataset. Absolute central moment, Brenner's measure, image contrast, image curvature, Hemli's and Scherer's mean, steerable filters-based features, spatial frequency measure, and Vollath's autocorrelation were among the eight characteristics of this miscellaneous family. When compared to machine learning, however, transfer learning, notably AlexNet, produced the best results. To compare these approaches, we used the same database to test the application of the stated architectures as a transfer learning methodology.

4 | CONCLUSION

The responses to the following two questions were studied in this study: (1) Can different texture features such as GLCM, GLRLM, NGDTM, FMOs, image statistics, and traditional classifiers be used to successfully differentiate informative frames from non-informative frames? (2) When compared to conventional machine learning, does transfer learning provide superior results? In this setting, a detailed study like ours has never been published before. The detection of non-informative frames using the methods described above is the first step toward our long-term aim of automatically detecting colon diseases, which necessitates dealing with only informative images. We provided the most comprehensive study to detect non-informative frames and our work will serve as a pioneer model for subsequent research, especially related to real-time colon disease detection studies.

ACKNOWLEDGMENTS

This study was supported by the Turkish Higher Education Council (100/2000 Scholarship).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in el salvador atlas of gastrointestinal video endoscopy at https://www.gastrointestinalatlas.com/english/english.html. These data were derived from the following resources available in the public domain: [Source Name], https://www.gastrointestinalatlas.com/english/english.html.

ORCID

Rukiye Nur Kaçmaz https://orcid.org/0000-0002-3237-9997

REFERENCES

1. Yao H, Stidham RW, Soroushmehr R, Gryak J, Najarian K. Automated detection of non-informative frames for colonoscopy through a combination of deep learning and feature extraction. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology* Society 2019. doi:10.1109/embc.2019.8856625
2. Khan MA, Rashid M, Sharif M, Javed K, Akram T. Classification of gastrointestinal diseases of the stomach from WCE using the improved saliency-based method and discriminant features selection. *Multimed Tools Appl*. 2019;78:27743-27770. doi:10.1007/s11042-019-07875-9
3. Khan MA, Majid A, Hussain N, et al. Multiclass stomach diseases classification using deep learning features optimization. *Comput Mater Contin*. 2021;67(3):3381-3399. doi:10.32604/cmc.2021.014983
4. Majid A, Khan MA, Yasmin M, Rehman A, Yousafzai A, Tariq U. Classification of stomach infections: a paradigm of the convolutional neural network along with classical features fusion and selection. *Microsc Res Tech*. 2020;83(5):562-576. doi:10.1002/jemt.23447
5. Khan MA, Sarfraz MS, Alhaisoni M, Albesher AA, Wang S, Ashraf I. StomachNet: optimal deep learning features fusion for stomach abnormalities classification. *IEEE Access*. 2020;8:197969-197981. doi:10.1109/ACCESS.2020.3034217
6. Ballesteros C, Trujillo M, Mazo C, Chayes D, Hoyos J. Automatic classification of non-informative frames in colonoscopy videos. *Lect Notes Comput Sci*. 2017;10125:401-408. doi:10.1049/ic.2015.0307
7. Oh J, Hwang S, Tavanapong W, De Groen PC, Wong J. Blurry frame detection and shot segmentation in colonoscopy videos. *Int Soc Opt Eng*. 2004;5307:531-542. doi:10.1117/12.527108
8. Oh J, Hwang S, Lee J, Tayanapong W, Wong J, De Groen PC. Informative frame classification for endoscopy video. *Med Image Anal*. 2007;11:110-112. doi:10.1016/j.media.2006.10.003
9. Sun Z, Li B, Zhou R, Zheng H, Meng MQ. Removal of noninformative frames for wireless capsule endoscopy video segmentation. IEEE International Conference on Automation Logistics, 2012, 294–299. 10.1109/ICAL.2012.6308214.
10. van Dongen NC, van der Sommen F, Zinger S, Sekoon EJ, de With PHN. Automatic assessment of informative frames in the endoscopic video. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016, 119–122. 10.1109/ISBI.2016.7493225
11. Tajbakhsh N, Sharma H, Wu Q, Gurudu SR, Liang J. Automatic assessment of image Informativeness in colonoscopy, abdominal imaging, computational and clinical applications. *Lect Notes Comput Sci*. 2014;8676:51-158. doi:10.1007/978-3-319-13692-9_14
12. Tong H, Li M, Zhang H, Zhang C. Blur detection for digital images using wavelet transform. IEEE International Conference on Multimedia Export, 2004, 17–20. 10.1109/ICME.2004.1394114
13. Arnold M, Ghosh A, Lacey G, Patchett S, Mulcahy H. Indistinct frame detection in colonoscopy videos. 13th International Machine Vision Image Processing, 2009, 47–52. 10.1109/IMVIP.2009.16
14. An Y, Kang G, Kim IJ, Chung HS, Park H. Computer-aided diagnosis system for colon abnormalities detection in WCE shape from focus through Laplacian using 3D window. *Second Int Confer Fut Generat Commun Netw*. 2008;2:46-50.
15. Cho M, Kim JH, Kong HJ, Hong KS, Kim S. A novel summary report of colonoscopy: timeline visualization providing meaningful colonoscopy video information. *Int J Colorectal Dis*. 2018;33:549-559. doi:10.1007/s00384-018-2980-3
16. Wang T, Zhang X, Jiang R, Zhao L, Chen H, Luo W. Video Deblurring via spatiotemporal pyramid network and adversarial gradient prior. *Comput Vis Image Understand*. 2021;203:103135. doi:10.1016/j.cviu.2020.103135
17. Eskicioğlu AM, Fischer PS. Image quality measures and their performance. *IEEE Transact Commun*. 1995;43:2959-2965. doi:10.1109/26.477498
18. Wee CY, Paramesran R. Measure of image sharpness using eigenvalues. *Inform Sci*. 2007;177:2533-2552. doi:10.1109/ICOSP.2008.4697259
19. Xie H, Rong W, Sun L. Wavelet-based focus measure and 3-D surface reconstruction method for microscopy images. *Int Confer Intelligent Robots Syst*. 2006;229-234. doi:10.1109/IROS.2006.282641
20. Pertuz S, Puig D, Garcia MA. Analysis of focus measure operators for shape-from-focus. *Patt Recogn*. 2013;46:1415-1432. doi:10.1016/j.patcog.2012.11.011
21. Caruso D, Polici M, Zerunian M, et al. Radiomic cancer hallmarks to identify high-risk patients in non-metastatic colon cancer. *Cancer*. 2022;14:3438. doi:10.3390/cancers14143438
22. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data*. 2016;3:1-40.
23. Islam AB, Alammari MR, Oh J, Tavanapong W, Wong J, De Groen PC. Non-Informative Frame Classification in Colonoscopy Videos Using CNNs. Proceedings of the 3rd International Conference on Biomedical Imaging, Signal Processing, 2018, 2402–2406. 10.1145/3288200.3288207
24. Van Der Putten J, De Groof J, Van Der Sommen F, et al. Informative frame classification of endoscopic videos using

convolutional neural networks and hidden Markov models. IEEE International Conference on Image Processing, 2019, 380–384. 10.1109/ICIP.2019.8802947
25. https://www.gastrointestinalatlas.com/english/english.html [Online], August, 2022.
26. Kurt B, Nabiyev VV. Dijital Mamografi Görüntülerinin Kontrast Sınırlı Adaptif Histogram Eşitleme ile Iyileştirilmesi. VII Ulusal Tıp Bilişimi Kongresi, 2010, 67–79.
27. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern*. 1973;3:610-621. doi:10.1109/TSMC.1973.4309314
28. Patel H, Prajavati P. Study and analysis of decision tree-based classification algorithms. *Int J Comput Sci Eng*. 2018;6:74-78. doi:10.26438/ijcse/v6i10.7478
29. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Neural Inform Process Syst*. 2012;13:1097-1105. doi:10.1145/3065386
30. Yurdusev A, Adem K, Hekim M. Detection and classification of microcalcifications in mammograms images using difference filter and Yolov4 deep learning model. *Biomed Signal Process Contr*. 2023;80:104360. doi:10.1016/j.bspc.2022.104360
31. Fan Y, Meng M, Baopu L. A novel method for informative frame selection in wireless capsule endoscopy video. International Conference IEEE Engineering in Medicine, Biology, Society 2011. doi:10.1109/iembs.2011.6091205
32. Arivazhagan S, Sylvia L, Jebarani W, Jenifer Daisy V. Categorization and segmentation of intestinal content and pathological frames in wireless capsule endoscopy images. *Int J Imaging Robot*. 2014;13:134-147.
33. Grega M, Leszczuk M, Duplaga M, Fraczek R. Algorithms for automatic recognition of non-informative frames in video recordings of Bronchoscopic procedures. *Adv Intell Soft Comput*. 2010;2:535-545. doi:10.1007/978-3-642-13105-9_53
34. Rangseekajee N, Phongsuphap S. Endoscopy video frame classification using edge-based information analysis pre-processing. *Comput Cardiol*. 2011;38:549-552.
35. An YH, Hwang S, Oh JH, et al. Informative-frame filtering in endoscopy videos. *Progr Biomed Opt Imaging*. 2005;5747:291-302. doi:10.1117/12.595622

How to cite this article: Kaçmaz RN, Doğan RS, Yılmaz B. A comprehensive study on automatic non-informative frame detection in colonoscopy videos. *Int J Imaging Syst Technol*. 2024;34(1): e23017. doi:10.1002/ima.23017