CHAPTER-6 AUTOMATIC ABNORMALITY DETECTION IN CAPSULE ENDOSCOPY USING CONVENTIONAL MACHINE LEARNING AND DEEP LEARNING

This chapter presents a multi-class medical image analysis problem using image processing and machine learning techniques. It presents a CAD system based on the hybrid confluence of the transfer learning technique of deep learning and conventional machine learning technique for automatic abnormality detection in the GI tract from the CE procedure.

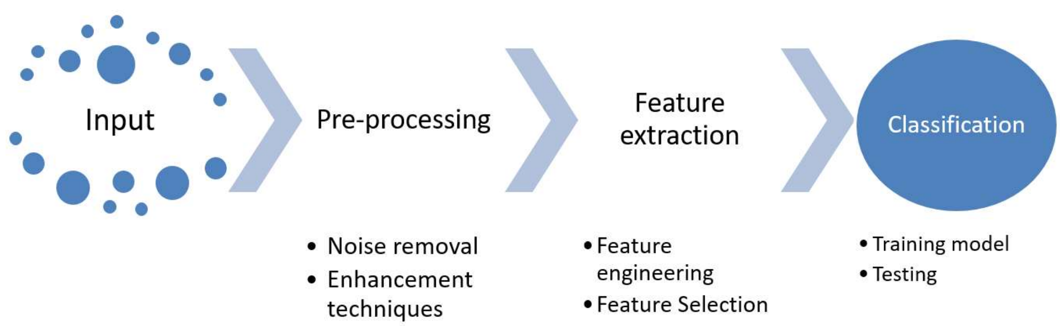
6.1 Introduction

The manual analysis of CE video or images is tedious and time-consuming. Also, decision making is a subjective process, and so a decision may change from person to person. Thus, a computer-aided diagnosis (CAD) system is a must for a fast and accurate diagnosis. CAD systems play an essential role in training inexperienced clinicians and assisting the medical experts in improving the accuracy of medical diagnosis [37]. A CAD system capable of analyzing and understanding the visual scene will undoubtedly help the doctor with a precise, fast, and accurate diagnosis. After the manual analysis of the CE video, CAD can also provide a second opinion to a gastroenterologist [35]. In medical imaging, CAD is a prominent research area capable of delivering precise diagnosis [36]. The ultimate goal of CAD is to reduce interpretation errors, reduce search errors and, reduce variation among observers [37]. In particular, a computer-aided medical diagnostic system for CE can consist of the following units: (1) a data capturing and transmitting unit – the capsule (2) a data receiver and storage unit – the waist belt (3) a data processing unit for pre-processing and feature extraction (4) a machine learning-based classification unit or decision support system (5) a user interaction unit for final diagnostic report.

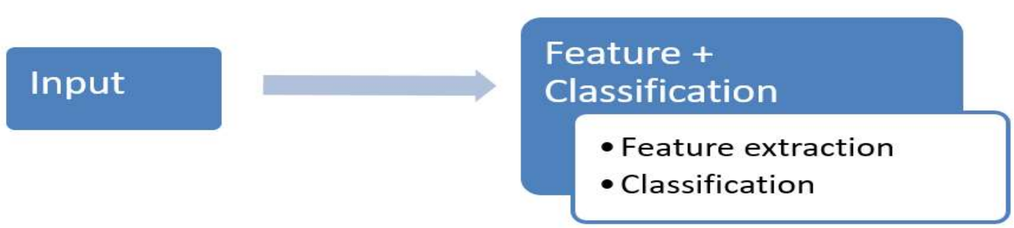
CAD in CE is a multi-class classifier problem. A general drawback of the CAD system in CE is the lack of annotated datasets in the public domain. Due to this reason, a common comparison between proposed and existing systems is a challenge. In the recent past, a few CAD systems for disease detection in CE are proposed; however, only two systems are found capable of detecting multiple abnormalities. Both these systems have been implemented on different datasets-not available in public domain and different abnormalities. Nawarathna et al. [60] proposed a system with 130 texton histogram features from Leung and Malik (LM) and local binary pattern (LBP) to detect bleeding, erythema, erosion, ulcer and, polyp using k-nearest neighbor (KNN). The system performed with a recall of 92% and specificity of 91.8% on a dataset of 1750 images. The method is found computationally intense due to the convolution of image blocks with LM-LBP filter bank. Also, it focuses on the textural information while ignoring other features such as shape from the spatial domain and completely ignores the wavelet domain features. Yanagawa et al. [10] proposed a system with Kanade-Lucas-Tomasi (KLT) feature points to detect red spot, phlebectasi, angiodysplasia, lymphangiectasia, erosion, erythematous, ulcer, and white-tipped villi using affine transform. Their system showed 114 correct predictions, and 06 predictions were lost out of 120 images. The prior art shows four major disadvantages of the existing systems. Firstly the important information is ignored as wavelet domain features and shape features are ignored. Secondly, the systems do not explore full potential of computer vision and machine learning techniques as the learning process is compromised by skewed data, insufficient data and lack of validation of proposed systems. Thirdly, the performance in terms of accuracy and finally the limited size of the dataset. Also other important performance criterion such as f-measure and Matthew's correlation coefficient (MCC) are not discussed.

6.2 Materials and Methods

This study presents automated abnormality detection for CE images using a hybrid approach of conventional machine learning and deep learning. Figure 6.1 gives an overview of how both these approaches differ functionally.



(a) Conventional machine learning



(b) Deep learning

Figure 6.1: The workflow of the (a) conventional machine learning system and (b) deep learning system

As seen in Figure 6.1, the conventional machine learning approach broadly includes pre-processing, feature extraction, and classification. Depending upon the input quality of the image, noise removal and enhancement techniques such as Adaptive contrast diffusion, CLAHE, TV minimization and many more are used. The feature extraction phase aims to derive the most discriminative features. A few such features are GLCM, LBP, color moments, shape features, and wavelet-based features. Subsequently, the classification phase classifies between abnormalities using techniques such as SVM, kNN, ensemble, and many more. Every stage is equally important to achieve successful results in the final stage. On the other hand, the deep learning-based approach is more kind of a black box. One can fine-tune various parameters, and then the model does the feature learning and classification part. Jani et al. [122] presented both these systems individually for CE image analysis. The proposed system is a hybrid of both these approaches, and the results section will show how the proposed method outperforms both approaches individually. The core contribution of this study is as follows:

- The acceptable solution even with the limitation of the size of data.
- Discriminative and robust feature set to classify between CE images.
- Reduced false alarm rate.

Figure 6.2 shows the overview of the proposed solution for automatic abnormality detection in CE and the details of each component of the system is explained subsequently.

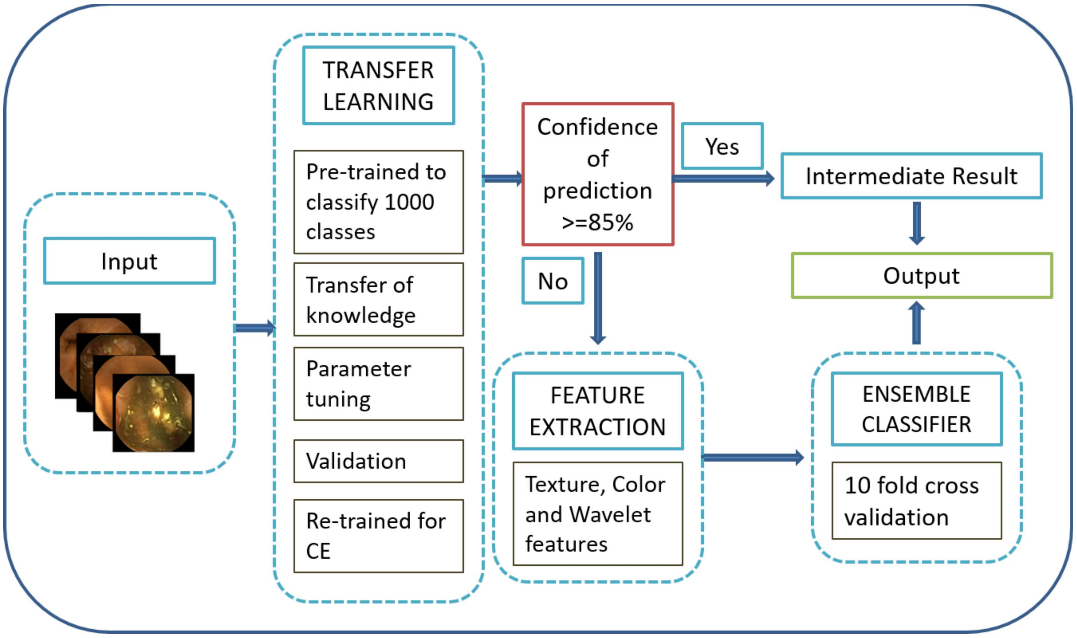


Figure 6.2: Overview of the proposed system

Steps involved in the proposed system:

1. Extract images from CE videos
2. Segregate CE images as per defined class namely angeostecia, bleeding, ulcer and normal
3. Create separate training and testing data
4. Set various parameters for transfer learning such as learning rate, batch size, training, testing, and validation percentage and many more
5. Input CE image training date to re-train transfer learning model
6. Collect intermediate results by considering all samples with the confidence of prediction >= 85%
7. Refer samples with the confidence of prediction <85% to the conventional system
8. Extract features CE image data
9. Train the classification algorithm on training data

10. Classify the referred samples let out from transfer learning technique

11. Merge the results with intermediate results to get the final result

12. Derive various performance metrics of the system

The proposed hybrid system takes into account the threshold for the confidence of prediction as 85%. This value is obtained by performing experiment on four different threshold values namely 80%, 85%, 90%, and 95%. The total misclassifications and overall system accuracy at different threshold values indicate that the lower value tends to overfitting while higher value tends to poor results. Thus to balance between both the threshold value is chosen as 85%.

6.2.1 Conventional Machine Learning

A CAD system for medical image analysis consists of pre-processing for image enhancement, feature extraction to discriminate between classes, feature selection for reduced and relevant search space and classification to reach a diagnosis. The CE images of GI tract diseases exhibit a wide range of color and texture. To address all the diseases under this study, the proposed feature set comprises features from the wavelet domain and color and texture features from the spatial domain. Texture features include GLCM and LBP. The GLCM features are very close to human inference from texture and describe image texture well [123]. After obtaining GLCM matrices for every pixel with offset [0 1], statistical features namely contrast, correlation, energy & homogeneity are calculated. Also, the mean, standard deviation, entropy, RMS, variance, smoothness, kurtosis, skewness, inverse difference moment (IDM) is extracted from the image. Uniform local binary patterns are fundamental properties of the texture of an image, and their occurrence histogram is a compelling textural feature [13]. It is learned that LBP performs robustly to illumination variations. Colour moments for HSI and RGB color space are computed as it provides a measurement for the color similarity between images [110]. Statistical features namely mean standard deviation, kurtosis and, skewness for each of the color channels are computed. The color coherence vector (CCV) shows coherent pixels versus incoherent pixels for each color [111]. CCV consists of 27 different colors and thus, it will create a feature of 54 values indicating coherent and noncoherent pixels for each color. An autocorrelogram captures the spatial correlation between identical colors [124]. Gabor features have been successful in many image processing and computer vision applications. It extracts local pieces of information which are then combined to recognize an object or region of interest [112]. Here we have utilized two features from Gabor response: mean-squared energy & mean amplitude. DWT is a high-level feature extraction technique [125]. In this study, we have considered the first two moments of DWT coefficients as features. The feature vector is then obtained by combining all features. This feature vector is then fed to the classifier to get results. GI tract disease detection in CE is a multi-class classification problem. To perform classification, we have deployed an ensemble classifier: subspace discriminant. An ensemble of classifiers is a set of classifiers, whose individual classification decisions are combined typically by a weighted or non-weighted voting system to classify new samples [126]. The performance is compared with various other classifiers.

6.2.2 Deep Learning

Amongst various deep learning techniques, the transfer learning technique is employed in this study to classify GI tract abnormalities. Deep learning has shown commendable progress in the field of computer vision in general and classification tasks in particular. Most of the deep learning techniques work with an assumption that the training and testing data belong to the same feature space and follow an identical distribution. However, in all real-world applications, this assumption may not always hold. It is challenging to obtain training data. Under such situations, transfer learning is desirable [127]. In transfer learning, a model is trained on base dataset and base task, and then the learned features are transferred to train the target dataset and target task [128]. Two popular approaches to transfer learning are (a) develop a model approach and (b) pre-trained model approach. We have selected the second approach with the MobileNets model. MobileNets are low-latency, low-power, and small models parameterized to meet the resource constraints of a variety of use cases. They can be built upon for classification, detection, embeddings, and segmentation tasks [129]. This study has adopted a pre-trained model approach to transfer learning with the following steps:

- Select a source model
- Re-use the model
- Tune the model

We reused a model MobileNet_v1_1.0_224 using TensorFlow libraries, which was trained on a similar task to design a multi-class CAD system. This architecture-MobileNet was pre-trained on the ImageNet dataset. MobileNet has proven effectiveness in various applications such as object detection, face recognition, and many more [130]. Essentially, we transferred the pre-learned values of the model and added our dataset parameters to the model to re-train the model according to our use. This is effective in our case because the dataset is comparatively small, so we used our dataset to train only the last layer, and utilized pre-learned parameters on the ImageNet dataset. In our case the following simulation parameters are used: MobileNet version=1.0, learning rate = 0.005, number of training steps = 5000, training batch size=20, training, testing and validation percentages as 60, 20 and 20 respectively. As seen in Figure 6.3, the feature extraction part is done by a pre-trained model, and the classification task is performed for CE images. Essentially it is called transfer learning.

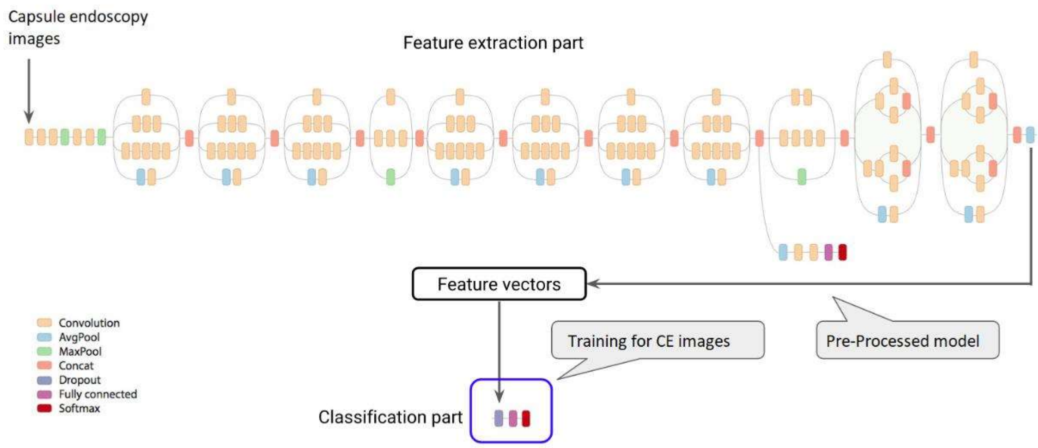


Figure 6.3 Workflow of deep learning-based system

6.2.3 Dataset

A total of 900 images from CE videos [9] is extracted out of which 100 images are of angioectasia, 200 images are of ulcers, 200 images are of bleeding, and 400 images are normal. The dimension of each image is 576x576 pixels. All the images were manually diagnosed and annotated by experts providing the ground truth.

6.3 Results and Discussion

The ultimate objective of the CAD system is to provide a timely and accurate diagnosis with a negligible or no false prediction rate. Both these objectives are nicely met by the proposed system. This experiment with transfer learning shows that how knowledge obtained from a different domain can be transferred altogether to another field and does wonders. Despite different statistical properties, features, and objectives, the knowledge turns out to be very consistent. The marvelous ability of the MobileNet model to classify 1000 different classes gives us an edge over other models. The re-training of the pre-trained model on four class CE image dataset triggered the idea to reach a highly accurate CAD system. This confluence of transfer learning and conventional machine learning overcomes the limitations of both these approaches individually. This proposed model works in two phases; one for coarse tuning and other for fine-tuning. The transfer learning approach is capable to produce the intermediate results with reasonably acceptable accuracy and time at a prediction probability confidence level of 85%. The average evaluation time per image is 0.192 seconds. This first phase left only 18 out of 225 samples to be re-evaluated by the second phase. In the fine-tuning process, the second phase produced results by utilizing the discriminative feature set to classify the left out images. This approach employes an ensemble classifier with 10-fold cross-validation to predict the unknown samples. The average prediction time per image for this phase is 0.001 seconds. The final result is obtained by merging both these results. Figure 6.4 shows the final confusion matrix of the proposed approach.

		Predicted			
		A	B	C	D
Actual	A	24	0	1	0
	B	0	49	1	0
	C	0	0	50	0
	D	2	0	3	95

Figure 6.4: Confusion matrix of the proposed system

The performance of the proposed system is compared with six other approaches on six different criteria. Jani and Srivastava [131] discussed these approaches in designing a CAD system for CE by reviewing the previous works. Table 6.1 shows the statistical comparison and Figure 6.5 shows the graphical representation of the obtained results. It is observed that the system outperforms other approaches in all aspects.

Table 6.1: Comparative analysis of the system with other approaches

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-measure	MCC
SVM cubic	91.33	88	96.74	91.35	89.31	86.41
SVM linear	85.89	79.69	94.46	85.88	82.03	77.42
Tree	79.56	74.37	92.66	74.57	74.43	67.16
Ensemble RUSBoosted Trees	86.33	86.06	95.48	83.38	84.01	79.74
Ensemble subspace discriminant	92.67	89.87	97.31	91.83	90.75	88.26
Deep learning [122]	95.11	95.25	98.42	92.83	93.9	92.26
Proposed	**96.89**	**97.25**	**99.04**	**95.80**	**96.45**	**95.41**

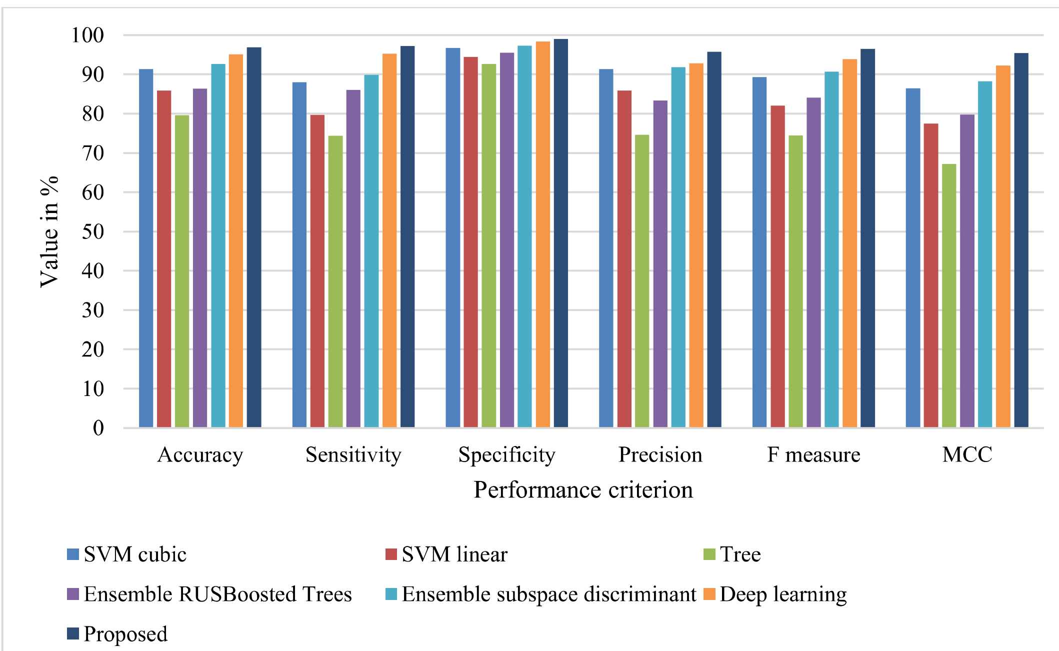


Figure 6.5: Graphical representation of results

In addition to the above results, this study also observes the evaluation time taken by each method. Table 6.2 presents the evaluation time in terms of observations per second.

Table 6.2: Comparison of evaluation time between different approaches

Classifier	Observations evaluated per second
SVM cubic	1200
SVM linear	2700
Tree	3000
Ensemble RUSBoosted Trees	3300
Ensemble subspace discriminant	1200
Deep learning [122]	5.2
Proposed	5.18

6.4 Conclusion

The proposed system is a hybrid system of transfer learning and conventional machine learning technique. It performs with an accuracy of 96.89% and an F-measure of 96.45%. The system outperforms not only in accuracy but also MCC, which is a balanced measure and Kappa coefficient, which is a more robust measure than simple percentage agreement measures. Kappa coefficient of the proposed system is 0.917, while amongst the rest of the systems, the best kappa co-efficient value is 0.804. Notably, the false positive is drastically reduced, and most importantly, none of the abnormal images, i.e., angioectasia, ulcer, or bleeding, are misclassified as normal. Thereby the false negative is nil. The total time of evaluation of each sample is approximately 0.193 seconds empowering the system to provide a fast and accurate diagnosis of GI tract abnormalities from CE images.