

## A Pipeline for Computer Aided Polyp Detection and Segmentation in Ct Colonography

Janapriya.A.S<sup>1</sup>, Mythili.C<sup>2</sup>, Nishanthi.P<sup>3</sup>, Sunitha.K<sup>4</sup>

<sup>1</sup>(M.E.VLSI design, Kalaignar karunanidhi institute of technology/ Anna University, India)

<sup>2</sup>(M.E.VLSI design, Kalaignar karunanidhi institute of technology/ Anna University, India)

<sup>3</sup>(M.E.VLSI design, Kalaignar karunanidhi institute of technology/ Anna University, India)

<sup>4</sup>(M.E.VLSI design, Kalaignar karunanidhi institute of technology/ Anna University, India)

---

**Abstract:** CAD-CTC helps the radiologists for the automatic detection of suspicious polyps in the CT images of the colon. In this paper, a novel method is proposed in order to decrease the computational time taken by the CAD system. In the proposed CAD system, segmentation of the colon is done by Otsu's method of thresholding and clustered by k-means for extraction of the candidates. Classification is performed by SVM (Support Vector Machine) and ANN (Artificial Neural Network) from the training and testing images of candidate extraction. The proposed CAD system is evaluated with 16 slices (each) of normal and abnormal images. The CAD system achieved a minimal computation time (10 minutes) and an increased learning rate of about 1.2608 at an epoch 411. Thus the good results demonstrate that the proposed system may provide relevant additional information for the clinical decision process.

**Keywords:** ANN(Artificial Neural Network), CAD(Computer Aided Diagnosis), CTC(Computerized Tomographic Colonography), SVM(Support Vector Machine), Otsu's method.

---

### I. Introduction

CAD systems are used for the automatic detection of suspicious polyps in the CT images of the colon. There are different CAD systems developed by different authors, which were incorporated by making changes either in the segmentation phase or in the feature extraction phase for the identification of polyps.

Colorectal cancer is a cancer that occurs in the colon or rectum. It is the third most common form of cancer after lung and breast cancers. It affects mostly women and men above the age of 50. It is the second leading cause of cancer-related deaths in the United States (American Cancer Society, annual report)[1]. The progression from a cell to turn out to be cancerous will take 10 years. The guideline recommended the screening [2] of adults who are at average risk for colorectal cancer, since the detection and removal of adenomas has been shown to reduce the incidence of cancer and cancer-related mortality. Therefore, many have advocated screening programs to detect polyps[3]. Most of the people felt discomfort with the Virtual colonoscopy and inconvenience with the optical colonoscopy.

To encourage the people participating in screening test computed tomography was developed. Computerized Tomography Colonography (CTC) is a minimally invasive technique for the detection of colorectal polyps and carcinoma [2]. CTC involves acquisition and interpretation of original 2-D data. The data acquisition is fast and reliable; the interpretation of CTC data can be time-consuming and depends on human factors, such as experience, fatigue, and perception [1]. The most common methods used in identifying the presence of colorectal polyps in a clinical setting are the fecal occult blood test (FOBT), flexible sigmoidoscopy, barium enema, DNA stool test, and optical colonoscopy. The colonoscopy examination is performed by the gastroenterologists by using a flexible colonoscope to investigate the inner wall of the colon.

The colonoscopies fitted with a miniature camera and light unit to allow the gastroenterologist to evaluate any abnormal changes on the colon tissue. If abnormal growths on the colon are discovered, the gastroenterologist can use the channels of the colonoscope to preserve biopsies (that can be used for later examination) or even remove the colorectal polyps if this is necessary. Prior to the scheduled colonoscopy investigation, the patient is instructed to follow a strict diet for 24 h in order to perform bowel cleansing and typically the colonoscopy examination takes between 20–60 min while the patient is sedated. The main disadvantage associated with the optical colonoscopy is the invasive nature of the procedure, although the risk of bowel perforation by the colonoscope is small (0.005% of the total number of investigated cases). For 5% of the patients the colonoscope is not able to reach the caecum point, which results in an incomplete colon examination.

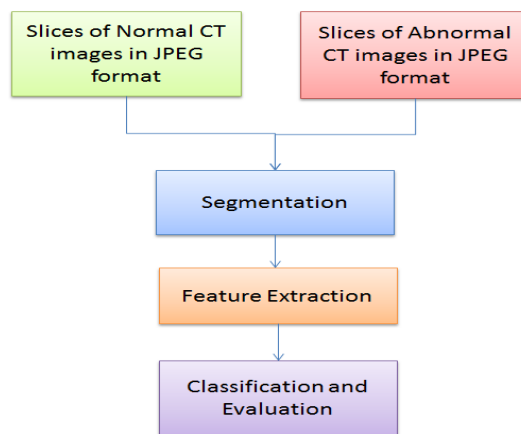
With the introduction of the multi slice CT scanners, the amount of data to be examined by the radiologists has increased dramatically and a dataset can consist of 800 to 2000 axial images depending on the patient height and the resolution of the CT data. Thus, the visual examination of such a large amount of data is a time consuming procedure, and the results are not always reproducible since they are often affected by eye

fatigue and the subjectivity of the human operator [4]. This has led to a sustained effort to develop computer-aided diagnosis (CAD) schemes that can help radiologists with locating colorectal lesions in CTC data in an efficient and accurate manner [1]. The aim of CAD systems were to automatically detect polyps and masses, and to provide the locations of suspicious regions of the colon. Radiologists can thus focus on a reduced number of small areas, while quickly surveying the larger portion of the colon [3].

### II. Overview Of The Proposed System

In this paper, the CAD pipeline consists of various operations such as segmentation, candidate extraction, classification and evaluation [4]. The operations are explained in detail. CT images consists of different slices of normal and abnormal images which are obtained from the datasets where they are in DICOM format which are converted into JPEG format and given as the input for the CAD system. The data set contains up to 700 slices of image. The thickness of the normal and abnormal CT dataset is 0.6mm. Initially the dataset will be in the DICOM format. Then the DICOM format image is converted to JPEG format for the convenience to process the CAD system.

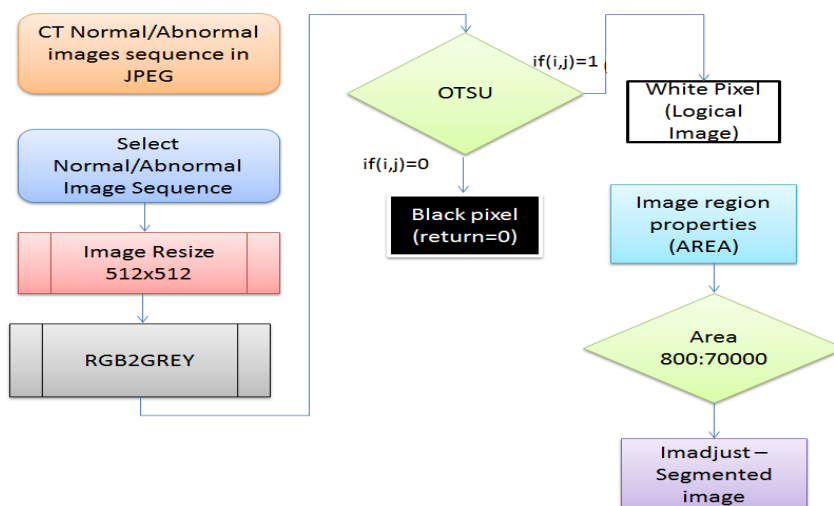
We are extracting the images in which the colon part clearly appears or visible. Colon has three different parts. They are transverse, ascending and descending colon.



**Fig.1:** Overview of Proposed System

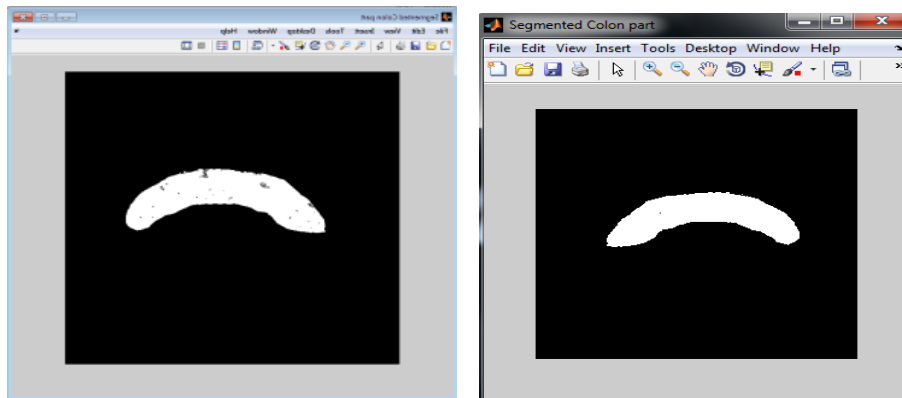
### III. Segmentation And Candidate Extraction

Segmentation is performed by using the Otsu’s method of thresholding. The flowchart of segmentation is shown Fig.1 and it explains the entire process of segmentation. The original images are resized to 512x512 then the images are converted into gray scale after which Otsu’s thresholding is performed. If  $im(i, j) = 0$  then it returns black else it returns white. Then binary components are connected. An area is chosen from the properties of the image (800:70000). Then the image is contrast adjusted and the segmented colon is obtained as shown.



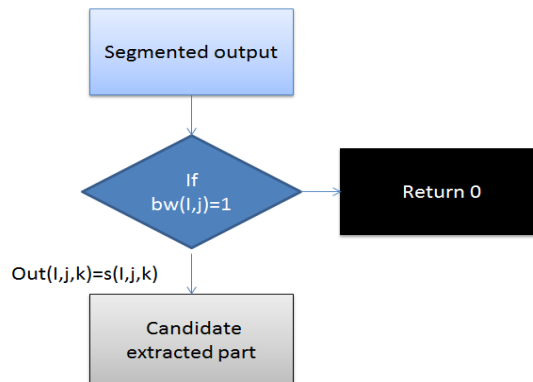
**Fig.2:** Flowchart for Segmentation

The segmented colons for normal and abnormal images are shown in the Fig.2 (a) and (b).



**Fig.3: (a)** Segmented colon of normal image **(b)** Segmented colon of abnormal image

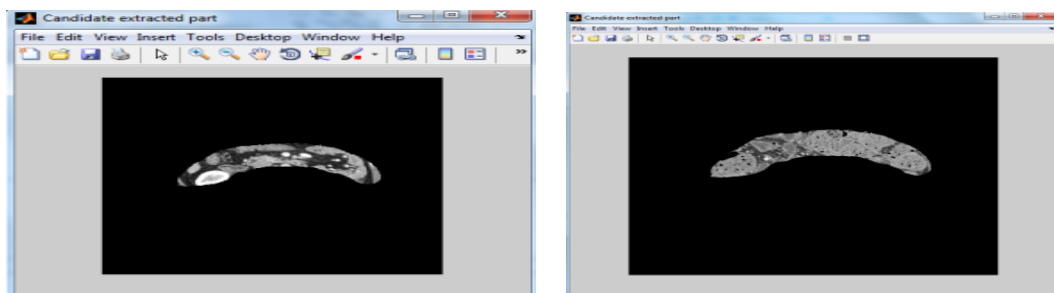
The segmented colon will be given as input to the feature extraction stage. The flow chart for feature extraction is shown in Fig.4. K-means clustering [4] is used because it is simple and has relatively low computational complexity. In addition, it is suitable for biomedical image segmentation as the number of clusters is usually known for images of particular regions of human anatomy.



**Fig.4:** Flowchart for Candidate Extraction

If the data point is not closest to its own cluster, it will have to be shifted into the closest cluster. If the data point is already closest to its own cluster, we will not shift it. The process continues until cluster means do not shift more than a given cut-off value or the iteration limit is reached.

In hierarchical iteration, on each level, the spatial constrained region growing is carried out, whose results are used as a feed forwarded to the next level of k-mean algorithm. By this algorithm we handle over segmentation by combining color similarity with edge information, which is fast and generic. Segmentation of medical image using k-mean clustering and marker controlled watershed algorithm, uses the conservative watershed algorithm for medical image analysis which does the entire division of the image. But its disadvantages are the over segmentation and sensitivity to false edges. For that here k-mean algorithm is used before using the watershed algorithm, to reduce the number of false edges and over-segmentation. The candidate extracted images of both normal and abnormal are shown in the Fig.5.



**Fig. 5:** Candidate extracted colon for normal and abnormal images

#### IV. Feature Extraction

Feature extraction is to detect and isolate desired portions or shapes of an image. Colon lesion can be classified to their size, measured in diameter, according to their morphology, into pedunculated, sessile or flat.

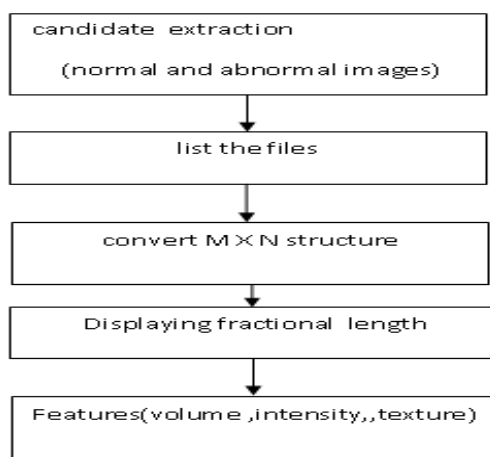
When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant, then the input data will be the input data into set of features is called feature extraction. The quantitative features are used for detecting the normal and abnormal condition of the colon.

An important task in CTC CAD is to extract features that characterize the colonic anatomic structures, such as polyps, haustral folds, and colonic walls in the CTC image data. The structural features popularly used in CTC CAD include principal curvatures of surface, surface curvature measures by shape index and curvedness, surface normal overlap, lines of curvature, and surface normal and sphere fitting.

While these techniques commonly rely on the successful extraction of local surface structures, the irregularities of the colon wall surface and image noise frequently cause erroneous results, which in turn leads to relatively large number of FPs in surface-based CAD schemes

Feature extraction can be used with different techniques tissue intensity characteristics, volumetric and surface shape characteristics, texture characteristics, geometrical characteristics, fecal characteristics [5].

In this paper we are using with three techniques. they are surface intensity characteristics, volumetric and surface shape characteristics and texture characteristics. The flow chart for feature extraction is shown in Fig.6 where features such as volume, intensity and texture are determined [5].



**Fig.6:** Flowchart for Feature Extraction

The aim is to use a polyp specific volumetric feature in the detection of polyps in ctc. shape features uses circularity, irregularity, area, perimeter, shape index. We employ the polyp specific volumetric feature and conduct a full scale parameter sweeping and interaction on detection performance and to determine an optimum setting for the parameters. The materials allocated to training and test datasets. A computer program was used to ensure that the number of cases assigned to the training and test datasets was similar.

In this study, we presented a straightforward approach to polyp detection in CTC. The underlying hypothesis of our study was that a single polyp-specific volumetric feature could effectively distinguish polyps from the other colonic structures.

To test this hypothesis, we introduced a set of volumetric structure filters based on the Eigen values of a Hessian matrix, and discovered a most powerful descriptor for the polyp-specific intensity structure, the blob-likeness. The color volume rendering of blob-likeness superimposed on the CTC image demonstrated its polyp-specificity by distinguishing polyps from the other colonic structures.

A feature that is derived from the thresholded protrusion field should therefore include the size of a candidate as well as the ratio between the largest diameter and the maximum protrusion. Moreover, the feature should characterize the whole segmented area instead of the extrema (like the largest diameter or the maximum protrusion).

We designed a feature that takes into account both the protrusion as well as the lateral size of the object. Effectively, it measures the percentage of the area of the candidate that has a protrusion larger than a certain threshold. A large circumference as well as shallow edges lead to relatively large areas with protrusion and result in a low response.

Shape was previously described by the shape index and curvedness mean curvature, average principal curvatures and sphericity ratio, and spherical harmonics. An alternative method to measure shape, which is based on the protrusion field.

The internal intensity of the candidates has been found before to be a discriminative feature to discard a large number of false detections. It may be expected that due to the partial volume effect false detections arise that have low internal intensity. False detections that are stool often have air inside, which also lowers the intensity. Such information about the candidates will be included through statistics on the object's internal voxel intensities [5].

Texture is a powerful regional descriptor that helps in the retrieval process. Texture on its own does not have the capability of finding similar images, but it can be used to classify textured images from non-textured ones and then be combined with another visual attribute like color to make the retrieval more effective. For the 16 images (each) normal and abnormal images the characteristics such as intensity texture and volume are determined (Table.1).

**Table.1:** Intensity, Texture and Volume characteristics

INTENSITY	TEXTURE	VOLUME
144.8343	0.1726	9.1079
144.8091	0.1752	8.7934
141.0676	0.1940	8.2764
173.5791	0.2621	17.6064
140.5700	0.2138	7.2801
142.6903	0.1830	8.0159
149.3837	0.1744	11.9550
174.6292	0.2365	18.9187
175.0985	0.2533	20.6533
141.3648	0.2242	6.6865
182.1687	0.2537	30.6506
145.9801	0.1732	12.7163
174.2520	0.2797	17.9961
141.3648	0.2242	6.6865
142.8857	0.2326	8.0287
179.5538	0.2539	27.3661
179.8729	0.2460	25.6179
144.4931	0.2211	9.2726
174.9976	0.2607	18.5944
182.2652	0.2750	29.7447

### V. Classification And Evaluation

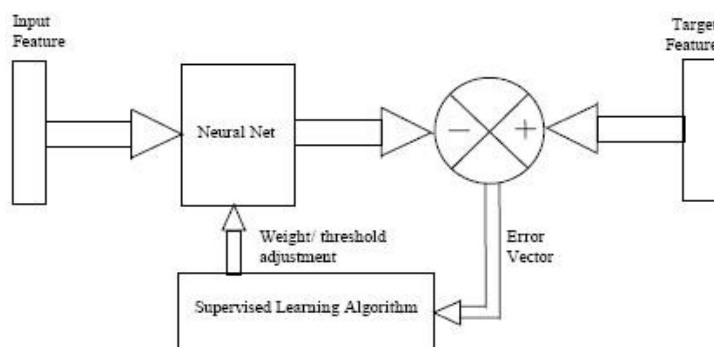
Classification of CT image data is used to assign corresponding levels with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image. Classification will be executed on the base of spectral or spectrally defined features, such as density, texture etc. in the feature space. It can be said that classification divides the feature space into several classes based on a decision rule.

In many cases, classification will be undertaken using a computer, with the use of mathematical classification techniques. Classification will be made according to the following procedures.

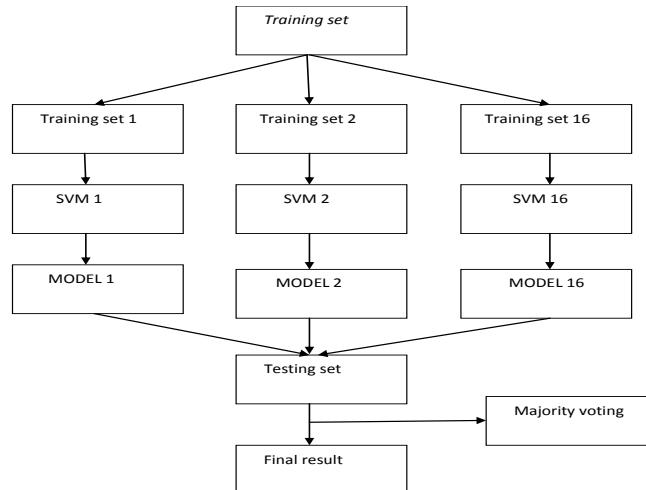
Candidate uniquely identified with the above associated features are then fed to a classifier for final evaluation with the help of training set or test set. The basic flow of training a dataset is shown. The classification for evaluation is performed by using two methods. They are

- SVM classification(support vector machine)
- ANN (Artificial Neural Networks)

The flow charts for SVM and ANN are shown

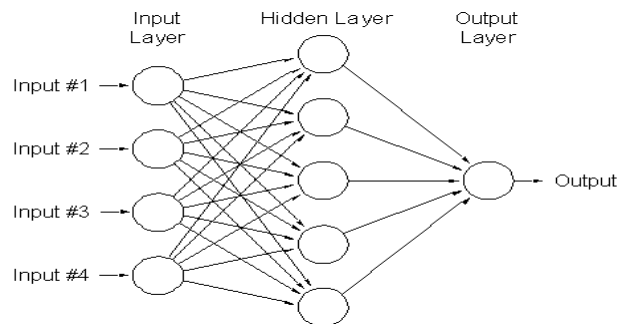


**Fig.7 :**Flow chart for Training a dataset



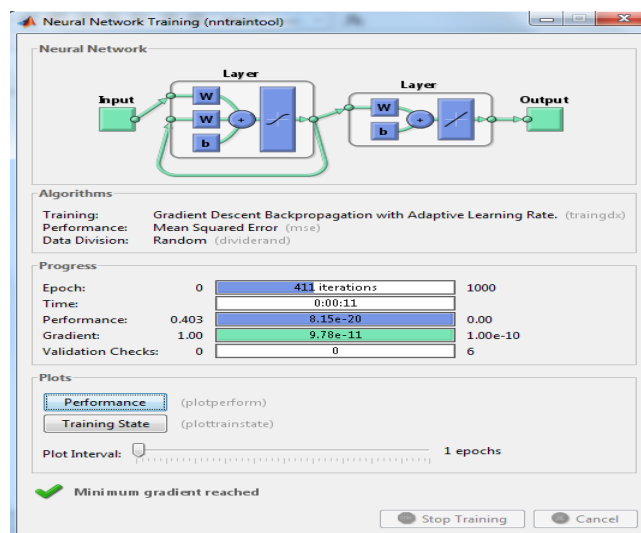
**Fig.8:** Flowchart for Support Vector Machine

The numerical results were obtained by classifying with SVM using cost sensitive learning after normalizing the data. The ANN (artificial neural network) result which are exactly performed with the accurate performance of the training and testing state output. In the SVM, results are obtained by classifying the pixels as non-cancerous pixels (1) and cancerous pixels (2). Thus the cancerous cells are differentiated from the non-cancerous cells.



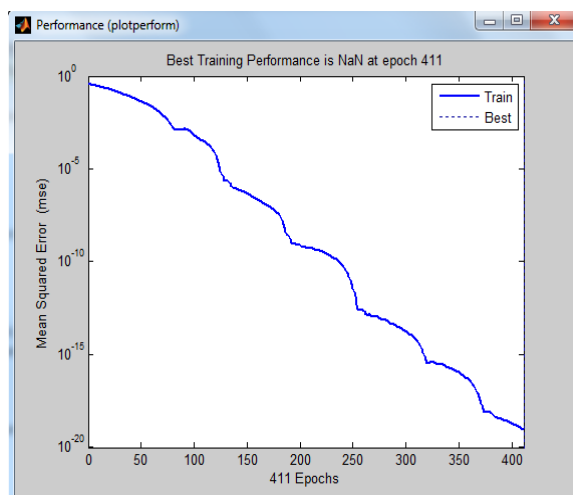
**Fig.9** Flow chart for ANN

The Fig.10 shows the output for neural network training state output. It has been divided into three layers as input layer, hidden layer and the output layer. It is performed with back propagation method that contains about 1000 iteration levels and the gradient value is  $1.00e-10$ .



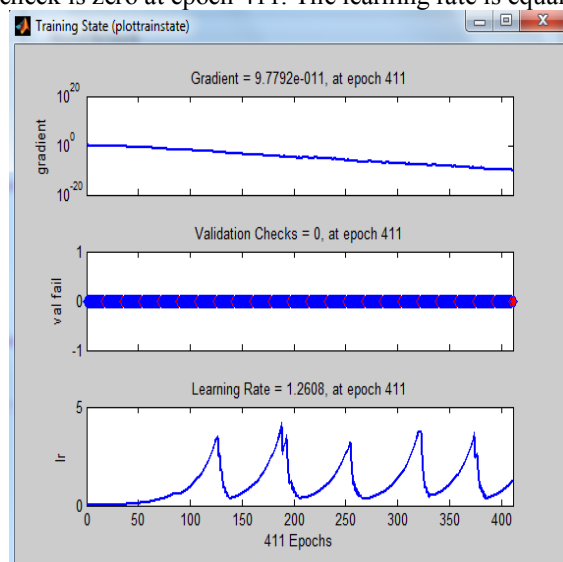
**Fig.10:** Training state output

The train state and the best state are compared and the results are plotted graphically Fig.11. The best training performance is NaN at epoch 411 as shown in figure below.



**Fig.11: Performance**

Also the graph for training state is determined Fig.12; this shows that the gradient value obtained is  $9.7792e-011$  at epoch 411. The validation check is zero at epoch 411. The learning rate is equal to 1.2608 at epoch 411.



**Fig.12 Training State Graph**

## VI. Conclusion

Thus in this paper, both normal and abnormal images were extracted from the datasets went through the pipeline which consisted of various operations such as the segmentation of the colon, candidate extraction and the feature extraction is done based on values of texture, volume and intensity and classification were performed.

The colon has been detected and segmented by Otsu's thresholding and candidate extracted with the original slice. Features such as texture, volume and intensity were determined and training and testing phases were carried out by using the specified set of images from the datasets. Then the classification is carried out by applying the Support Vector Machine (SVM) and Artificial Neural Networks (ANN), thereby distinguishing the cancerous cells from the normal cells.

Thus this method of screening technique of CAD-CTC(Computer Aided Diagnosis- Computer Tomography Colonography) with variations in the segmentation and classification has been really effective in identifying the exact polyps. Thus, this paper helps the radiologists to detect the exact polyp by distinguishing it (polyps) from the folds and false positives. The existing system and the proposed system is compared by using the parameter as shown in the table.

Thus the proposed system demonstrated good results such as the low computation time and high learning rate and also low mean square rate when compared to the existing systems and this system will provide additional information to the clinical decision process.

**Table 2.** Comparison of Automated ,Straight Forward Approach System and Pipeline System in CAD-CTC

PARAMETERS	EXISTING SYSTEM-I[3]	EXISTING SYSTEM – II[29]	PROPOSED SYSTEM
Computation Time	30 Mins	25 Mins	10 Mins
Mean Square Error	1023	974	411
Learning Rate	1.032	0.8032	1.206

## VII. Summary

Thus in this chapter, the image having the colon part are extracted from the dataset. Then the colon part is segmented using the OTSU's thresholding method. The candidates have been extracted from the segmented part of colon. Feature extraction is done with the texture ,intensity and volume characteristics. The classification is performed with the help of SVM(Support Vector Machine) and ANN(Artificial Neural Network) where the performance and computational timing were found.

## References

- [1]. M. Miller, M. Franaszek, and R. M. Summers, J. Yao, (2011) "Colonic polyp segmentation in CT Colonography-based on fuzzy clustering and deformable models," IEEE Transaction on Medical Imaging, vol. 23, no. 11, pp. 1344–1352.
- [2]. I. Bitter, D. Brickman, R. Choi, M. Franaszek, et al, (2010), "Computed Tomographic Virtual Colonoscopy Computer-Aided Polyp Detection in a screening population," Gastroenterology, vol. 129, no. 6, pp. 1832–184.
- [3]. C.D. Johnson, J. D. Malley, L. M. Pusanik, J. E. Reed, R.M. Summers and A.M. Youssef, (2010) "Automated polyp detection at CT colonography: Feasibility assessment in a human population," Radiology, vol. 219, no. 1, pp. 51–59.
- [4]. B. Acar, C. Beaulieu, S. Gokturk, R. B. Jeffrey, et al, (2010), "Edge displacement field-based classification for improved detection of polyps in CT Colonography". IEEE Transactions on Medical Imaging, vol. 21, no-9, pp. 1461–1467.
- [5]. G. Kiss, G. Marchal, P. Suetens and J. Van Cleynen breugel, (2010) "Computer Aided Detection in CT Colonography, via spin images," Lecture Notes in Computer Science, in Proc. MICCAI Conference, pp. 804–812.
- [6]. B. Acar, C. F. Beaulieu, S. B. Gokturk, Jeffrey, Jr, et al. (2009) "A statistical 3-D pattern processing method for Computer-Aided Detection of polyps in CT Colonography," IEEE Transaction in Medical Imaging, vol. 20, no. 12, pp. 1251–1260.