



# A New Method Based on Convolutional Neural Networks and Discrete Wavelet Transform for Detection, Classification and Tracking of Colon Polyps in Colonoscopy Videos

Hüseyin Kutlu<sup>1\*</sup>, Fatih Özyurt<sup>2</sup>, Engin Avci<sup>2</sup>

<sup>1</sup> Computer Tech. Dep., Besni A.E. Voc. Sch., Adiyaman University, Besni / Adiyaman 02300, Turkey

<sup>2</sup> Software Engineering Dep., Firat University, Elazig 23119, Turkey

Corresponding Author Email: [hkutlu@adiyaman.edu.tr](mailto:hkutlu@adiyaman.edu.tr)

<https://doi.org/10.18280/ts.400116>

## ABSTRACT

**Received:** 12 September 2022

**Accepted:** 28 January 2023

### Keywords:

*CNN, DWT, SVM, faster R-CNN, colonoscopy, deep learning, polyp tracking, polyp detection, polyp classification*

In this study, a new method based on Convolutional Neural Network (CNN), Discrete Wavelet Transform (DWT) and Support Vector Machine (SVM) is presented for polyp detection, classification and tracking during colonoscopy. The proposed method is constructed in 3 parts. 1) Detection of polyps with deep learning based Faster R-CNN for detection of polyps 2) Classification of detected polyps by CNN-DWT-SVM. 3) Tracking for polyps counting. The proposed method was trained and tested with the Colonoscopy Dataset, a public data set. In the first step of the method, polyp detection was carried out with pre-trained ResNet 50 CNN architecture with 92.6% precision. The regions identified in the second step of the method were classified for four classes adenoma, hyperplastic, lumen, serrated and 94.7% classification accuracy was obtained. With the proposed method, the detection sensitivity of Faster R-CNN was increased from 92.6% to 99.2%, and the accuracy of 95.4% was achieved by using DWT in the classification of polyp classes. In the classification process, 98% correct adenoma, 95% hyperplastic, 90% luminal intestine, 96% serrated polyp were reached. The proposed method reached an average of 94% MOTA in polyp tracking and was able to detect polyp frames with their classes with 99.2% precision.



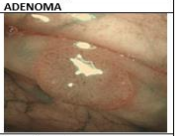



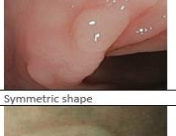

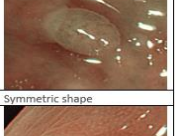
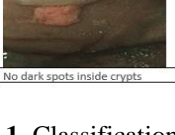
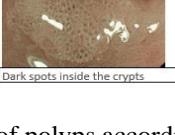
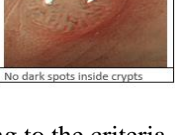
## 1. INTRODUCTION

Colon and rectal cancers (CRC) develop from cells that form the layer that covers the inner surface of the large intestine. According to statistics, bowel cancer is among the top 5 most common cancers. Polyp is a benign tumor. Polyps originate from the inner wall of the colon or rectum. Some polyps (adenomas) may become cancerous. In this case, the polyp should be detected and removed from the body due to the risk of cancer and the body should be checked at regular intervals. Early diagnosis and removal of polyps reduces the risk of CRC [1]. An increase of 1% in adenoma polyp detection was associated with a 3% decrease in the incidence of CRC [2].

CRC is the most beneficial type of cancer. Even with screening tests, the disease can be detected before the disease occurs, and if it has occurred, the symptoms can be detected and the necessary treatment can be made. Because 95% chance of colorectal cancers develop from polyps [3]. If these polyps are detected in screening tests and removed by polypectomy before the cancer develops, a cancer that may occur in the future or may be formed at a very small level will be removed from the intestine. In colorectal cancers, the definitive diagnosis is made by endoscopic imaging (rectoscopy, flexible sigmoidoscopy, colonoscopy) of the polyps in the intestine and microscopic diagnosis of the part to be taken by the pathologist.

Polyps are divided into three classes according to the criteria set by the Workgroup Serrated Polyps and Polyposis (WASP) working group. These classes are called Adenoma, Serrated, Hyperplastic. The WASP classification is based on specific criteria. The brown color of the polyp, the color of the

veins in the polyp, the surface of the polyp, the boundaries, the shape, and the dark spots in it determine the WASP classes. These criteria are shown on the polyp in Figure 1 [4].

KRITER	HYPERPLASTIC	SERRATED	ADENOMA
Surface	 Smooth surface	 Cloud-like surface	 Smooth surface
Border	 Sharp demarcated border	 Indistinct/vague border	 Sharp demarcated border
Shape	 Symmetric shape	 Irregular shape	 Symmetric shape
Crypts	 No dark spots inside crypts	 Dark spots inside the crypts	 No dark spots inside crypts

**Figure 1.** Classification of polyps according to the criteria determined by the Workgroup Serrated Polyps and Polyposis (WASP) working group

As can be seen from the images in Figure 1, polyp classification is a complex process, although it does according to certain criteria. Polyp detection is technically difficult because the same type of polyp may vary in size, color and

texture, and many polyps are not clearly visible in the intestinal mucosa. Colonoscopy is a highly operative procedure and may miss 25 or 28% polyps in colonoscopy [5]. In this study, a computerized diagnostic system was proposed by detecting, classifying and tracking polyps in endoscopic video images used for diagnostic or screening. In this study, a new method, which uses the pre-trained CNN architectures, detects the polyp object from the colonoscopy videos, extracts and classifies the feature vector and tracking and counts them with the help of the features is proposed. Pre-trained CNN architectures were used to classify polyps. With a large data set, the features of the image can be obtained by passing the image in one go from pre-trained CNN architectures. This method is a quick and easy method since it does not require training during the feature extraction phase. There are studies in the literature that make image classification by extracting features from pre-trained CNN architectures [6]. In some of these studies, image classification was performed by fusing the features obtained from CNN architectures. Features that fused reduced to increase the performance of the classifier [7].

DWT was applied to the feature vector obtained from CNNs, meaningful information was selected and the size of the feature vector was reduced. This increased the performance of the classifier even further, as in the fusing of CNN features. There have been studies in the literature where DWT has been used as a feature reduction method [8]. Reducing the size of the feature vector is an important strategy to improve classification performance. As the size of the feature vector increases, it is stated that the sample size should increase exponentially in order to obtain an effective estimation of multivariate densities [9]. In the literature, DWT has been used in many studies to obtain the feature vector and reduce the size of the feature vector. DWT has been frequently used in microarray data analysis [10-12]. Microarray vectors, such as CNN feature vectors, are linear vectors that do not contain any open time or field variables. It is also used to find continuity in DWT object shape information [13, 14]. In the experimental studies, discontinuities were observed in the image and signal representations of the CNN features of detected polyp regions, and DWT was applied to the CNN features to detect and reduce these discontinuities and the feature vectors were classified with SVM.

The proposed method consists of three basic steps. The first step is the detection of polyps. Polyps were detected with Faster R-CNN in this step. The Faster R-CNN algorithm is based on the CNN architecture. In this experimental study, it has been tested with the success of pre-trained CNN architectures with transfer learning in polyp detection with Faster R-CNN. ResNet50 architecture which uses the highest performance in polyp detection is used. In the second step, the polyp regions detected by Faster R-CNN were cropped from the video images. The features of the cropped images are extracted from the last fully connected (FC) layers of the pre-trained CNN architectures. Features were compared with SVM in terms of classified accuracy rates and feature extraction times. ResNet101 architecture has reached the highest rate in terms of classification accuracy. With ResNet101 architecture, feature extraction time is higher than other architectures. Therefore, by fusing different CNN architectures, it is desired to reduce the feature extraction time without reducing performance. For this purpose, AlexNet, ResNet18, SqueezeNet architectures were fused to achieve the classification performance achieved with ResNet101. This method saved about 55% of the feature extraction time. In

addition, feature reduction methods, which are frequently used to improve classifier performance, have been tried. Relief, PCA, correlation-based feature selection (CFS), DWT methods have been tried on the fused feature vectors. Among these methods, DWT yielded better results than other algorithms in increasing accuracy. In addition, the feature vector (ARS+DWT) obtained by applying DWT to the feature vector (ARS) which is formed by fusing the features of AlexNet, Resnet18, SqueezeNet CNN architectures, was used in the cost stage of object tracking algorithms which is the third stage of the proposed method. DWT contributed positively to tracking performance in terms of both time and mean accuracy. This resulted in a 30% reduction in the IDsw score. This decrease caused an increase in the MOTA score. The main contributions of the study are listed below.

1. According to our research, there are no studies or practices that tracking by classifying polyps into three classes. In this study, polyp images were identified and classified into three classes.

2. Polyp features that extracted from pre-trained CNN architectures are classified. Object tracking was performed using the affinity ratio of the features.

3. The most efficient features are obtained by comparing the features in the last FC layers of the eleven pre-trained CNN architecture and combining the appropriate ones.

4. Using DWT, feature size is reduced, classifier mean accuracy value is increased.

The rest of the study is organized as follows, In Chapter 2, previous studies are reviewed. Section 3 describes the proposed method. Chapter 4 gives detailed information about the experimental process. Finally, Chapter 5 presents the results of our study.

## 2. RELATED WORKS

The classification of polyps in colonoscopic images is a current issue and many studies have been carried out on it. In the studies carried out with handcrafted features, in tissue analysis, only adenoma polyps were classified with Color Wavelet Covariance (CWC) features and LDA in their dataset and obtained 97% specificity. They also used histograms of the text on obtained from the filter bank and LBP of the image fragments in their dataset and identified the polyp images with 90% specificity. In their studies according to the shape or appearance of polyps, they classified the images in their dataset as polyp and non-polyp images according to their geometric radii. They proposed the least squares analysis method to increase the distinguishability of polyp non-polyp classification and obtained a recall value of 71.67%. It reached 82 true positive rate values using Edge cross-section profiles (ECSP). Polyp non-polyp classification was made by segmentation. They obtained 71.66% recall value in the study.

Deep learning has gained superiority to handcrafted feature extraction methods in image processing. Many CAD systems such as brain, liver, lung, chest, skin with deep learning algorithms have been proposed in the literature [15, 16]. Polyp sensing studies with deep learning are usually conducted on polyp non-polyp sensing. They obtained 91.76% sensitivity success using AlexNet CNN architecture. They identified polyps with a series of CNNs, each of which specialized in a polyp feature in the polyps localized by feature such as color, texture, shape. They classified adenoma vs hyperplastic polyp images from the studies that classified polyp types. In the first

stage, polyp non-polyp classifier was proposed. In the second stage, transfer learning and polyp images were classified according to their types. They obtained 85% accuracy in PWH classification and 98% accuracy in polyp non-polyp classification. In their study, they obtained CNN features of fixed size image fragments with a sliding window on the images that make up the colonoscopy video. Polyp species can also be classified according to CNN features, but reported only polyp non-polyp classification results in their studies. In this classification, they obtained an accuracy rate of 98.65%. Byrene et al. [17] classified adenoma and hyperplastic polyps by CNN. They designed the model with a 50ms Delay and achieved 94% success in classifying polyp types. They classified in vivo differentiation of polyp species by SSD object detection method with 85% accuracy. Mo et al. [18] used Faster R-CNN in their study. They have shown that they can work quite well for polyp detection with transfer learning. Shin et al. [19] proposed a post-learning scheme to improve the detection performance of Faster R-CNN. The post-learning scheme automatically re-trained the Faster R-CNN with false positives and increased the correct detection rate. Wang et al. [20] proposed a new anchor-independent method for detecting polyps. They stated that the study could work with high speed and accuracy in detecting polyps of various sizes. Urban et al. [21] proposed polyp detection method with YOLO object detection algorithm. Wang et al. [22] proposed a SegNet-based detection algorithm, one of the deep learning semantic pixel-based segmentation methods for polyp detection. Qadir et al. [23] proposed the Mask B-ESA architecture for polyp detection in their study.

Most of the studies in the literature independently detect polyps in the video frame. This creates vibration due to the small intensity fluctuation between video frames. In order to overcome this problem, there are also some studies that tracking polyps. Zhang et al. [24] proposed a two-stage method for polyp tracking. In the first stage, they used pre-trained ResYOLO algorithm which is one of the deep learning object detection algorithms for polyp detection. In the second stage, they proposed a discriminant correlation filter-based monitoring approach.

### 3. PROPOSED METHOD

In recent studies, deep learning algorithms have shown higher performance than traditional methods. Deep learning algorithms work with many parameters. They need a lot of training data to optimize these parameters. When it comes to medical images, it is time-consuming to obtain high amounts of data. In this study, instead of reducing the amount of data by dividing polyp images into classes, polyp detection path was firstly taken into consideration. Due to lack of data, pre-trained CNN architectures have been used with large data sets such as ImageNet, Cifar10. Medical image classification can be used with transfer learning [25]. Features extracted from CNN architectures represent an image. CNNs with different architectures can represent one image better than another. Good representation brings good classification.

In this study, a new method for polyp detection, classification and tracking was proposed. The proposed method consists of three steps.

In the first step of the method, polyp regions were detected by Faster R-CNN architecture without polyp classification. In Faster R-CNN architecture, the pre-trained Resnet-50 CNN

architecture were used as the CNN architecture. Features were extracted from the 40th relu layer (activation\_40\_relu layer). At this stage, the amount of data is increased by not dividing the data set into classes. The performance of the deep learning object detection method is increased with high amount of data.

The second step of the method is the process of classifying the detected polyp regions after crop and resizing. The second step consists of three steps (i) In the first step, the feature vectors obtained from the pre-trained CNN architectures are fused to form a powerful feature vector. (ii) In the second step, DWT, which is a process that processes the whole feature vector, reduces the size of and sub-sampling the feature vector, was applied to the feature vector. (iii) In the third step, the DWT-processed feature vector derived from CNN architectures is classified with the SVM classifier. The first and second steps of the proposed method are shown in Figure 2.

In the third step of the proposed method, the object tracking process was performed with Kalman filter and Hungarian algorithm.

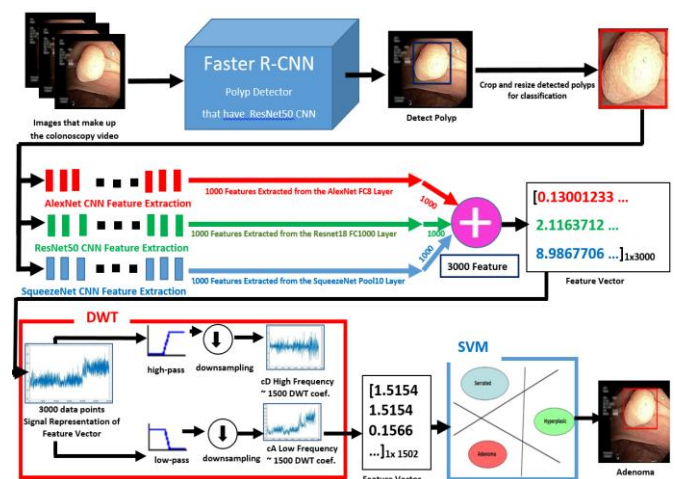


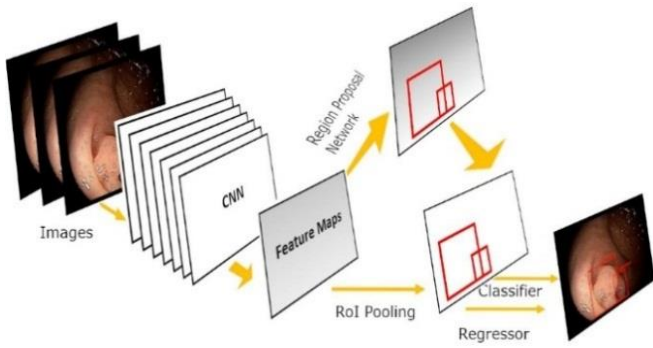
Figure 2. Diagram of the first and second steps of the proposed method

#### 3.1 Object detection with faster regional convolutional neural networks (Faster R-CNN)

The purpose of Faster Region Convolutional Neural Networks (Faster R-CNN) is to take the image as input and accurately determine the coordinates of the objects in the image. If the Faster R-CNN is considered as a system, the input of this system is the image, the output is the bounding box that frames the object in the image and the labels indicating the object's class. Faster R-CNN can detect a objects in an image, as well as objects of different classes. The first release of R-CNNs was announced in 2013 [26]. Following the first release, Fast R-CNN was announced in 2015 [27]. In the same year, a faster version of the Faster R-CNN version was described [28]. Figure 3 shows the Faster R-CNN architecture.

Apart from the CNN architecture, the Faster R-CNN consists of Region Proposal Network (RPN), Region of Interest Pooling (RoI), Classifier and Regression units. In the RPN volume, the image is hovered on the nxn-sized floating window on the Feature Map. For each window case, 9 anchors are produced using 3 different scales, all with the same center. Each of the anchors calculates the intersection value with the ground-truth boxes specified by the programmer, which is the actual location of the objects in the image. This value is given

to the Regresor network for training. The purpose of the ROI unit is to bring the anchors of different sizes from the RPN unit to a fixed size. The RoI Pooling output is assigned to the classifier for classification.



**Figure 3.** Faster R-CNN architecture

There are many object-detection algorithms based on deep learning. YOLO [29], SSD [30] are some of these algorithms. In this study, Faster R-CNN algorithm was preferred because high accuracy classification was targeted.

In this study, polyp detection was performed with Faster R-CNN object detection algorithm without any classification. The Faster R-CNN algorithm consists of three network structures. The first of these networks is the feature extraction network. This network is followed by a region recommendation network (RPN) trained to generate object suggestions and a trained network to predict the real class of object suggestions. In this study, pre-trained Resnet50, ResNet101 and InceptionResNet CNN architectures are tested as feature extraction network. Pre-trained ResNet50 CNN architecture was used in this study as it outperforms ResNet101 and InceptionResNet architectures with 92.81% sensitivity, with ResNet50 CNN architecture. All layers and weights of the Resnet50 CNN architecture are transferred except the last three layers. The last three layers, which are classification layers, have been replaced and trained with three new layers supporting two classes of polyps and non-polyps. The Activation\_40\_relu layer, which is the fortieth relu layer, was used as the feature extraction layer. Stochastic gradient descent with momentum (SGDM) optimizer was used in network training. Max Epochs 10, Mini Batch Size 2, Learning Rate was taken as 0.001. The Negative Overlap Range was taken as 0-0.3, while the Positive Overlap Range was accepted as 0.6-1.

Faster R-CNN detected polyp regions were cropped and resized respectively. Deep features of lumen, adenoma, hyperplastic and serrated polyp images obtained from crop and resize procedures were extracted. Features are classified with SVM with cubic kernel function. The performance of the deep feature extraction process does not depend on the amount of data, and this method does not require training to remove the feature.

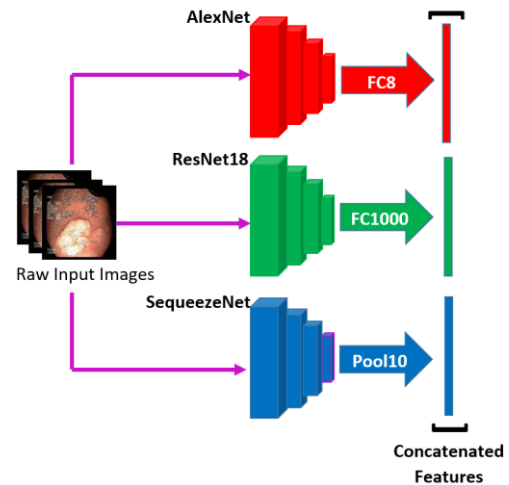
### 3.2 Deep feature extraction and combination with transfer learning

Transfer learning is often used when the data set to be used is not large enough. Transfer Learning can be done by transferring all layers of a CNN architecture, or by transferring (freezing) some layers and retraining some layers. Deep feature extraction is based on obtaining the features of an

image passed through a pretrained CNN architecture. It is not dependent on the amount of data and saves time as it does not require training.

In this study, the most widely used CNN architectures of deep neural networks in the classification and detection of polyps in images that make up the colonoscopy video were used with transfer learning. Images were passed through these pre-trained architectures with a large database, such as the ImageNet database, and their features were extracted.

The features of colonoscopy images in the study were AlexNet [31], SqueezeNet [32], GoogleNet [33], ResNet18 [34], MobileNetV2 [35], Vgg16 [36], inceptionV3 [37], DenseNet201 [38], ResNet50 [39], ResNet101 [40] is derived from the last fully connected layer of CNN architectures such as inceptionResNetV2 [41]. Deep feature extraction and feature fusion with the proposed method is shown in the diagram in Figure 4.



**Figure 4.** Deep features extraction and feature fusion

CNN architectures consist of layers. While the first layers obtain the low-level features of the image such as color, texture, shape, the last layers extract the high-level features of the image as a whole. Therefore, CNN architectures with different architectures and different number of layers extract different features. For this purpose, CNN architectures are combined to take advantage of different features. In experimental studies, many combinations have been tried to obtain a feature vector with high representation power and short generation time. At the end of the studies, 1000 features were obtained from AlexNet's FC8 layer, Resnet18's FC1000 layer and SqueezeNet's pool10 layer, and then a feature vector with 3000 features was obtained by combining them.

### 3.3 DWT as feature extraction and reduction of feature vector dimension

The wavelet transform was proposed by Grossmann and Morlet [42]. The wavelet analysis of CNN feature vectors can be said to be the sum of wavelets in variable time shifts and scales. DWT extracts native properties by separating components of feature vectors into both time and scale. One-dimensional DWT produces two sets of coefficients: approximation coefficients (cA) and detail coefficients (cD). cA is the most important part of the signal for many signals. It defines the cA identity of the mark. On the other hand, descriptive information is also available in the cD. For

example, when cD is removed from a human audio signal, the sound changes, but remains understandable. However, if the cA information is removed, the sound becomes incomprehensible. Thus, cA is used in DWT analysis and sub-samples the signal. Figure 5 shows the filtration in the DWT.

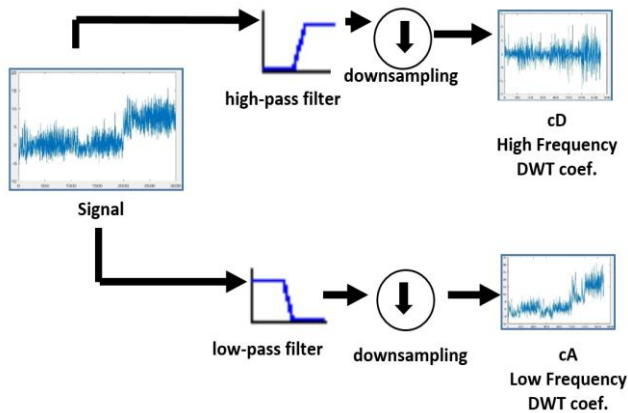


Figure 5. Diagram of DWT

In this study, as mentioned above,  $1 \times 3000$  dimension vector is obtained by fusing the features of different CNN architectures. Daubechies basis 7 (db7) wavelet DWT was applied to this feature vector once. As a result of this process cA and cD vectors with 1507 data were obtained. The cA vector containing important information is given to the SVM classifier as a feature. This process reduced the  $1 \times 3000$  CNNs feature vector to  $1 \times 1507$ . This improved the classifier performance. The reason for using Db7 wavelet analysis is that Daubechies wavelets may have a wide range of problems, such as self-similarity characteristics of a signal or fractal problems, signal discontinuities, and so on. Figure 6 shows the signal and image representation of the fused feature vector of the CNNs belonging to three classes. The image and signal representations of the cA vectors obtained after applying DWT to these image and signal representations are also shown in the figure.

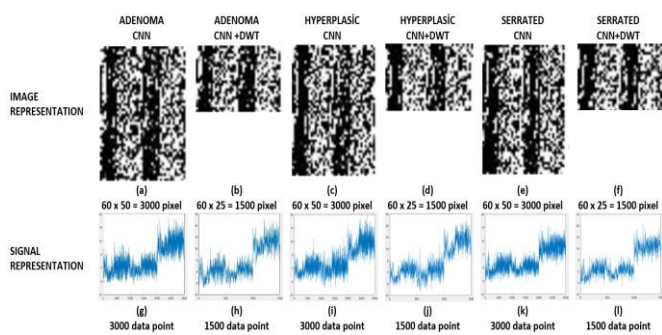


Figure 6. Signal and image representation of the fused feature vector of CNNs belonging to three classes and image and signal representations of the cA vectors obtained after applying DWT to the fused feature vector of CNNs

As shown in Figure 6, DWT has well sub-sampled and reduced vector size of the fused feature vector of CNNs for each class. The white dots in the images indicate discontinuities in the signals. In addition to the dimension reduction task, DWT was used to detect these discontinuities and to enhance the feature vector. In addition, DWT has

reduced the number of features by half, saving approximately 55% time in the classification period. DWT has shown higher performance in terms of classification performance with different feature selection techniques such as ReliefF [43] and Correlation-based feature selection [44]. This comparison table is presented in experimental studies.

### 3.4 Support vector machine (SVM) classifier

SVM is a classification method based on statistical learning. It was developed by Vapnik [45]. It aims to separate classes by specifying a separating boundary between classes on the plane. The learning data closest to Hyperplane is called a support vector. It is used in many fields such as image processing and sound processing with high classification performance.

There are many SVM core functions. Cubic kernel function was used in this study. The polyp region candidates identified by Faster R-CNN were removed from the video frame, their features were extracted with Pretrained CNN architectures, sub-sampled with DWT and classified with SVM. Region candidates were grouped into 4 classes. Training and test procedures were performed with 400 images from each class. Figure 7 illustrates the SVM classifier symbolically.

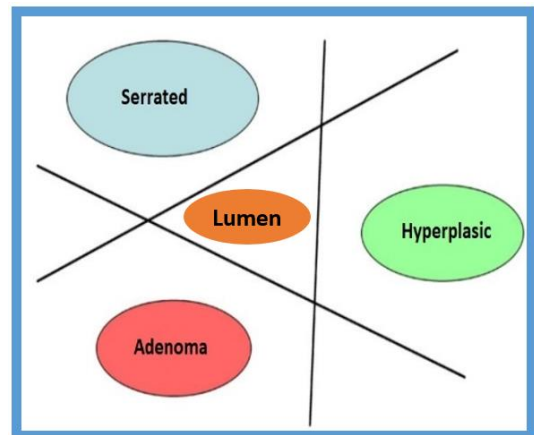


Figure 7. Symbolic representation of the SVM classifier

### 3.5 Hungarian algorithm

It is the method proposed by Hungarian researchers for the solution of the assignment problem [46]. Also known as the Kuhn-Munkres algorithm. It is widely used in object tracking algorithms to assign object detections to tracks. The Hungarian algorithm creates a cost matrix based on the previously calculated affinity score of detected objects. It associates objects from frame to frame with this cost matrix. The score used to generate the cost matrix can be the intersection over Union (IoU) value of the detected object and the predicted object. It can be also CNN features or shape features. These scores, which are commonly used in object tracking, are summarized below.

**IOU (Intersection Over Union) Score:** The overlap ratio of the bounding boxes.

**Shape Score:** It is the similarity ratio of shapes or sizes of objects.

**Convolution Cost Score:** The vectors of the CNN attributes of the object bounding boxes are obtained and compared. If there is a similarity as a result of the comparison, it is decided that the objects are the same.

In Figure 8, a tracking example of objects 1, 2 and 3 is given in the t0 frame and the t1 frame. A new detection (4) occurred at time t1 and the detection of object 3 was kept in memory. According to the example in Figure 8, the Faster R-CNN algorithm will detect objects in t0 and t1 frames. These object detections create track list t0 and detect list t1. Detected images were passed from pre-trained AlexNet, ResNet18, SqueezeNet CNN architecture without any training. Then, the features of the image were taken from the last fully connected layer of these architectures. An ARS feature vector with 3000 features was created by fusing 1000 features obtained from each CNN architecture. DWT is applied to the feature vector to strengthen the feature vector and reduce its size. Thus, the ARS+DWT feature vector was obtained. ARS+DWT feature vectors of images detected in consecutive frames were created and cosine similarity scores of these feature vectors were calculated. Detections with a cosine similarity score greater than 0.7 were matched. When the cosine similarity scores of the feature vectors are transferred to the table, an association problem that can be solved with the Hungarian algorithm arises. Figure 9 shows the pseudocode of the Hungarian algorithm used in this study.

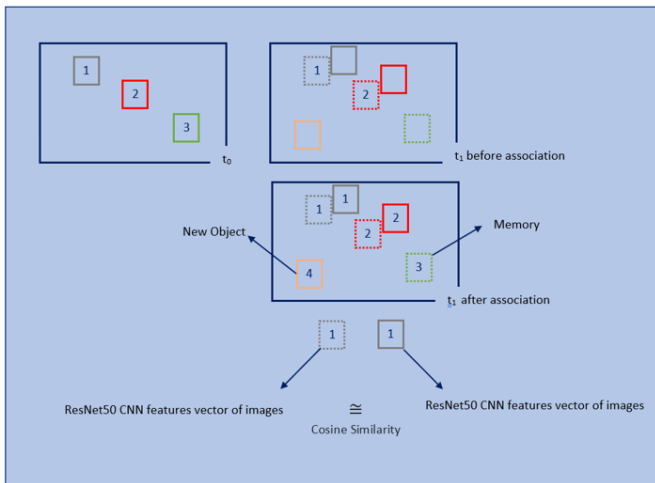


Figure 8. Hungarian algorithm assignment phase

```

FUNCTION assign_detections_to_trackers (trackers, detections, cosSim_thrd = 0.7)
SET cosSim_mat= zeros((len(trackers), len(detections)))

WHILE t, trk in enumerate(trackers)
WHILE d, det in enumerate(detections)
cosSim_mat[t, d] = box_cosSim2(trk, det)
END WHILE
END WHILE

SET matched_idx = FUNCTION linear_assignment (cosSim_mat)
SET unmatched_trackers = []
SET unmatched_detections = []

WHILE t, trk in enumerate(trackers)
IF t not in matched_idx[:0] then
unmatched_trackers.append(t)
END IF
END WHILE

WHILE d, det in enumerate(detections)
IF d not in matched_idx [1] then
unmatched_detections.append(d)
END IF
END WHILE

SET matches = []

WHILE m in in matched_idx
IF cosSim_mat [m [0], m [1]] <cosSim_thrd
unmatched_trackers.append (m [0])
unmatched_detections.append (m [1])
ELSE
matches.append(m.reshape(1,2))
END IF
END WHILE

IF len(matches)==0 THEN
matches = empty (0,2)
ELSE
matches = concatenate (matches, axis=0)
END IF
RETURN matches, unmatched_detections, unmatched_trackers
END FUNCTION

```

Figure 9. Pseudocode of the Hungarian algorithm used in this study

### 3.5.1 Cosine similarity

The relationship between the two vectors is expressed at an angle. While the cosine similarity of the two identical vectors is close to 1, the cosine similarity of the two opposing vectors will be 0. Figure 10 shows the angle difference between the two vectors. In Eq. (1), cosine similarity ratio is given.

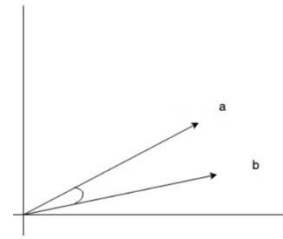


Figure 10. Cosine Similarity

Cosine similarity equality is given in equality 5.1.

$$\text{Cosine Similarity} = \frac{a \cdot b}{\|a\| \|b\|} = \frac{\sum_1^n a_i b_i}{\sqrt{\sum_1^n a_i^2} \sqrt{\sum_1^n b_i^2}} \quad (1)$$

### 3.6 Kalman filter

The Kalman filter is a filter that can predict the future status of input and output information along with the previous information of the system [47]. In addition, the Kalman filter is capable of estimating the unmeasurable values of the system. Kalman filter is an algorithm used to predict current and future situations in object tracking. A Kalman filter is used in each bounding box for object tracking. After the Hungarian algorithm correlates the determinations with the tracks, the Kalman filter's prediction and update functions are called up. These functions determine the state average and covariance. If the mean and covariance values of the randomly distributed random variable (Gaussian) are known, the probability of any range can be known. The Kalman filter is based on this principle.

## 4. EXPERIMENTAL RESULTS AND DISCUSSION

### 4.1 Experimental design and environment

The ARS-DWT-SVM algorithms were implemented on the MATLAB R2022b in a notebook with Intel Core i7-4510U processor, 16 GB RAM and Windows 10 operating system. Average values are given in time measurements.

### 4.2 Dataset

In this study, colonoscopy dataset, which is a public dataset, was used. Training and test data were obtained from this data set classified by experts. In the data set, 76 videos are presented in both White Light (WL) video and Narrow-Band Imaging (NBI) video format. NBI video format was used in the study. The data set contains 15 serrated, 21 hyperplastic and 39 adenoma polyps. These videos contain 20948 adenomas, 7423 hyperplastic and 5902 serrated polyp images, which differ from each other in different angles. For image classification process, a total of 1329 images were taken from 443 each class in a different view. Detailed information about the data set is given in the article in which the data set was

announced [48]. For the image classification and detection models proposed in this study, k-fold cross validation method was preferred for dividing the data sets as training and test data. In this study, 5 fold cross validation methods were used.

### 4.3 Evaluation metrics

This study was evaluated with three different metrics: object detection, classification of images detected object and object tracking.

#### 4.3.1 Metrics for object detection

The metrics used in the MICCAI 2015 Endoscopic Vision Challenge [34] were used as evaluation criteria for object detection.

The metrics used and their descriptions are listed below:

- True Positive (TP): If the center of the box that predicts the boundaries of the object is within the ground-truth boxes, the detection is considered a TP. If there are two estimates that cross with ground-truth boxes, one TP is counted.
- False Positive (FP): Any detection that falls outside the ground-truth boxes is considered an FP. The image contains polyps, which have been detected but incorrect estimates have been made. FP is the number of incorrect estimates.
- True Negative (TN): The number of cases in which there is no predicted box on a non-polyp image is considered TN.
- False Negative (FN): Number of undetected polyps. The image contains polyps, but this is the number of no guesses.

The following metrics are calculated from TP, FP, TN, FN:

$$\text{Accuracy (Acc.)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Specificity (Spec.)} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Precision (Prec.)} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Sensitivity or Recall Rate(Rec.)} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1 Score} = \frac{2 * \text{Prec} * \text{Rec}}{\text{Prec} + \text{Rec}} \quad (6)$$

$$\text{F2 Score} = \frac{5 * \text{Prec} * \text{Rec}}{4 * \text{Prec} + \text{Rec}} \quad (7)$$

#### 4.3.2 Metrics used in detaching and classification of detected polyps from image

In this study, to evaluate the classification performance; In relation to sensitivity, Precision, Specificity and Accuracy. True Positives –TP, False Positives - FP, True Negatives - TN and False Negative - F are described as follows:

- True Positive (TP): A correctly classified image is considered a TP.
- False Positive (FP): An image that is not correctly classified is considered FP.
- True Negative (TN): The classifier estimates that the image class is not X, but actually represents the

number of evaluations TN, which is not the image class X.

- False Negative (FN): The classifier estimated the image class not X, in fact the image class represents the number of evaluations in the form of X.

#### 4.3.3 Metrics used in evaluating the object tracking

CLEAR MOT (Multi-Object Tracking) metrics have been used for multiple object tracking [49]. These metrics are MOTA, Multi Object Tracking Accuracy (MOTA). These metrics are a summary of the other metrics that create them. Below is a list of these metrics.

- False Positive (FP): Number of detections in the entire video that fall outside the object area.
- False Negative (FN): Number of undetectable objects in the entire video.
- Fragmentation (Frag.): Number of object traces interrupted in the entire video.
- ID Change (IDSW): Number of incorrect ID number that changes wrongly over the entire video.

### 4.4 Experiment results

The proposed method consists of three steps. The first of these steps is the detection of polyps in colonoscopy videos. Faster R-CNN, SSD, YOLOV3, one of the deep learning algorithms for region detection, has been used in the literature for polyp detection. In the studies, Faster R-CNN showed higher success in polyp detection compared to YOLOV3 and SSD like this study. For this reason, Faster R-CNN was used as object detector with ResNet-50 CNN architecture. Due to the small number of data, transfer learning has been utilized in Faster R-CNN training. Only the last three layers of ResNet-50 CNN architectures have been trained.

**Table 1.** Polyp detection performance table of algorithms after 20 epoch training

Object Detector	Pre.(%)	Rec.(%)	F1(%)	F2(%)
FASTER R_CNN (ResNet50)	92.6	83.5	87.8	85.2
SSD (ResNet50)	77.4	82.0	79.6	81.0
YOLO V3 (ResNet50)	81.3	59.2	68.5	62.6

As seen in Table 1, Faster R-CNN architecture showed higher performance under SSD and YOLO algorithms under 20 epoch training and other equal conditions in polyp detection. For this reason, Faster R-CNN object detection algorithm is used in object detection step of the application.

Table 2 shows the polyp detection performance of Faster R-CNN object detection algorithms with different CNN architectures.

**Table 2.** Polyp detection performance of faster R-CNN object detection algorithms with different CNN architectures

Faster R-CNN Object Detector	Pre.(%)	Rec.(%)	F1(%)	F2(%)
<b>Different Backbone</b>				
<b>FASTER R_CNN (ResNet50)</b>	<b>92.6</b>	<b>83.5</b>	<b>87.8</b>	<b>85.2</b>
FASTER R_CNN (Resnet101)	92.2	80.8	86.1	82.8
FASTER R_CNN (Inception ResNet)	91.1	77.3	83.7	79.8

As seen in Table 2, Faster R-CNN object detection algorithms with different CNN architectures are compared. ResNet50 CNN architecture showed the highest performance in object detection. After the detection step of the polyp regions, the detected polyp regions were cropped from the image and resized to the first layer of CNN architectures. To reduce the amount of FP in polyp detection, a lumen class was established. Thus, by using the object detection method, a data set consisting of 400 images from each of adenoma, hyperplastic, lumen, serrated classes and a total of 1600 images was created. The data set was passed through pre-trained CNN architectures and classified with SVM by extracting features from the last layers of the architectures. In Table 3, the classification performance and feature extraction times of the SVM classifiers of the features obtained from the pre-trained different CNN architectures are given.

**Table 3.** Feature extraction times of CNNs and classification performance of features extracted from different CNN architectures

Pretrained CNN Model	Av. Acc. (%)	Ad. Acc. (%)	Hyp. Acc. (%)	Lum. Acc. (%)	Ser. Acc. (%)	Time Cons. (sec.)
AlexNet	91.7	91	95	95	86	0.02
DenseNet201	94.3	95	97	96	89	0.42
GoogleNet	78.2	70	87	87	72	0.07
InceptionRes Netv2	93.9	93	98	96	89	0.69
Inceptionv3	93.3	92	97	96	89	0.27
Mobilenetv2	93.2	92	97	96	89	0.09
<b>ResNet101</b>	<b>94.6</b>	<b>95</b>	<b>98</b>	<b>96</b>	<b>90</b>	<b>0.32</b>
ResNet18	93.1	93	96	96	88	0.07
ResNet50	94.5	95	98	96	89	0.18
SqueezeNet	92.8	92	96	96	88	0.03
VGG16	92.4	92	96	95	88	0.33

**Table 4.** Feature extraction times of fusing CNNs and classification performance of fusing CNN features

Combined Pre-trained CNN Architectures Features	Av. Acc. (%)	Ad. Acc. (%)	Hyp. Acc. (%)	Lum. Acc. (%)	Ser. Acc. (%)	Time Cons. (sec.)
AlexNet						0.026
ResNet18						0.072
SqueezeNet	<b>94.7</b>	95	98	96	90	0.038
<b>+</b>						<b>+</b>
ARS						<b>0.136</b>
DenseNet						0.426
Resnet18						0.072
SqueezeNet	95.0	96	98	97	90	0.038
<b>+</b>						<b>+</b>
DRS						0.536
DenseNet						0.426
Resnet101						0.321
Resnet50	95.3	96	99	97	90	0.189
<b>+</b>						<b>+</b>
DRR						0.936

As seen in Table 3, pre-trained ResNet101 CNN architecture showed the highest classification performance. However, the feature extraction time of Resnet101 architecture is quite high. In order to overcome this situation, it is aimed to fusion different CNN features and shorten the feature extraction time without losing the classification

performance. Table 4 shows the classification performance with SVM classifier and feature extraction times of the feature vectors obtained by fusing different architectures.

As seen in Table 4, the feature vector obtained by the fusion of AlexNet, ResNet18 and SqueezeNet (ARS) architectures achieved 94.6% classification success and caught ResNet101, the highest classification performance shown in Table 1, and saved about 55% of the feature extraction time. Although the classification success of DenseNet, ResNet 18, SqueezeNet (DRS) architectures and DenseNet, ResNet 101, ResNet 50 (DRR) architectures is higher, the cost of computation times has increased. In this study, a fusion of AlexNet, ResNet 18 and SqueezeNet architectures was used in order to achieve a fast result.

Another method frequently used in the literature to increase the amount of performance is feature selection or feature reduction algorithms. In this study, DWT was proposed as a feature reduction method. Table 5 shows the classification performance of different feature selection algorithms in the SVM classifier and time measurements for a feature vector.

**Table 5.** Comparison of classification accuracy and feature extraction times of Feature selection algorithms applied to feature vector

Feature Selection Method	A. Acc. (%)	Ad. Acc. (%)	Hyp. Acc. (%)	Lum. Acc. (%)	Ser. Acc. (%)	Time Cons. (sec.)
DWT	95.4	96	98	96	90	0.00
Relieff [7]	94.8	95	98	96	89	0.03
Genetic Algorithm [50]	94.7	96	98	96	89	0.03
Correlation-based [51]	94.6	95	97	96	89	0.28
Differential evolution Algorithm [50]	94.7	96	98	96	89	0.03
PCA [50]	93.6	93	97	96	89	0.11
LDA [50]	94.7	96	98	96	89	0.11
PCC (Pearson Correlation Coefficient) 1500/3000 [51]	94.4	95	98	96	89	0.03
F-score 1500/3000 [52]	94.6	95	97	96	89	0.03

As seen in Table 5, DWT increased the classification performance of ARS CNN architectures from 94.8 to 95.4. At the same time, the implementation time of DWT architecture is shorter than other methods. Figure 11 shows the complexity matrix and the ROC curve of the proposed method.

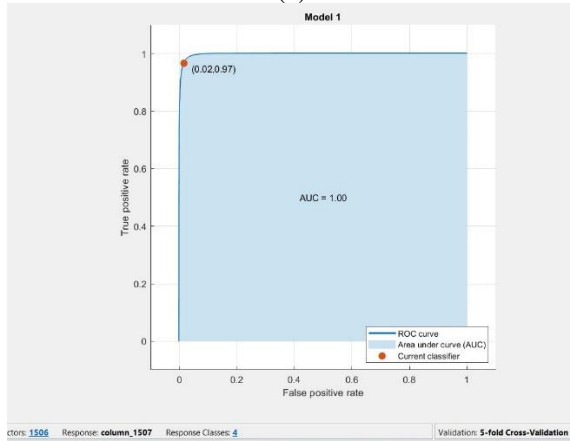
Figure 12 illustrates graphically the mean accuracy values of ARS + DWT feature vectors obtained by applying DWT to ARS feature vector and ARS vector consisting of the combination of ResNet101 feature vector and AlexNet, SqueezeNet, ResNet18 architectures and SVM classifier.

As seen in Figure 12b, the process of fusion features yielded a 55% gain in time. It was also found that feature fusion classifier increases the average accuracy rate. DWT, on the other hand, worked faster than all other feature size reducers and increased the classifier mean accuracy.



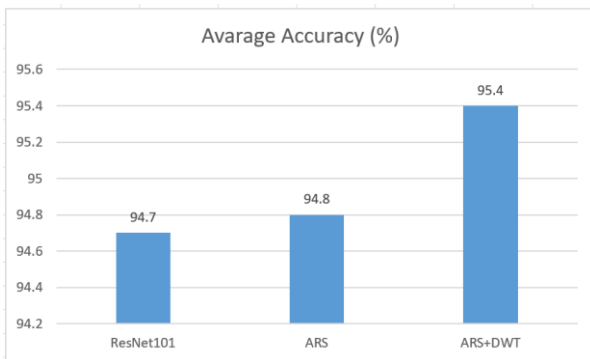


(a)

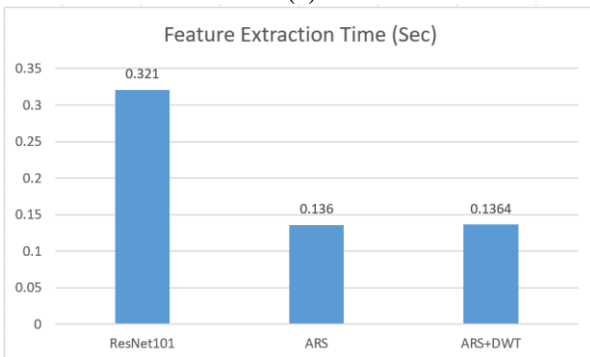


(b)

**Figure 11.** Complexity matrix (a) and ROC curve (b) obtained by SVM Classification of the proposed ARS + DWT method



(a)



(b)

**Figure 12.** Comparison of feature vectors in terms of mean accuracy (a) and total feature extraction times

In addition, the ARS + DWT feature vector was used in object tracking. In the object detection step of the Hungarian algorithm, cosine similarity ratios of the traces and detections were used in the cost matrix stage. This increased the performance of the object tracking algorithm. Table 6 shows the tracking results of the proposed method.

**Table 6.** Tracking performance of proposed method

	FN ↓	FP ↓	Frag. ↓	IDsw ↓	MOTA ↑
<b>Adenoma</b>	13	0	7	7	0.96
<b>Hyperplastic</b>	17	1	9	8	0.95
<b>Serrated</b>	20	3	10	10	0.94

As shown in Table 6, the proposed method resulted in a decrease in FP. This decline is related to the strategy developed to reduce the amount of FP. Using the cosine similarity ratio of the ARS + DWT feature vector in the cost matrix resulted in a decrease in the IDsw ratio. Developed strategy and proposed method increased object tracking performance.

In order to see the effect of the proposed method on object tracking, the metrics were recalculated without cropping and classifying the determinations obtained as a result of not polyp detection, and the ratio IoU as known as Jaccard index between cost matrix detection and track. The results obtained are shown in Table 7.

**Table 7.** Tracking performance metrics without proposed method

	FN ↓	FP ↓	Frag. ↓	IDsw ↓	MOTA ↑
<b>Adenoma</b>	13	5	7	10	0.95
<b>Hyperplastic</b>	17	7	9	13	0.93
<b>Serrated</b>	20	9	10	14	0.92

Table 7 shows object tracking metrics obtained when the strategy that decreases the number of FNs is abandoned and the cost matrix is generated with IoU.

## 5. CONCLUSIONS

There are two important problems in image processing using deep learning algorithms. One of these problems is that the cost of calculation is high due to the fact that the transactions performed are in the order of pixels, thus it requires high processing performance hardware. The second important problem is that deep learning algorithms need millions of parameter updates during the training phase. This increased the need for big data to train deep learning algorithms. When it comes to medical images, it is very difficult to find labeled data in medical images. Labeling images takes a lot of time for medical professionals and requires a large number of expert opinions to overcome human error. Another difficulty is the high number of rare images in the medical field. This leads to imbalance between data classes. Failure to balance the data causes the image not to be adequately represented in the data set. These problems make it difficult to use high performance algorithms such as deep learning in some areas of medical imaging. In this study, it worked with polyp images which are examples of the limitations mentioned above. In the beginning, polyp, non-

polyp classification was made with Faster R-CNN object detection algorithm so that polyp images could not be divided into classes. Due to the small number of data, transfer learning has been utilized. Only the last three layers are trained in the CNN architecture used in the Faster R-CNN. The other layers were transferred. Faster R-CNN object detection algorithm with ResNet-50 CNN architecture was able to detect polyp regions with 92.6% precision rate. Subsequently, a database was created by cropping the polyp sections from the image to reduce the FP number and classify the polyp images in the regions proposed as polyps. The data set consisted of 400 adenomas, hyperplastic, serrated and lumen images. The features of the images of this database were extracted with popular CNN architectures and classified with SVM with cubic kernel function. ResNet-101 showed the best performance with 94.7% average classification accuracy. In order to save the feature extraction time of ResNet-101 CNN architecture, AlexNet, ResNet18, SqueezeNet features were fused to form a vector called ARS. The ARS vector achieved the average accuracy of ResNet101 and saved about 55% in feature removal time. In order to increase the classification performance of the ARS vector and to perform operations with lower features in the tracking stage, the feature size reduction methods were applied to the ARS vector. In experimental studies, it was found that applying DWT to ARS vector decreased the number of features by 50% and increased the classification average accuracy value from 94.8% to 95.4%. In addition, applying DWT to the ARS vector not only contributed to the classification phase but also contributed to the tracking step by increasing the affinity rate. The ARS + DWT vector was used in the cost matrix generation step, which is the first step of the Hungarian algorithm. The similarity score between tracks and detections was established using cosine similarity coefficients of ARS + DWT features vector of tracks and detections. This resulted in a 30% reduction in the IDsw score. This decrease caused an increase in the MOTA score.

## REFERENCES

- [1] Atkin, W.S., Saunders, B.P. (2002). Surveillance guidelines after removal of colorectal adenomatous polyps. *Gut*, 51(suppl 5): v6-v9. [https://doi.org/10.1136/gut.51.suppl\\_5.v6](https://doi.org/10.1136/gut.51.suppl_5.v6)
- [2] Azer, S.A. (2019). Challenges facing the detection of colonic polyps: What can deep learning do?. *Medicina*, 55(8): 473. <https://doi.org/10.3390/medicina55080473>
- [3] Julka, M., Cherukuri, M., Lamah, R. (2011). Screening for cancerous and precancerous conditions of the colon. *Primary Care: Clinics in Office Practice*, 38(3): 449-468. <https://doi.org/10.1016/j.pop.2011.05.009>
- [4] IJspeert, J.E., Bastiaansen, B.A., Van Leerdam, M.E., et al. (2016). Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. *Gut*, 65(6): 963-970. <https://doi.org/10.1136/gutjnl-2014-308411>
- [5] Heresbach, D., Barrioz, T., Lapalus, M.G., et al. (2008). Miss rate for colorectal neoplastic polyps: A prospective multicenter study of back-to-back video colonoscopies. *Endoscopy*, 40(04): 284-290. <https://doi.org/10.1055/s-2007-995618>
- [6] Özyurt, F., Sert, E., Avci, D. (2020). An expert system for brain tumor detection: Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine. *Medical Hypotheses*, 134: 109433. <https://doi.org/10.1016/j.mehy.2019.109433>
- [7] Özyurt, F. (2020). Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. *The Journal of Supercomputing*, 76(11): 8413-8431. <https://doi.org/10.1007/s11227-019-03106-y>
- [8] Kutlu, H., Avci, E. (2019). A novel method for classifying liver and brain tumors using convolutional neural networks, discrete wavelet transform and long short-term memory networks. *Sensors*, 19(9): 1992. <https://doi.org/10.3390/s19091992>
- [9] Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- [10] Wang, X.H., Istepanian, R.S., Song, Y.H. (2003). Application of wavelet modulus maxima in microarray spots recognition. *IEEE Transactions on Nanobioscience*, 2(4): 190-192. <https://doi.org/10.1109/TNB.2003.816230>
- [11] Wang, J., Ma, J.Z., Li, M.D. (2004). Normalization of cDNA microarray data using wavelet regressions. *Combinatorial Chemistry & High Throughput Screening*, 7(8): 783-791. <https://doi.org/10.2174/1386207043328274>
- [12] Bennet, J., Arul Ganaprakasam, C., Arputharaj, K. (2014). A discrete wavelet based feature extraction and hybrid classification technique for microarray data analysis. *The Scientific World Journal*, 2014: 1-9. <https://doi.org/10.1155/2014/195470>
- [13] Phinyomark, A., Nuidod, A., Phukpattaranont, P., Limsakul, C. (2012). Feature extraction and reduction of wavelet transform coefficients for EMG pattern classification. *Elektronika ir Elektrotechnika*, 122(6): 27-32. <https://doi.org/10.5755/j01.eee.122.6.1816>
- [14] Sun, G., Dong, X., Xu, G. (2006). Tumor tissue identification based on gene expression data using DWT feature extraction and PNN classifier. *Neurocomputing*, 69(4-6): 387-402.
- [15] Özyurt, F., Sert, E., Avci, E., Dogantekin, E. (2019). Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy. *Measurement*, 147: 106830. <https://doi.org/10.1016/j.measurement.2019.07.058>
- [16] Özyurt, F., Tuncer, T., Avci, E., Koç, M., Serhatlioğlu, İ. (2019). A novel liver image classification method using perceptual hash-based convolutional neural network. *Arabian Journal for Science and Engineering*, 44(4): 3173-3182. <https://doi.org/10.1007/s13369-018-3454-1>
- [17] Byrne, M.F., Chapados, N., Soudan, F., Oertel, C., Pérez, M.L., Kelly, R., Iqbal, N., Chandelier, F., Rex, D.K. (2019). Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*, 68(1): 94-100. <https://doi.org/10.1136/gutjnl-2017-314547>
- [18] Mo, X., Tao, K., Wang, Q., Wang, G. (2018). An efficient approach for polyps detection in endoscopic videos based on faster R-CNN. In 2018 24th International Conference on Pattern Recognition (ICPR), pp. 3929-3934. <http://arxiv.org/abs/1809.01263>
- [19] Shin, Y., Qadir, H.A., Aabakken, L., Bergsland, J., Balasingham, I. (2018). Automatic colon polyp detection

- using region based deep CNN and post learning approaches. *IEEE Access*, 6: 40950-40962. <https://doi.org/10.1109/ACCESS.2018.2856402>
- [20] Wang, D., Zhang, N., Sun, X., Zhang, P., Zhang, C., Cao, Y., Liu, B. (2019). Afp-net: Realtime anchor-free polyp detection in colonoscopy. In 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 636-643. <http://arxiv.org/abs/1909.02477>
- [21] Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W., Baldi, P. (2018). Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*, 155(4): 1069-1078. <https://doi.org/10.1053/j.gastro.2018.06.037>
- [22] Wang, P., Berzin, T.M., Brown, J.R.G., Bharadwaj, S., Becq, A., Xiao, X., Liu, P., Li, L., Song, Y., Zhang, D., Li, Y., Xu, G., Tu, M., Liu, X. (2019). Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*, 68(10): 1813-1819. <https://doi.org/10.1136/gutjnl-2018-317500>
- [23] Qadir, H.A., Shin, Y., Solhusvik, J., Bergsland, J., Aabakken, L., Balasingham, I. (2019). Polyp detection and segmentation using mask R-CNN: Does a deeper feature extractor CNN always perform better?. In 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT), pp. 1-6. <https://doi.org/10.1109/ISMICT.2019.8743694>
- [24] Zhang, R., Zheng, Y., Poon, C.C., Shen, D., Lau, J.Y. (2018). Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognition*, 83: 209-219. <https://doi.org/10.1016/j.patcog.2018.05.026>
- [25] Raghu, M., Zhang, C., Kleinberg, J., Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32. <http://arxiv.org/abs/1902.07208>
- [26] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. *Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.1311.2524>
- [27] Girshick, R. (2015). Fast R-CNN. <https://doi.org/10.48550/arXiv.1504.08083>
- [28] Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. <https://doi.org/10.48550/arXiv.1506.01497>
- [29] Redmon, J., Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263-7271. <http://arxiv.org/abs/1612.08242>
- [30] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016). Ssd: Single shot multibox detector. In European Conference on Computer Vision, pp. 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [31] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks.
- [32] Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. <https://doi.org/10.48550/arXiv.1602.07360>
- [33] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9. <https://doi.org/10.48550/arXiv.1409.4842>
- [34] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778. <https://doi.org/10.3390/app12188972>
- [35] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510-4520.
- [36] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- [37] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818-2826.
- [38] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708.
- [39] Theckedath, D., & Sedamkar, R. R. (2020). Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Computer Science*, 1: 1-7. <https://doi.org/10.1007/s42979-020-0114-9>
- [40] Sun, C., Shrivastava, A., Singh, S., Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In Proceedings of the IEEE International Conference on Computer Vision, pp. 843-852. <https://doi.org/10.1109/ICCV.2017.97>
- [41] Ferreira, C.A., Melo, T., Sousa, P., Meyer, M.I., Shakibapour, E., Costa, P., Campilho, A. (2018). Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15 (pp. 763-770). Springer International Publishing. [https://doi.org/10.1007/978-3-319-93000-8\\_86](https://doi.org/10.1007/978-3-319-93000-8_86)
- [42] Grossmann, A., Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4): 723-736. <https://doi.org/10.1137/0515056>
- [43] Tuncer, T., Dogan, S., Ozyurt, F. (2020). An automated residual exemplar local binary pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image. *Chemometrics and Intelligent Laboratory Systems*, 203: 104054. <https://doi.org/10.1016/j.chemolab.2020.104054>
- [44] Wosiak, A., Zakrzewska, D. (2018). Integrating correlation-based feature selection and clustering for improved cardiovascular disease diagnosis. *Complexity*, 2018: 2520706. <https://doi.org/10.1155/2018/2520706>
- [45] Vapnik, V., Levin, E., Le Cun, Y. (1994). Measuring the VC-dimension of a learning machine. *Neural Computation*, 6(5): 851-876. <https://doi.org/10.1162/neco.1994.6.5.851>
- [46] Kuhn, H.W. (1955). The Hungarian method for the

- assignment problem. *Naval Research Logistics Quarterly*, 2(1-2): 83-97. [https://doi.org/10.1007/978-3-540-68279-0\\_2](https://doi.org/10.1007/978-3-540-68279-0_2)
- [47] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82: 35-45. <https://doi.org/10.1115/1.3662552>
- [48] Mesejo, P., Pizarro, D., Abergel, A., Rouquette, O., Beorchia, S., Poincloux, L., Bartoli, A. (2016). Computer-aided classification of gastrointestinal lesions in regular colonoscopy. *IEEE Transactions on Medical Imaging*, 35(9): 2051-2063. <https://doi.org/10.1109/tmi.2016.2547947>
- [49] Bernardin, K., Stiefelhagen, R. (2008) Evaluating multiple object tracking performance: The CLEAR MOT Metrics. *Eurasip J. Image Video Process*, 2008: 1-10. <https://doi.org/10.1155/2008/246309>
- [50] Too, J., Abdullah, A.R., Mohd Saad, N., Tee, W. (2019). EMG feature selection and classification using a Pbest-guide binary particle swarm optimization. *Computation*, 7(1): 12. <https://doi.org/10.3390/computation7010012>
- [51] Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., Vapnik, V. (2000). Feature selection for SVMs. *Advances in Neural Information Processing Systems*, 13: 668-674.
- [52] Song, Q., Jiang, H., Liu, J. (2017). Feature selection based on FDA and F-score for multi-class classification. *Expert Systems with Applications*, 81: 22-27. <https://doi.org/10.1016/j.eswa.2017.02.049>