



A shallow extraction of texture features for classification of abnormal video endoscopy frames

Hussam Ali^{a,b,*}, Muhammad Sharif^b, Mussarat Yasmin^b, Mubashir Husain Rehmani^c

^a Department of Computer Science and IT, University of Sargodha, Sargodha, Pakistan

^b Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan

^c Department of Computer Science, Munster Technological University (MTU), Ireland

ARTICLE INFO

Keywords:

Classification
Endoscopy
Texture analysis
Gastric cancer
Deep learning

ABSTRACT

Automated analysis of the gastric lesions in endoscopy videos is a challenging task and dynamics of the gastrointestinal environment make it even more difficult. In computer-aided diagnosis, gastric images are analyzed by visual descriptors. Various Deep Convolutional Neural Network (DCNN) models are available for representation learning and classification. In this paper, a computer aided diagnosis system is presented for the classification of abnormalities in Videos Endoscopy (VE) images based on Deep Gray-Level Co-occurrence Matrix (DeepGLCM) texture features. In our scheme, the convolutional layers of an already trained model are employed for acquisition of the statistical features from responses of filters to estimate the texture representation of VE frames. A learning model is trained on these features for gastric frames classification. The results obtained by using public datasets of endoscopy images to calculate the performance of the proposed method. In addition, we also use a private endoscopy dataset which is acquired from the University of Aveiro. The DeepGLCM outperforms by achieving the average accuracy of $\approx 92\%$ and 0.96 area under the curve (AUC) for the chromoendoscopy (CH) dataset and $\approx 85\%$ accuracy for Confocal Laser Endomicroscopy (CLE) and white light video endoscopy datasets. It is evident that the DeepGLCM texture features provide a better representation than the traditional texture extraction methods by efficiently dealing with variance in images due to different imaging technologies.

1. Introduction

Endoscopy is a less invasive way of screening human organs. An endoscope is a wire like instrument and it has a camera, light source, channel (for cleaning with air or water), and an accessory channel (for biopsy or ligation). Endoscopy is a routine imaging technique for the detection, diagnosis, and treatment of diseases in hollow-organs; the esophagus, stomach, colon, uterus, and the bladder. Endoscopic-assisted ear surgery has its advantages and disadvantages in otology and neurotology [1]. Endoscopy has recently been extended to the upper aerodigestive tract, revealing good results in different clinical fields. From top to bottom, multispectral imaging (such as Narrow Band Imaging) seemed to deliver meaningful details drawn from endoscopic images [2].

Currently, video endoscopy has been widely used by medical experts for screening gastric patients. The literature confirms its benefits for a non-invasive early diagnosis of gastric cancer [3,4]. The inspection of

long videos or a huge number of endoscopic frames is a laborious job for a gastroenterologist [5]. Given the rapid advancement in endoscopy technology and image processing, machine-vision based techniques help in developing automated systems for better diagnosis. The computer-aided diagnosis systems (CADs) are very helpful for providing a second opinion to medical experts. Also, CADs can be used to train the clinical staff and medical students [6].

In a real clinical setting, the automated method will help doctors in detection of abnormal frames in video endoscopy and also provides a second opinion to doctors especially when they are tired and there is high chance of miss detection.

CADs can facilitate both the patients and doctors. An endoscopy session outputs a large number of endoscopic images. Inspection of every endoscopic-frame in the whole video sequence is an exhaustive task for medical experts. In contrast with manual inspection, automated analysis of gastroenterology images using computer vision techniques can assist gastroenterologists by finding out abnormal frames from the

* Corresponding author.

E-mail address: hussamalic@gmail.com (H. Ali).

<https://doi.org/10.1016/j.bspc.2022.103733>

Received 8 August 2020; Received in revised form 18 March 2022; Accepted 27 April 2022

Available online 25 May 2022

1746-8094/© 2022 Elsevier Ltd. All rights reserved.

whole endoscopic sequence. Such systems can automatically highlight, detect, or classify the abnormalities in the endoscopy videos. This paper aims to design efficient computer-assisted diagnostic methods for the recognition of gastric abnormalities, hence requiring features that are robust to scale, rotation, and illumination variations.

On the other hand, various advancements of endoscopy technology are available. For instance, Chromoendoscopy (CH), Confocal Laser Endo-microscopy (CLE), and Wireless Capsule Endoscopy (WCE) are available to provide a visual aid to a gastroenterologist for observing the mucosal structures with great detail.

The dynamics of image acquisition in the gastric environment poses novel challenges in machine-vision. Therefore, the endoscopy frames normally suffer from the illumination, spatial, and scale invariance caused by uncontrolled camera movements. Moreover, the endoscopic frames may suffer from poor cleansing of the gastric tract and lens distortions. Furthermore, gastric lesions do not have a specific geometrical structure which makes the gastric surveillance more challenging.

Colors are good visual descriptors but the gastric frames have minimal usage of color spaces [7,8]. Due to similar color characteristics, it is hard to distinguish abnormal and normal frames based on only colors. In contrast with colors, texture features are often used in medical image analysis and in many other domains. Texture refers to repetitive intensities, distributed over the images and can also be described as the property of a surface (e.g., smoothness and roughness) [9]. Given a reduced color space in endoscopy frames, texture can be a good candidate for the representation of CH images for classification of gastric frames [10,11].

Several techniques are proposed for texture extraction from images. Texture descriptors are used for the detection of abnormalities in gastric frames. Some widely used texture extraction methods are local binary pattern (LBP) [12], homogeneous texture (HT) [13] and statistically computed texture features [14]. Similarly, the texture descriptors can be estimated from the Gray-Level Co-occurrence Matrices (GLCM) [15] are used for various classification tasks.

Deep learning has revolutionized almost every field. The deep convolutional neural networks (DCNN) can be used for representation learning instead of using explicit traditional techniques [16]. The learned features from different DCNNs can be fused to achieve a better performance than a single learning model [17]. Similarly, LBP and DCNN features are concatenated to form feature set for hyper-spectral images classification [18]. In [19], a co-occurrence matrix is formed by DCNN and GLCM features are extracted for detection of polyps. In this study, we fuse deep learning method with traditional GLCM features by extracting GLCM features from activations of convolutional layers instead of using full deep convolutional neural network [20].

For experimentation, we have used 4 publicly available and 1 private images datasets. The problem with most of the datasets is that most of the datasets contain a small number of images because from an endoscopic video, a small amount of abnormal frames are collected. Problem of small number of training examples can be dealt with data augmentation but, the low inter frame differences can cause over learning of network. In this study, we will not use an entire DCNN, instead, we only use initial DCNN layers for the computation of texture features.

Previously, we extracted the homogeneous texture (HT) descriptors by using Gabor filters by changing statistical responses in GHT [21]. Also, we extracted GLCM features from Gabor responses [22] to see the impact of hybrid approach. It has been known that the convolutional layers filters are similar to Gabor filters and can be replaced by Gabor filters [23]. By using this idea, we extract the GLCM features from the activation of a convolutional layer of a DCNN.

A DCNN is made of many layers followed by the input layer, such as convolutional, pooling (min, max or average), Rectified Linear Unit (ReLU), fully connected, and classification layer. Each layer has its importance in performing the classification task. Fully connected layers contained all the features and mostly situated where the classification is performed. ReLU is used as the activation function. In some cases, the

leaky ReLU is used for activation of the neurons. The pooling layers are used for extracting important information and the core purpose of this layer is downsampling. We have used convolutional layers activations to capture the texture information from the responses of filters in the convolutional layer. For experimentation, we have used a basic DCNN, the Alexnet [20].

In this paper, we proposed a technique for texture extraction by joining of the DCNN and GLCM. In the proposed technique, we have calculated texture features from the activation of filters in convolutional layers of DCNN and computed GLCM features. For additional experiments, we also estimated local GLCM texture features to see the impact of locally computed texture features vs global features.

For local texture extraction, the VE frames are divided into equal sized blocks, then the GLCM from each sub-image is calculated. Also, deep features of these blocks are computed. Finally, the texture of every VE image is represented by a vector of second-order statistics, locally computed from co-occurrence matrices. Moreover, multi-sized blocks are selected to see the impact of different sizes on the classification performance. Another reason for selecting different numbers of blocks is to find the optimal image-dividing criteria. Then the extracted features are used for training state-of-the-art classifiers.

GLCM texture features are traditionally used for the classification of images. Similarly, deep learning methods are used to learn features from given images and can use these features for classification. It is intuitive to combine both methods to get more robust texture descriptions of gastric images that can lead to better discrimination of gastric images for abnormalities.

Existing CADs are based on traditional hand-crafted features or deep learning methods. Both approaches work well for image classification. However, endoscopic frames classification is a difficult task due to the complex nature of gastric environment. A new hybrid approach is adopted by combining the traditional feature extraction method GLCM and the deep learning method to obtain better image discrimination. Traditional methods explicitly define the procedure for feature extraction and a perfect definition exists about features. It is already known that what type of information will be extracted from gastric images. On the other hand, modern deep learning methods are mostly based on assumptions (by observing empirical results to decide whether a model is converging or not) and it is expected that (for example second layer or third layer capturing texture and shape information respectively) DCNN model is learning high level texture or shape features from images.

1.1. Contributions

The main reason behind computing GLCM locally is the distribution of intensities that are not uniform in the whole image. Moreover, the computation of features over the entire image using traditional statistical methods suppresses the original response. The deep texture features give us more local responses on different filters like Gabor filters [21]. The proposed feature extraction technique provides a scale, rotation, and spatial invariant feature set. The major contributions of this paper are highlighted as follows:

- DeepGLCM texture extraction method is presented where the statistical features are extracted from the activations of convolutional layers to provide more robust features.
- A L-GLCM features extraction method is used by extracting features from sub-images by dividing images into multiple blocks for the representation of local texture and achieve spatial invariance. The SVM classifier is used to classify images based on local texture features.
- Both traditional and current state-of-the-art classification models are trained for experimentation and their results are compared.
- Both traditional feature extraction methods and current state-of-the-art deep learning models are trained for experimentation and their results are compared.

– Multiple publically available datasets of different modalities are used to validate the performance of the proposed method.

In this study, we will address the following research questions:

Q1. Which layer of a DCNN like Alexnet contains texture information which is useful for gastric frame classification? (for answer see Section IV–C)

Q2. Are there any data-dependent factors involved in convolutional layers for feature learning? (for answer see Section IV–D)

Q3. How the layers of a trained DCNN can be efficiently used for the classification of endoscopy images? (for answer see Section III–A1)

Q4. What will be the impact of locally computed texture and the performance gain that cannot be achieved by using globally computed features? (for answer see Section IV–A)

Q5. Does the proposed method works for different gastric modalities and images other than simple video endoscopy for abnormalities detection? (for answer see Section IV–D)

Q6. What will be the impact on the performance of proposed method if it is used for the detection of different abnormalities? (for answer see Section IV–D)

The comparison is done by extracting GLCM features from the whole image and proposed deep texture features and L-GLCM. The answers to the above questions are given later sections *in italic font*.

1.2. Organization of the paper

This paper is structured as follows: The upcoming section gives a review of existing methods. The GLCM, L-GLCM and DeepGLCM texture features and classification of CH, VE, and CLE images are discussed in Section III. The experiments and materials are also presented in this section. The results of various classifiers are presented in section IV, results are compared with existing methods discussed in Section IV–D, and this paper is concluded in Section V.

2. Related Work

There are several methods developed so far for the discovery of abnormalities in endoscopy frames. These methods classified into four categories based on their features as follows:

2.1. Statistical Methods for Computing Textures Descriptors from Endoscopy Frames

The higher-order local auto-correlation (HLAC) based descriptors are used for the detection of ulcerative colitis and also used for image retrieval tasks. Moreover, this method uses multi-level HLAC features [24]. In the same way, the statistical descriptors (e.g., mean and variance) extracted from images for computation of texture representation. Then, KNN and SVM classifiers used for classifying cancerous frames [25].

Statistics calculated from the image's histogram are also widely used for the representation of color features. The histogram features of endoscopic images are used for representing the texture of images and combined with wavelet-based texture features to classify frames with celiac disease [26]. Likewise, a multi-layer perceptron is trained on several statistical descriptors for the detection of cancer in endoscopy [27].

2.2. Analysis of Images in Frequency Domain for Texture Features Extraction from Endoscopy Images

Gabor filters can be used for the multi-resolution analysis of gastric images and rotation invariant texture descriptors extracted by exploiting the auto-correlation property of Gabor filters. Moreover, naïve Bayes

(NB) and SVM classifiers used for classification [7]. Wavelet-based LBP texture obtained from WCE images for detection of cancer frames [28,29]. Similarly, the color wavelet covariance computed for texture representation of WCE images in [30]. Pit-patterns of mucosal surface analyzed by computing Gabor wavelet and dual-tree complex wavelet transforms for the detection of cancer [31].

2.3. Hybrid Approaches for Texture Extraction from Endoscopy Images

In [16], a review of different models and methods is presented for detection of gastric disease from endoscopy and it is concluded that most of the methods are based on a hybrid approach for better discrimination of gastric lesions.

Gray-level and global texture features are extracted for the early detection of carcinoma from endoscopy frames by using a higher-order graph matching kernel SVM classifier in [32]. Similarly, the GLCM with a fusion of color features is extracted from endoscopic frames in [33] for the detection of stomach gastritis and used SVM for training and classification of endoscopic-frames. GLCM texture features combined with temporal features in [34] for retrieval of images with similar clinical conditions from gastric images database. Color co-occurrence can compute the color-texture features from the wireless capsule endoscopy (WCE) image for bleeding detection by assuming the blood has a texture. Moreover, the dominant color features computed from hue, saturation, and value (HSV) frequency of images [35]. In the same way, LBP texture with the color histogram features used to classify the endoscopy frames having cancerous regions in [36]. The ensemble method (gentle boost) used for training over texture features after training, it employed for classification of the endoscopy frames. LBP texture is used for the detection of polyps by training an artificial neural network in [37]. Likewise, LBP and GLCM features used for segmentation of polyp in colonoscopy frames with SVM classifier [38].

The contraction of the gastric tract is detected using wrinkle-based features that comprised of fourteen attributes: a set of 8 orientation features, four local entropy related descriptors, two descriptors linked to edge sharpness [39]. Moreover, the SVM classifier used for the categorization of gastric frames. The texture spectrum employed for the discovery of polyps from colonoscopy images in [40], where a wide range of texture features extracted to train SVM classifier for classification.

Colors and textures are important visual cues for the detection of gastric abnormalities like ulcers, bleeding, and cancer. In the case of polyps, shape features are also important therefore, the geometry, texture, and color of a polyp give sufficient information for detection. The geometry features of polyps are extracted by pyramid histogram of oriented gradient (PHOG), a fractal weighted local binary pattern (FWLBP) is used for texture representation and it also provides partial illumination invariance. Feature fusion is used by feature ranking algorithm based on the fuzzy entropy then SVM classifier is trained for classification of gastric frames [41]. Features in different color spaces are computed and combined with LBP then SVM classifier is trained for classification of CH image [8].

2.4. Deep learning Methods for Feature Learning from Endoscopy Images

Bleeding in the gastrointestinal tract is a common symptom for many abnormalities. Therefore, finding bleeding frames from the endoscopic sequence is important for the early detection of gastric diseases. A DCNN model is designed by modifying MobieNet named as BIR (bleedy image recognizer) to classify WCE bleeding frames [42]. Three pre-trained models (VGG19, InceptionV3, and ResNet50) are used to learn features from WCE frames. Further, these features are fused using the minimum redundancy maximum relevance method to obtain an efficient feature vector then the SVM classifier is trained on selected features for the classification of WCE frames [43]. The full potential of DCNN is utilized by designing an Esophageal Lesion Network (ELNet) for the classification and segmentation of gastric images. In ELNet,

contextual lesion information is used to extract both local and global features for classification and segmentation of gastric frames [44].

Texture features can be used for segmentation and low-level statistical textures are useful for this purpose [45]. Gabor texture features are combined with pre-trained DCNN features for content-based image retrieval [46]. Content-based image retrieval can be used in medical imaging for retrieving images with similar abnormalities against a query image [16]. Five gradient boosting models are trained and features selection is performed for the detection of tumors. Along with different endoscopy techniques, some CADs are based on different types of imaging modalities such as CT scan [47]. Similarly, GLCM texture features are extracted from gastric lesions from CT scan images to detect metastasis of gastric stromal tumors [48].

In some cases, endoscopy is used for head and neck cancer surveillance and detection [49,50]. Surgical access from the neck to the jugular foramen is feasible with endoscopy and reduces the risk of trauma and various complications. It is an effective way of detecting jugular foramen tumors [51]. Colonoscopy is used for screening of lower part of the gastric tract. It is also useful for the detection of polyps, ulcers, and other abnormalities. Images obtained from colonoscopy are used for automated detection of any of the given abnormalities [52]. There are some modalities like WCE where endoscopy is a non-invasive way for screening the whole human gastric tract. In [53] WCE video summarization is done using CNN transfer learning. Similarly, DenseNet is used for the detection of ulcer colitis in [54].

There are two shortcomings with existing texture feature extraction methods: (1) The traditional methods like LBP, Gabor wavelet-based, and GLCM are well-known for texture analysis. To the best of our understanding, no single technique can cope all the complexities of image acquisition in the gastric environment, and (2) most of the exiting techniques created for general recognition applications which are not specific to gastric frames. Still, there is a need to develop efficient representation for gastric frames that provides rotation, scale, spatial, and illumination invariance.

3. Materials and Methods

In this section, DeepGLCM texture extraction approach is discussed along with GLCM texture extraction and architecture of Alexnet is discussed in detail.

3.1. Deep Gray-Level Co-occurrence Matrix (DeepGLCM) Texture features

In the proposed scheme, texture features extracted from the entire image by computing the GLCM of every image. Also, a DCNN used for learning features from traditional GLCM methods. These traditional descriptors extracted from the convolutional layer activations of the DCNN as shown in Fig. 1. A feature set is formed by computing the statistics of co-occurrence metrics histogram. For the local texture features, images divided into patches and the GLCM computed from these sub-images.

3.1.1. Gray-Level Co-occurrence Matrix (GLCM)

The texture of endoscopic frames represented by first calculating GLCM. The GLCM contain histograms of a pair of pixels, located at a specified distance. The GLCM can be interpreted as a matrix $C_{L \times L}$ and all elements contain the frequency c of set of pixels at a defined location (distance) D where L symbolizes gray values [15].

$$C_{ij}(D, R)_{L \times L} = \frac{1}{N} \sum_{l=0}^{L-1} \left(\sum_{x=1}^X \sum_{y=1}^Y c_{x,y} \begin{cases} 1, & \text{if } I(x, y) = i \\ & \text{and } I(x', y') = j \\ 0, & \text{otherwise} \end{cases} \right) \quad (1)$$

In Eq. 1, x and y are horizontal and vertical coordinates of image I and $D = (\Delta x, \Delta y)$ are calculated by computation of differences in both

dimensions (horizontal $x' = x \pm \Delta x$ and vertical $y' = y \pm \Delta y$. Here, I signifies the input frame and \pm sign specifies the offset in both orientations, i, j are horizontal and vertical coordinates of the GLCM. Where, C (in GLCM) indicates the frequency on a specific distance D and R indicates rotation factors. The pixels $I_{x,y}$ and $I_{x',y'}$ are positioned at a particular distance on a specified angle.

The GLCM encompasses the joint likelihood of gray values occurrence. The matrix has count of pair of pixels like a histogram. Afterward, the second-order statistical features are calculated from the GLCM to represent the texture of frames. The statistical measurements for instances, variance, mean, covariance, etc., are obtained from the estimated co-occurrence matrices. The features represent information globally about the gastric wall of an image. The correlation, contrast, energy and homogeneity described in [15] are extracted from GLCM to describe the texture of gastric images. The contrast computes the sum of differences of GLCM elements, and it preserve the local variation in image intensities as presented in Eq. 2.

$$\text{Contrast} = f^1 = \sum_{i=0}^M \sum_{j=0}^N C_{ij}(i-j)^2 \quad (2)$$

Where, C_{ij} represents the GLCM and GLCM elements are denoted by i and j . The correlation is the occurrence of the joint probability of certain gray values. This relationship is only calculated for a specific pairs of pixel, as described in Eq. 3. M and N represent horizontal and vertical size of the GLCM matrix.

$$\text{Correlation} = f^2 = \sum_{i=0}^M \sum_{j=0}^N C_{ij} \left(\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{\sigma_i^2 \times \sigma_j^2}} \right) \quad (3)$$

In Eq. 3, μ_i and μ_j are mean values of GLCM and σ_i and σ_j are the standard deviation in horizontal and vertical directions. The measure of uniformity of intensities in an image is computed by taking the squared sum of all GLCM elements. Therefore, the energy or angular second moment of an endoscopy frame calculated as given in Eq. 4.

$$\text{Energy} = f^3 = \sqrt{\sum_{i=0}^M \sum_{j=0}^N C_{ij}^2} \quad (4)$$

A measure of similarity of elements of the GLCM is defined by Eq. 5 [15]. This quantity provides insight into the repetition of specific frequency contents in the endoscopic images.

$$\text{Homogeneity} = f^4 = \sum_{i=0}^M \sum_{j=0}^N \frac{C_{ij}}{1 + (i-j)^2} \quad (5)$$

The texture representation is extracted from GLCM. The GLCM is computed from the activation of all convolutional layers. Also for experimentation in this method, we have divided images into small blocks of equal sizes to extract GLCM features from each block of endoscopic images¹.

$$\vec{S} = \sum_{i=1}^M \sum_{j=1}^N F(\widetilde{a_{ij}})^\ell \quad (6)$$

DeepGLCM feature vector \vec{S} is formed by extracting GLCM from each activation $\widetilde{a_{ij}}$ of a convolutional layer ℓ . In this scenario, ℓ can be range from 1 to 5.

$$f: \vec{S} \rightarrow C \quad (7)$$

A DCNN is used to perform various tasks. In this particular case, we used it for feature learning instead of using the whole network. We have used the convolutional layers of DCNN and get activation from these layers by providing the input data as defined in Eq. 6. The classification

¹ Answer to Q3.

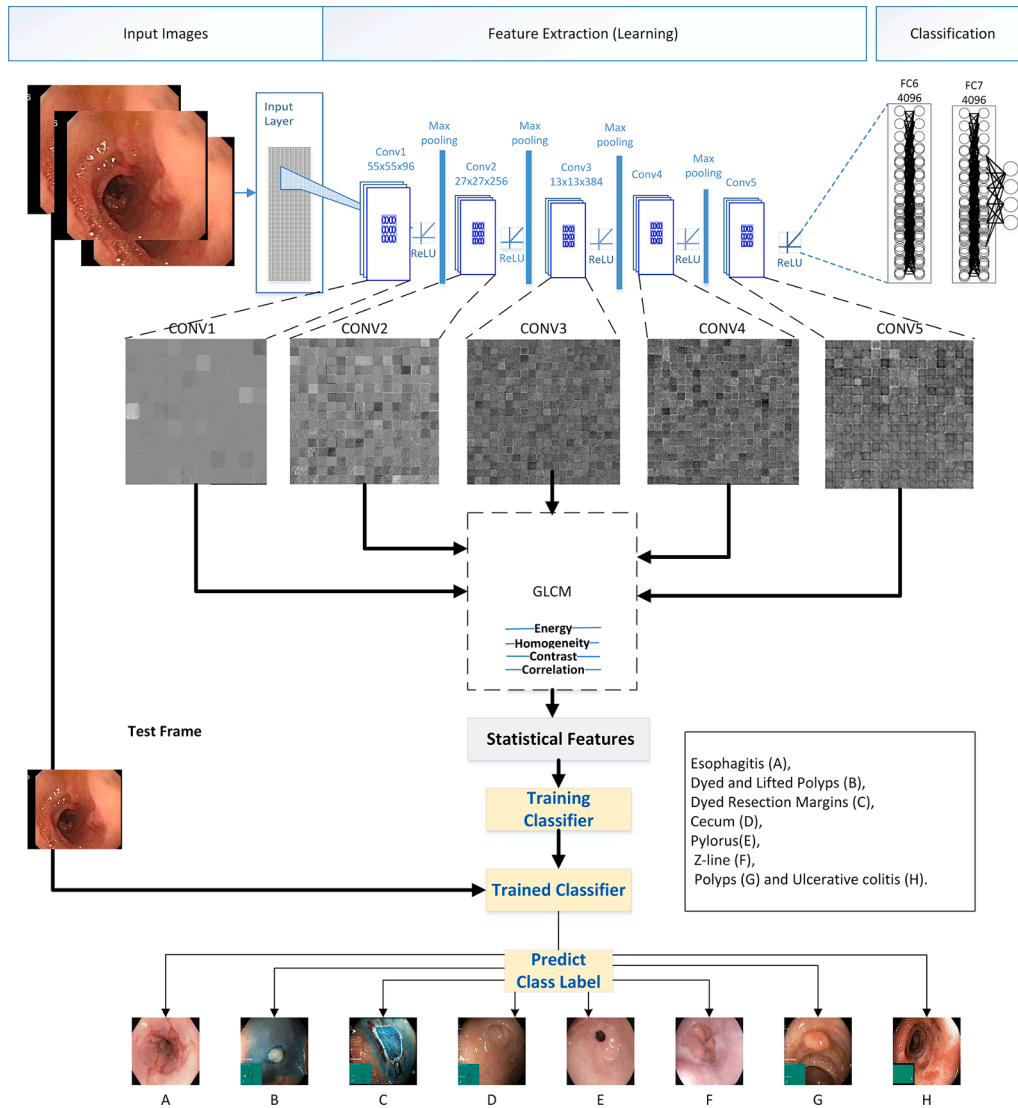


Fig. 1. Flow diagram of DeepGLCM technique where GLCM texture features (obtained by using activation of DCNN) are extracted from different types of VE images and classified them into different categories.

problem can be defined as a function which takes feature vector as input and output class label. In Eq. 7, \vec{S} is the features vector and $C \in \mathbb{Z}$ labels.

3.2. Architecture of Alexnet

DCNN models are available with variable depths and widths. Some layers store all training information while other layers help to optimize and enhance the performance of DCNN. Alexnet [20] is one of the publicly available pre-trained models designed for image classification into 1000 categories. It has 25 layers in total with eight learning layers (5 convolutional and three max pooling). Authors of Alexnet also admitted that removing any convolutional layer from the DCNN (each containing no more than 1% of the model parameters) resulted in low performance. The convolution layer works like filtering an image with multiple small sized filters and records responses to these filters. The convolution layers deal with the local transformation of images and help parameter reduction, which allows more massive data inputs.

Each convolutional layer has the same dimensions as the input connections, but each pixel is only activated by a region of pixels centered around the pixels at the same location in input images. The weights are also shared for each output pixel. Each map in a convolutional layer performs a convolution of input images with a learned

kernel.

Traditionally, neurons are activated through a hard limiting function. However, recent researches show that ReLU has better performance over large datasets because when DCNN is trained using extensive data, there is always the issue of overfitting (when a training model best performs on train data and shows low performance on test data). Dropout layers are used for regularization. The regularization is an essential step for removing any bias which causes overfitting.

3.2.1. Convolutional Layer

The DCNN we are using for experimentation, contains 5 convolutional layers. Each layer has a different number and sizes of filters and different strides. Normally, in transfer learning features are extracted from fully connected layers but it has a large dimension and information received from the near classification layer. Computing all these features requires all previous layers to pass information and present it to fully connected layers.

In Alexnet, we have conv1 with 96 11x11x3 convolutions with (padding [0 0 0] and stride [44]). Second, the conv2 layer with 256 5x5x48 convolutions with padding [2 2 2] and stride [1 1], conv3 with 384 3x3x256 filters by using padding of 1 and stride [1 1 1], conv4 with 384 3x3x192 learning parameters and same strides as its previous

layer, and conv5 with 256 3x3x192 filters with one padding and stride [1 1 1 1].

For deep learning, we input different types of endoscopy frames into a DCNN. We got different responses from each filter and by getting each response we computed GLCM of each response of each filter from the activations of convolutional layers.

3.3. Local-Gray-Level Co-occurrence Matrix (L-GLCM)

To compute texture descriptors locally from images, the frames are divided into sub-images. After the division of images into sub-images, GLCM is calculated from every sub-image.

Analysis of sub-images provides texture information from a small region of the image. We have an assumption, that it will capture the texture of images more precisely in contrast with the extraction of GLCM from whole images. GLCM computed from whole images fail to capture texture with high detail. Moreover, gray values having more repetition in GLCM dominates the gray-levels with less appearing values. Therefore, images are divided into sub-images or blocks to extract GLCM-based texture features.

The images are first scaled to 256×256 and then converted from RGB to gray-scale for texture extraction. As we have described earlier, there is a limited utilization color space in gastric frames and we assume that it is not sufficient for discrimination of gastric lesions as shown in Fig. 2. Because abnormal gastric lesions have a coarse texture, whereas normal lesions possess a uniform texture. Thus, colors are not as much important as textures for description of gastric lesions [7].

Furthermore, all the images are divided into 2×2 , 4×4 , and 8×8 blocks. It means that 2×2 has four sub-images, 4×4 has 16 sub-images, and 8×8 has 64 sub-images. Therefore, the numbers of blocks are 4, 16, and 64 respectively. There are three main reasons for selecting these block sizes are as follow:

- First, smaller blocks restrict the use of GLCM because of distance parameter D .
- Secondly, the dimension of features will increase and classifiers will face the issue of the curse of dimensionality.
- The third reason for selecting different sizes of blocks is to see the impact of different divisions on the performance of calculated texture features.

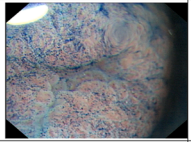
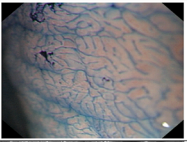
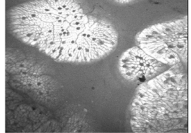
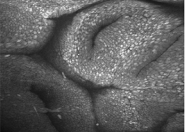
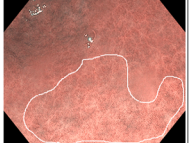
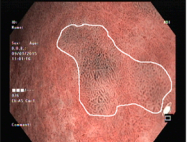
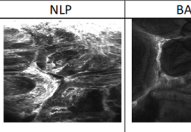
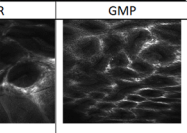
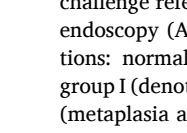
| Dataset | Endoscopy | Disease | Class Labels | |
|--------------|-----------|----------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| | | | Normal | Abnormal |
| Chromogastro | CH | Cancer |  |  |
| CLE celiachy | CLE | Celiac Disease |  |  |
| Private | VE | Cancer |  |  |
| 'CLE barrett | CLE | Barrett | NLP | BAR |
| | | |  |  |
| | | | GMP |  |

Fig. 2. Examples of different types of endoscopy images from datasets used in experimentation with normal and abnormal mucosal texture.

After computing the GLCM features from every sub-image, statistics are computed to create a feature set. Then, these locally extracted textures features from endoscopy images are used to train classifiers.

3.4. Training of Learning Models for the Classification of Video Endoscopy Images

After the extraction of texture representation, a multi-class SVM classifier is trained on these features by using the 10 cross-validation method.

Subsequently, the trained classifiers are used for the classification of gastric images. The reason for training multiple classifiers is to compare their respective performance on GLCM, L-GLCM, and DeepGLCM texture features.

The SVM classifier is trained with a quadratic function. The main reason for selecting the quadratic function is to classify non-linearly separable data. Box constraints are set to one and standardization is done on input features for better convergence. For multi-class classification, one on one method is used in which one class is considered positive, other is negative, and remains is ignored. This technique exhausts all combinations of class pair tasks. Further, we have used two more classifiers KNN and Random Forest (RF) for comparison. In KNN, 10 neighbors are used and euclidean distance is used by squared inverse distance weights. In random forest, bagging is used with 30 learning cycles and tree is used as a weak learner.

Visually there is a very small difference between two frames this can also be an indication of less differences in features of two distinct images. However, the SVM model maximizes the gap between the decision boundary and distinct features which leads to better discrimination between gastric images belonging to different classes.

For the classification of gastric images, a multi-class SVM classifier trained and tested with multiple 10 cross-validation iterations. The DeepGLCM features are used for training the multi-class classification and the classification results are depicted in the next section.

3.5. Materials and Experimentation

The SVM is trained on obtained texture descriptors and the k-cross-validation scheme is used for data selection. In k-cross-validation, each time, different data is chosen randomly for both training and testing. There are two reasons for selecting $k = 10$, the first reason is the size of the dataset, and (2) the other reason is the methods used for comparison also have these criteria for training classifier. Therefore, the same parameter should be used for a fair performance comparison.

The accuracy of the SVM classifier changes in every training and testing session, due to a random choice of training and testing data in each run. Accordingly, the method replicated 100 times to get average results for all features. Performances of the SVM classifier has shown in the results section with the performance compared with existing image representation techniques.

3.6. Endoscopy Data-sets

Publicly available datasets containing CH, VE and, CLE images of multiple patients are used for experimentation. The CH dataset has 176 (experts annotated) CH image in two sets (normal and abnormal). Moreover, these images can be analyzed according to the taxonomy given by [55]. This image dataset is a part of an open biomedical challenge referred as classification of images to detect abnormalities in endoscopy (AIDA-E). Images in the dataset have three possible conditions: normal, metaplasia, and dysplasia. Normal images are in the group I (denoted as the normal class) and the frames with two conditions (metaplasia and dysplasia) belong to group II and III respectively. Images referred to as negative and positive (normal and abnormal) classes have shown in Fig. 2. In CH dataset, 56 CH images belong to normal groups whereas 120 images are from abnormal class. It contains 24

images with dysplasia and 96 images with metaplasia².

The kvasir-dataset includes 4,000 frames, 8 categories, 500 jpeg images per category. The endoscopic-frames saved in the distinct directories labeled respectively to the name of the group frames belong to them [56,57].

Along with public data, a private dataset obtained from INSTITUTO DE TELECOMUNICAÇÕES - IT, Non-Profit Association, headquartered in the University of Aveiro, University Campus of Santiago, Aveiro, also tested for the verification of proposed method. The Private endoscopy dataset composed of 160 images 80 normal and 80 images with abnormal conditions.

3.7. CLE datasets

Two CLE datasets are used for experimentation. First, CLE-Celiac dataset has 181 confocal pictures from 30 patients and 50 various GI areas gathered from screening managed by the Gastroenterology and Liver Services of the Bankstown-Lidcombe Hospital (Sydney, Australia). Every image split as crypt hypertrophy, villous atrophy, or normal, to raise celiac disease damage to the gastric mucosa. Labeled as: VACH (mucosa showing both villous atrophy and crypt hypertrophy), VA (mucosa showing villous atrophy), CH (mucosa showing crypt hypertrophy), NR (normal mucosa) [58]. All datasets are used separately for training and testing.

In the second CLE-Barrett dataset, 262 frames of 81 different are of 32 patients were obtained at the European Oncological Institute (IEO, Milan, Italy) and Veneto Institute of Oncology (IOV, Padova, Italy). Each image classified according to disease outcome: metaplasia (GMP), intestinal metaplasia or proper Barrett's esophagus (BAR), or neoplasia (NPL) [59].

While regular clinical screening of patients with Barrett's esophagus, using a confocal laser endoscope (EC-3870CIFK; Pentax, Tokyo, Japan), aiding concurrent video endoscopy and endomicroscopy.

3.8. Experimental Setup

The publicly available images datasets have a division of by their authors, into train and test sets. Nevertheless, when the data randomly partitioned, the performance of classifier fluctuates. Therefore, after the feature extraction, a multi-class SVM classifier is trained via 10-cross validation with 100 iterations to get good estimation of results. In k-cross-validation, the images divided into randomly selected k subsets, and one subset is employed for training and leftover data is used for testing. It is worthwhile to mention that the accuracy of classifier will also vary when data is selected randomly. Subsequently, this experiment repeated over one hundred times to get the mean accuracy. The discussion section shows the average results. All the experiments performed on an Intel Core i7 octa-core with Nvidia Quadro K3100M GPU in MATLAB.

The classification performance of the SVM classifier is computed on DeepGLCM features and locally computed GLCM features. The classification performance of this system is measured by the following accuracy measures described in [60–62]:

$$Accuracy = \frac{TP + TN}{P + N} \quad (8)$$

$$AUC = \int_{-\infty}^{\infty} TPR(T)FPR'(T)dT \quad (9)$$

The area under the curve (AUC) and accuracy(ACC) are calculated to measure the performance of the proposed system. For calculation of these performance measures first the false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) are computed. The

ACC and AUC of the proposed system are calculated as described in Eq. 8 and 9 respectively.

4. Results

The results of the DeepGLCM method and the existing feature extraction techniques are presented in this section. For verification of methods, datasets of various imaging techniques are used for classification. The SVM learning model is used for the analysis of endoscopic images of different datasets.

4.1. Results on Chromoendoscopy Dataset

One dataset belongs to chromoendoscopy, the dataset contains whole frames and lesions area marked by the medical experts. Both lesion and image-level classifications are performed employing the DeepGLCM and the-state-of-the-art existing features extraction techniques. Table 1 shows the average results of the SVM classifier on different descriptors. The locally computed GLCM features LGLCM_(2x2) achieved higher performance than existing texture extraction methods³. It can be seen from Table 1, GLCM features extracted from activation of the 2nd convolutional layer achieved the highest performance for both lesion-level and image-level classification in terms of accuracy and AUCs. Fig. 3 shows the box-plot with a margin of error for lesion level accuracy of SVM classifier on proposed and existing methods. Where Fig. 4 shows the image-level average classification accuracy of the SVM classifier by using 100 iterations of training and testing cycles.

4.2. Results on CLE Datasets

In the chromoendoscopy dataset, there are only two classes so the binary SVM was trained and tested. For more than two categories multiclass SVM is used. The CLE datasets, CLE-Barret and CLE-Celiac both have 3 class labels. Table 2 shows the performance of SVM on CLE dataset. The CLE images contain only gray-scale images. Images with gray-scale lack colors but contain some patterns or texture which can be estimated through statistical features efficiently.

In Fig. 6, the average performance of different features has been depicted on CLE-Celiac datasets. Similarly, Fig. 5 shows the box-plot of the average results of the SVM classifier using different texture descriptors on the CLE-Barret dataset. The results confirm good performance of the proposed technique using both accuracy and AUCs for CLE

Table 1

Average Accuracy and AUC of SVM Classifier by using different Traditional and DeepGLCM texture features on Chromoendoscopy Dataset.

| Dataset Features | CH-Lesion | | CH-Whole | |
|-------------------------|-------------------|-----------------|-------------------|---------------------|
| | ACC | AUC | ACC | AUC |
| GHT [21] | 83.01%±0.1 | 0.7072±0.001 | 85.98%±0.2 | 0.9167±0.001 |
| GLCM [63] | 85.38%±0.2 | 0.9102±0.002 | 81.63%±0.3 | 0.8554±0.002 |
| HT [13] | 84.03%±0.2 | 0.8998±0.001 | 81.19%±0.2 | 0.8821±0.001 |
| G2LCM [22] | 72.49%±0.2 | 0.6348±0.001 | 78.37%±0.2 | 0.6156±0.001 |
| Harlick [15] | 83.83%±0.2 | 0.8927±0.002 | 80.62%±0.2 | 0.8108±0.002 |
| LBP [64] | 88.76%±0.2 | 0.9461±0.001 | 81.45%±0.2 | 0.8975±0.002 |
| L-GLCM _(2x2) | 87.75%±0.2 | 0.9087±0.001 | 87.15%±0.2 | 0.9374±0.001 |
| L-GLCM _(4x4) | 78.77%±0.1 | 0.5327±0.001 | 84.89%±0.2 | 0.8156±0.001 |
| L-GLCM _(8x8) | 65.91%±0 | 0.5167±0 | 72.22%±0.1 | 0.6536±0.001 |
| GLCMConv1 | 90.87%±0.1 | 0.9473±0.001 | 87.47%±0.1 | 0.9494±0.001 |
| GLCMConv2 | 91.93%±0.1 | 0.9665±0 | 89.85%±0.2 | 0.9669±0.001 |
| GLCMConv3 | 88.11%±0.1 | 0.9649±0.001 | 88.09%±0.2 | 0.9658±0.001 |
| GLCMConv4 | 86.7%±0.1 | 0.9476±0.001 | 85.8%±0.2 | 0.9601±0.001 |
| GLCMConv5 | 82.15%±0.1 | 0.9078±0.001 | 84.26%±0.2 | 0.9241±0.001 |

² <https://aidasub-chromogastro.grand-challenge.org/description/>

³ Answer to Q4.

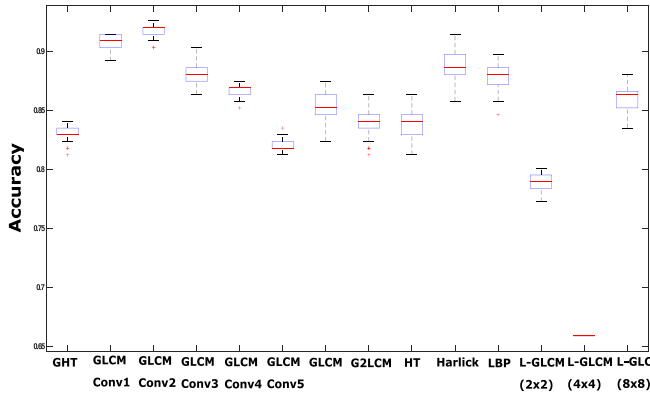


Fig. 3. Box plot showing average performance of traditional and DeepGLCM texture features on classification of chromoendoscopy lesions.

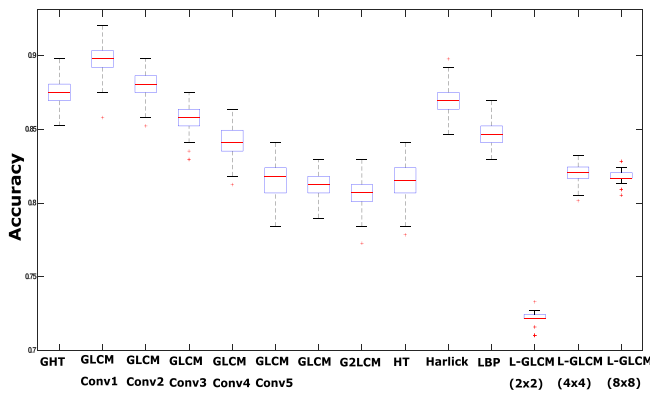


Fig. 4. Box plot showing average performance of traditional and DeepGLCM texture features on classification of chromoendoscopy images.

Table 2

Average Accuracy and AUC of SVM Classifier by using different Traditional and DeepGLCM texture features on CLE Datasets.

| Dataset Features | CLE-barrett | | CLE-Cleiac | |
|-------------------------|-------------------|---------------------|-------------------|--------------------|
| | ACC | AUC | ACC | AUC |
| GHT [21] | 82.1%±0.1 | 0.920±0.001 | 81.59%±0.2 | 0.908±0.002 |
| GLCM [63] | 77.47%±0.2 | 0.921±0.001 | 79.2%±0.3 | 0.849±0.003 |
| HT [13] | 81.23%±0.2 | 0.934±0.001 | 83.41%±0.2 | 0.910±0.001 |
| G2LCM [22] | 73.65%±0.1 | 0.684±0.001 | 72.54%±0.2 | 0.578±0.001 |
| Harlick [15] | 77.6%±0.2 | 0.923±0.001 | 75.91%±0.3 | 0.821±0.002 |
| LBP [64] | 80.71%±0.2 | 0.940±0.001 | 84.99%±0.2 | 0.917±0.001 |
| L-GLCM _(2×2) | 77.38%±0.2 | 0.938±0.001 | 80.8%±0.3 | 0.873±0.002 |
| L-GLCM _(4×4) | 75.25%±0.2 | 0.839±0.001 | 80.08%±0.3 | 0.808±0.002 |
| L-GLCM _(8×8) | 68.66%±0.1 | 0.463±0.001 | 74.83%±0.2 | 0.459±0.002 |
| GLCMConv1 | 81.8%±0.1 | 0.986±0 | 82.55%±0.2 | 0.901±0.002 |
| GLCMConv2 | 89.17%±0.1 | 0.9912±0.001 | 84.61%±0.2 | 0.927±0.001 |
| GLCMConv3 | 90.85%±0.1 | 0.9917±0 | 84.54%±0.2 | 0.946±0.002 |
| GLCMConv4 | 88.63%±0.1 | 0.9856±0.001 | 82.93%±0.3 | 0.939±0.002 |
| GLCMConv5 | 83.47%±0.1 | 0.9811±0.001 | 82.1%±0.3 | 0.942±0.001 |

datasets. Results shows that the DeepGLCM texture extraction worked well for color as well as in gray-scale images.

4.3. Results on Video Endoscopy Datasets

After testing the proposed method on CH and CLE datasets, white light video endoscopy frames (VE) are used for training and validation of proposed system.

Table 3 shows the performance of the trained classifier on different descriptors which are extracted from VE images. Three types of images

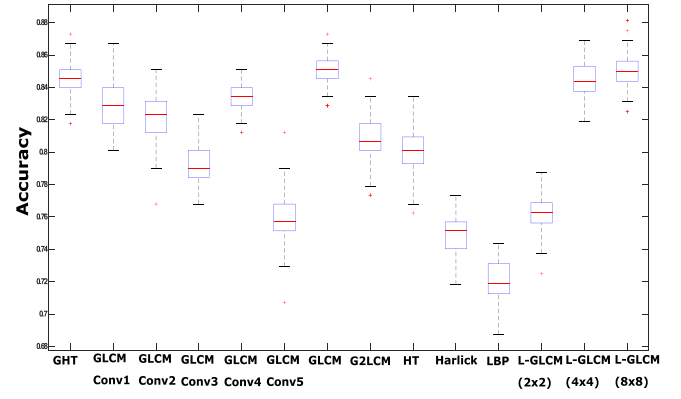


Fig. 6. Box plot showing average performance of traditional and DeepGLCM texture features on classification of CLE-Ceileac disease images.

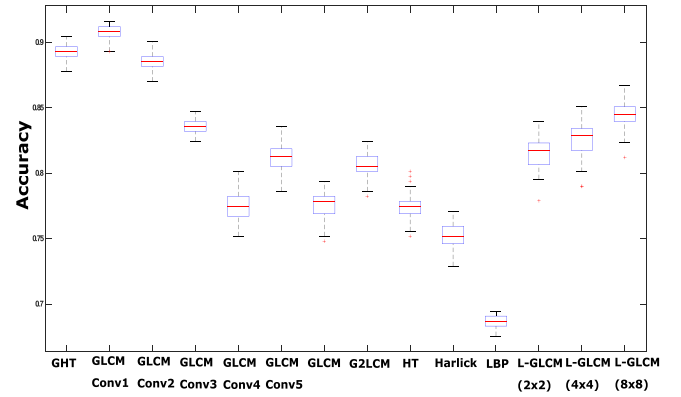


Fig. 5. Box plot showing average performance of traditional and DeepGLCM texture features on classification of CLE-Barrett esophagus images.

are used one is public dataset as described previously and lesions and whole frames from a private dataset. Fig. 7 shows the average accuracy of the proposed method along with existing methods. Similarly, a box plot of these methods with whole frames is shown in Fig. 8. The effectiveness of the proposed method is shown over publicly eight class data set in Fig. 9. It is clear from box plots and tables that the proposed technique is effective for different types of imaging modalities in terms of ACC and AUCs.

The GLCM texture features are extracted from whole images then we have divided image into 2×2 , 4×4 , and 8×8 blocks respectively and GLCM features extracted from every block. These features are denoted by L-GLCM_(2×2), L-GLCM_(4×4), and L-GLCM_(8×8), employed for the classification of endoscopy frames.

The GLCM texture features described in Eq. 1 are extracted from whole images. Then, the multi-class classifier is trained on extracted features for performance comparison.

Afterward, to extract the L-GLCM features, the frames are divided into sub-frames or patches and first, images are divide in 2×2 blocks and a total number of 4 patches. Then GLCMs are computed from every sub-image. After finding GLCM for every block, statistics as given in Eq. 2–5 are calculated to represent texture features from every endoscopy images. The features obtained from every sub-image are concatenated with each other to represent a feature vector for an endoscopy image.

The average classification performance of DeepGLCM is manifested in Tables 1 on the CH dataset. Table 1 shows the performance of the SVM classifier, it achieves $\approx 87\%$ ACC and 0.90 AUC and it has higher accuracy than classifier which is trained on traditional GLCM feature, as depicted in Table 1. Table 2 shows the effectiveness of L-GLCM_(2×2) which outperforms in terms of both ACC and AUC on the CLE dataset.

Table 3

Average Accuracy and AUC of SVM Classifier by using different Traditional and DeepGLCM texture features on Private and Public White Light Video Endoscopy Datasets

| Dataset | Private-Lesion | | Private-Whole | | Kvasir | |
|-------------------------|-------------------|---------------------|-------------------|---------------------|-----------------|-----------------|
| | ACC | AUC | ACC | AUC | ACC | AUC |
| GHT [21] | 72.04%±0.2 | 0.604±0.001 | 67.26%±0.3 | 0.7562±0.003 | 59.99%±0.1 | 0.9471±0 |
| GLCM [63] | 63.79%±0.3 | 0.6797±0.003 | 65.42%±0.4 | 0.6775±0.004 | 44.6%±0.1 | 0.8011±0 |
| HT [13] | 79.8%±0.2 | 0.8635±0.001 | 67.79%±0.3 | 0.7615±0.002 | 58.57%±0.1 | 0.9449±0 |
| G2LCM [22] | 63.54%±0.2 | 0.6096±0.001 | 54.21%±0.3 | 0.5746±0.002 | 36.23%±0 | 0.6131±0 |
| Harlick [15] | 62.68%±0.3 | 0.6818±0.003 | 65.12%±0.4 | 0.7004±0.003 | 41.4%±0.1 | 0.7865±0.001 |
| LBP [64] | 71.82%±0.3 | 0.7677±0.002 | 70.59%±0.3 | 0.7234±0.002 | 57.08%±0.1 | 0.9489±0 |
| L-GLCM _(2×2) | 62.42%±0.4 | 0.61±0.003 | 61.76%±0.4 | 0.663±0.003 | 48.08%±0.1 | 0.8078±0.001 |
| L-GLCM _(4×4) | 56.76%±0.2 | 0.4832±0.001 | 66.96%±0.4 | 0.6704±0.002 | 60.42%±0.1 | 0.8924±0 |
| L-GLCM _(8×8) | 50%±0 | 0.5±0 | 61.67%±0.2 | 0.429±0.001 | 20.81%±0 | 0.5671±0 |
| GLCMConv1 | 76.25%±0.2 | 0.873±0.001 | 72.1%±0.4 | 0.7939±0.002 | 74.42%±0.1 | 0.9678±0 |
| GLCMConv2 | 84.56%±0.2 | 0.9138±0.001 | 82.81%±0.3 | 0.8833±0.001 | 83.23%±0 | 0.9854±0 |
| GLCMConv3 | 85%±0.2 | 0.928±0.001 | 79.49%±0.2 | 0.8995±0.001 | 83.7%±0 | 0.9861±0 |
| GLCMConv4 | 85.58%±0.3 | 0.9283±0.001 | 80.37%±0.2 | 0.8984±0.001 | 84.23%±0 | 0.986±0 |
| GLCMConv5 | 87.54%±0.2 | 0.9388±0.001 | 78.14%±0.3 | 0.8635±0.002 | 84.34%±0 | 0.9862±0 |

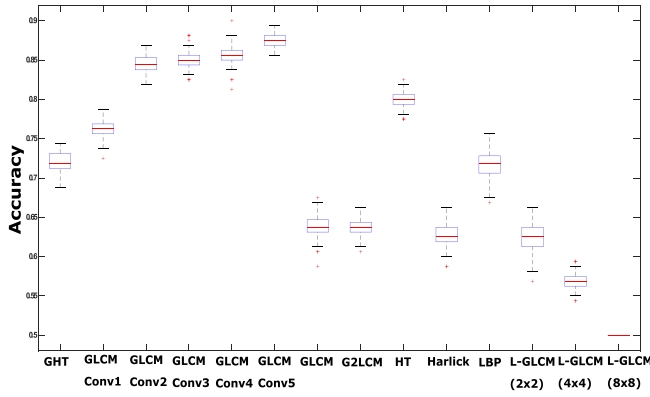


Fig. 7. Box plot showing average performance of traditional and DeepGLCM texture features on classification of the lesion of frames in private dataset.

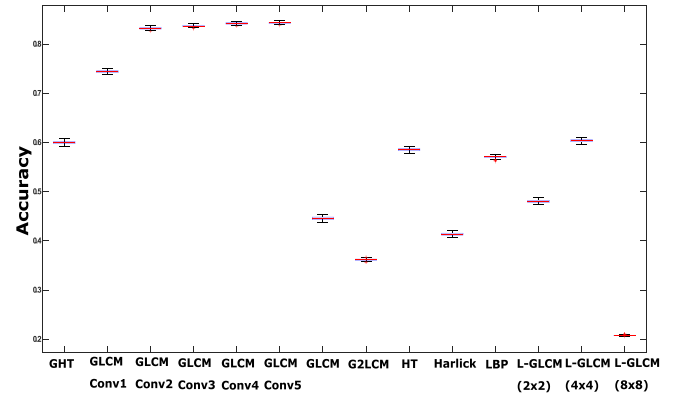


Fig. 9. Box plot showing average performance of traditional and DeepGLCM texture features on classification of Kvasir dataset.

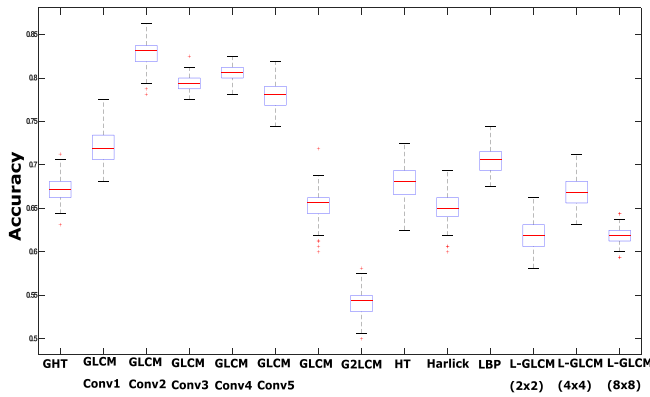


Fig. 8. Box plot showing average performance of traditional and DeepGLCM texture features on classification of the whole frames in private dataset.

The SVM classifier achieves $\approx 77\%$ and $\approx 90\%$ respectively for both CLE datasets.

Here, we have seen from the above division of images, the classification performance has improved to some extent. We were curious about how classification results will improve if we divide the endoscopy image into smaller sub-images. Therefore, for further verification, we have divided images into smaller blocks. Then the L-GLCM_(4×4) texture descriptors are extracted from the CH frames.

In the same way, every video image from the dataset is divided into 16 blocks. The texture features from every sub-image are extracted. We

have denoted these textures features by L-GLCM_(4×4). After computing texture features, the SVM classifier is trained on extracted descriptors.

The results show that the L-GLCM_(4 × 4) has a lower performance than its variant L-GLCM_(2 × 2) as presented in Table 1–3. We have divided the endoscopy images into 8×8 blocks. The GLCM is computed from every sub-image, then texture is computed statistically as it is mentioned earlier section.

The performance of L-GLCM_(8 × 8) is lower than the other two variant L-GLCM_(2 × 2) and L-GLCM_(4 × 4) when they are used for training of a SVM classifier as depicted in Table 1–3.

The SVM classifier has a good average performance as 88.7% ACC and 0.927 AUC. Moreover, all other classifiers on L-GLCM_(8×8) have low performance than the SVM classifier. However, we can dwell here on selection of block size, all classifiers have the highest ACC when the block size is selected as 8×8 . The drawback of dividing into smaller blocks is first the GLCM computation restriction from a small image. Secondly, the dimension of features can also increases causing problems such as under-fitting of classifiers.

After all above the experimentation, DeepGLCM texture features are extracted from endoscopy frames. The DeepGLCM texture features are computed from the convolutional layers of DCNN and the statistical values of each filter response in recorded as texture features.

Table 1 shows the performance of DeepGLCM texture features from the all convolutional layer of DCNN. The accuracy of these features computed by using SVM classifier. *The performance of DeepGLCM features*

extracted by using the second layer is better than other layers and 3rd convolutional layer of DCNN⁴. The performance of DeepGLCM features extracted by using the 3rd layer is not good enough than 2nd layer for the CH dataset. However, the accuracy on 3rd layer descriptors performs well on CLE dataset and traditional video endoscopy dataset shown in Table 2 and Table 3. The performance of DeepGLCM features extracted by using the 4th layer is similar to the results of 2nd convolutional layer. The performance of DeepGLCM features extracted by using the 5th layer is worst than all convolutional layers. The reason for this downgrade performance will be highlighted in the discussion section. The average results obtained by performing the simulation 100 times by training and testing the multi-class classifier using 10 cross-validations. It is clear from average results that the classification results on the second convolutional layer are better than other convolutional layers of DCNN. The mean accuracy of texture feature extracted from 2nd convolutional layer is higher than other layers. Conv2, Conv3, and Conv4 have a better performance than the traditional features.

Table 4 shows the classification performance of two classifiers k-nearest neighbors (KNN) and Random Forest (RF). This is clear from results that SVM classifier perform better than RF and KNN on different dataset by using DeepGLCM feature of various convolutional layers.

Table 6 shows the performance comparison of existing DCNNs and proposed method. Alexnet [20], Resnet-101 Resnet-18, Resnet-50 [65], Vgg16, and Vgg19 [66] are trained using transfer learning using 0.5 holdout method. The average accuracy shows the superior performance of the DeepGLCM in public and private endoscopy datasets.

4.4. Discussion

The classification results on different types of endoscopy images via the GLCM [63] texture features are presented to see the impact of texture features when computed from the whole images and lesions. The texture can be a local property of an image because a single image can have multiple texture. As mentioned earlier, the feature extraction method L-GLCM which is only a simple modification in the GLCM, outperforms the traditional (GLCM) feature extraction method by providing local texture information more precisely⁵. It is apparent from results that the L-GLCM_(8×8) have a lower performance in terms of AUC and ACC when it is used with the given image dataset. However, by using L-GLCM_(2×2) texture features with SVM classifier comparable classification accuracy can be achieved.

We have compared our results with the existing methods which developed for the classification of gastric frames. The results of traditional texture feature extraction methods, like LBP [64] texture features, (which locally computes the texture of images by comparing central pixel value with its neighbor), Gabor filter based HT defined in [13], a modified Gabor-based texture extraction methods geometric homogeneous texture features (GHT) [21], and auto-correlation homogeneous texture(AHT) features [7] are compared with proposed DeepGLCM. In our previous method, GLCM features are combined with Gabor filters to extract GLCM features from responses of Gabor filters [22].

The comparison of performances of existing methods has given in Table 1–3. It is evident from the results that the proposed method has a higher performance than all other existing methods. It is worth mentioning here, Gabor-based methods are computation extensive because of computing Fourier transform of images (for computing texture analysis).

The existing Gabor-based texture extraction methods such as HT, AHT [7], and GHT [21] have ≈81.0%, ≈83.0%, and ≈85.0% accuracy on CH whole frames. The GLCM and LBP are texture extraction methods are from the spatial domain of images processing have accuracies as ≈81% and ≈81.5% respectively. The L-GLCM texture feature extraction

method have different accuracies when used for the training of classifier with features from different frames dividing criteria. The proposed L-GLCM_(2×2) features achieve ≈87% ACC for both image-level and classification level accuracy when these are used to train the SVM classifier.

It is worth mentioning herethat the L-GLCM features and traditional GLCM features do not exhibit better performance than existing methods such as HT and LBP on the CLE datasets. However, it worked well on the CH dataset. The proposed DeepGLCM has superior performance than existing traditional methods on all datasets⁶.

The DeepGLCM features perform better than all above described methods. The GLCM features from the second layer of DCNN has the highest performance in terms of ACC and AUC. Linear SVM classifier outperforms when trained with the DeepGLCM texture features acquired from Conv1 and Conv2 layers of DCNN.

The texture features extracted from different convolutional layers of DCNN show different performances when they are used to train the SVM classifier. As shown in Table 1 the SVM achieves ≈90% accuracy for both frame-level and lesion-level classification of CH.

The texture features extracted from 2nd and 3rd convolutional layers of DCNN show best performance for CLE images. As in Table 2, the SVM achieves ≈90% accuracy for classification of CLE-Barret images and ≈85% accuracy, which is similar to LBP features, however, the DeepGLCM method beats the performance of the traditional methods in terms of AUC measure which is ≈0.92.

We can conclude from the average performance in Table 1 and 2 that the DeepGLCM texture features best performs than all traditional methods. Here, we can also infer that 2nd, 3rd, and 4th layer hold more texture information than other layers in the Alexnet when the CH and CLE frames are provided as input in DCNN⁴.

After obtaining the results on CH and CLE datasets, the VE datasets with 8 classes and 2 classes are used to extract texture features and a SVM classifier is trained. In Table 3, 100 iterations of training and testing steps with 10 cross-validations and average performance is recorded for VE datasets.

The average results shown in Table 3, depicts highest accuracy of DeepGLCM texture features on Conv2 to Conv5 layers. We can conclude from the average results that the performance of DeepGLCM is not affected by different types of imaging modalities and convolutional layers are data independent⁵. Therefore, DeepGLCM is a good option for performing classification of VE as well as CH and CLE frames with reasonable accuracy. It is also clear from the results that the DeepGLCM method beats the L-GLCM and existing texture feature extraction methods.

In the proposed method we have used a pre-trained deep neural network and only activation of each convolutional layer is obtained and GLCM features are extracted. The average time for extraction of only GLCM features from one image is 0.0004 s. Five convolutional layers of DCNN are used for the extraction of GLCM features. The average time for each layer is different as each layer has a different number of parameters (or activations), therefore, computation time for each layer is different for the computation of DeepGLCM. DeepGLCM Conv1 to Conv5 have an average time as follows: 0.0703, 0.1472, 0.2041, 0.2037, 0.1362 respectively. Because different datasets are used in experimentation therefore, computation time for each dataset is dissimilar because of uneven numbers of training examples in these datasets. In this research, our focus was to obtain features from shallow layers of DCNN by feed-forward step and it is not computationally extensive. We have shown the training (Tra) and validation (Val) time of SVM, KNN, and RF classifier in Table 5.

In a clinical setting, any kind of imaging modalities can be used from available options for screening of gastric tract. Existing methods are tested on a specific kind of endoscopy image for a specific kind of endoscopy frame. The proposed method is not restricted to a specific screening technique, rather, it can be used for the detection of

⁴ Answer Q1

⁵ Answer to Q2 and Q6.

⁶ Answer to Q5.

Table 4

Classification Performance of DeepGLCM with KNN and Random Forest Classifiers on Different Endoscopy Images Datasets.

| Classifier | KNN | | | | |
|----------------|------------|-------------------|-------------------|-------------------|------------|
| Datasets | GLCMConv1 | GLCMConv2 | GLCMConv3 | GLCMConv4 | GLCMConv5 |
| CH-Lesion | 79.92%±0.2 | 90.62%±0.1 | 87.45%±0.1 | 85.05%±0.2 | 83.56%±0.2 |
| CH-Whole | 85.6%±0.2 | 90.19%±0.2 | 86.23%±0.2 | 88.33%±0.3 | 86.75%±0.2 |
| CLE-barrett | 83.16%±0.2 | 87.17%±0.2 | 90.55%±0.1 | 90.25%±0.1 | 85.5%±0.1 |
| CLE-Cleiac | 84.14%±0.3 | 83.34%±0.3 | 84.38%±0.3 | 86.08%±0.3 | 84.58%±0.3 |
| Private-Lesion | 75%±0.3 | 83.17%±0.3 | 83.66%±0.3 | 80.83%±0.3 | 82.08%±0.3 |
| Private-Whole | 72.24%±0.3 | 76.94%±0.3 | 71.78%±0.3 | 78.25%±0.2 | 76.06%±0.4 |
| kvasir-dataset | 64.86%±0.1 | 74.66%±0.1 | 75.37%±0.1 | 73.14%±0.1 | 73.14%±0.1 |

| Classifier | Random Forest | | | | |
|----------------|---------------|-------------------|-------------------|------------|------------|
| Datasets | GLCMConv1 | GLCMConv2 | GLCMConv3 | GLCMConv4 | GLCMConv5 |
| CH-Lesion | 82.23%±0.4 | 85.49%±0.3 | 81.9%±0.3 | 75.93%±0.3 | 74.99%±0.4 |
| CH-Whole | 81.56%±0.3 | 82.71%±0.3 | 81.43%±0.3 | 77.23%±0.4 | 77.68%±0.3 |
| CLE-barrett | 81.31%±0.2 | 84.53%±0.2 | 77.65%±0.3 | 76.19%±0.3 | 75.86%±0.3 |
| CLE-Cleiac | 79.69%±0.4 | 78.36%±0.4 | 81.87%±0.3 | 77.06%±0.3 | 76.12%±0.4 |
| Private-Lesion | 72.4%±0.4 | 75.26%±0.4 | 72.71%±0.5 | 73.32%±0.5 | 73.03%±0.5 |
| Private-Whole | 68.32%±0.5 | 73.36%±0.4 | 70.64%±0.5 | 65.98%±0.6 | 64.77%±0.6 |
| kvasir-dataset | 64.88%±0.1 | 74.66%±0.1 | 75.32%±0.1 | 73.12%±0.1 | 73.14%±0.1 |

Table 6

Accuracy Comparison of Classification Performance with the-state-of-the-art exiting Models by using Transfer Learning.

| Ref. ↓ | Dataset → | CH-Lesion | CH-Whole | CLE_barrett | Cleiac | Private-Lesion | Private-Whole | kvasir-dataset |
|--------|------------|------------|------------|-------------|------------|----------------|---------------|----------------|
| Models | | Accuracy | | | | | | |
| [20] | Alexnet | 77% | 72% | 73% | 66% | 66% | 59% | 62% |
| [65] | Resnet-101 | 77% | 86% | 91% | 91% | 60% | 54% | 78% |
| | Resnet-18 | 70% | 88% | 84% | 64% | 74% | 79% | 56% |
| | Resnet-50 | 93% | 89% | 88% | 88% | 70% | 68% | 72% |
| [66] | Vgg16 | 88% | 82% | 76% | 80% | 71.25 | 69% | 58% |
| | Vgg19 | 83% | 63% | 89% | 76% | 73.75 | 51% | 59% |
| RF | DeepGLCM | 85% | 82% | 84% | 81% | 75% | 73% | 75% |
| KNN | DeepGLCM | 90% | 90% | 90% | 86% | 83% | 78% | 75% |
| SVM | DeepGLCM | 92% | 90% | 91% | 85% | 87% | 83% | 84% |

Table 5

Training and Validation Time (in seconds) of SVM, KNN, and Random Forest Classifiers on Different datasets.

| Seconds | SVM | | KNN | | RF | |
|----------------|------|------|------|------|------|------|
| | Tra | Val | Tra | Val | Tra | Val |
| CH-Lesion | 0.04 | 0.05 | 0.03 | 0.04 | 0.27 | 0.29 |
| CH-Whole | 0.05 | 0.05 | 0.02 | 0.05 | 0.29 | 0.29 |
| CLE-barrett | 0.10 | 0.13 | 0.04 | 0.05 | 0.33 | 0.30 |
| CLE-Cleiac | 0.05 | 0.05 | 0.04 | 0.04 | 0.29 | 0.28 |
| Private-Lesion | 0.05 | 0.05 | 0.02 | 0.04 | 0.25 | 0.29 |
| Private-Whole | 0.06 | 0.05 | 0.03 | 0.04 | 0.27 | 0.28 |
| kvasir-dataset | 6.77 | 5.60 | 0.20 | 4.08 | 3.01 | 0.47 |

abnormalities from images acquired from any of the three different options such as CLE, VE, and CH. Early detection of cancer or other abnormalities in the gastric tract reduces the mortality rate. However, in case of a doctor having too many patients for screening, there is a high chance of miss detection considering medical experts can be tired. Therefore, this method can be used by doctors for detection of abnormal frames in video endoscopy and can also provide a second opinion on doctor decisions.

From all above discussion, it can be concluded that traditional handcrafted features are equally important as modern deep learning models. In proposed the method, we have combined certainty with a little bit of uncertainty of deep learning models. This intuition has benefited us with a good performance. Therefore, the proposed deep learning methods can be helpful in diagnosis of gastric abnormalities with a certain confidence. *Proposed descriptors have good detection accuracy for gastric abnormalities from different kinds of endoscopy imaging*

techniques⁶.

Main findings of this research are summarized below:

- Shallow layers of Alexnet contains most of the texture information which helps in the detection of abnormalities in gastric images.
- Locally computed texture can be useful for the discrimination of gastric images. However, a certain division of subimages is acceptable. Too small sub-images can cause poor performance due to the high dimension of feature vectors.
- Texture features are good descriptors for the detection of gastric abnormalities from all kinds of gastric modalities or images.
- Our hybrid approach outperforms the state-of-the-art deep learning models.
- DeepGLCM texture features can be helpful in the diagnosis of gastric abnormalities with a certain confidence.

4.4.1. Strengths and Weaknesses of Proposed Solution

The proposed method gives rotation, scale, and spatial invariant texture features to represent endoscopy images. The proposed method is less computation-intensive and simpler than wavelet methods where Fourier analysis is the need for extraction of textures [22]. Training a DCNN is also computationally extensive and requires lots of data from training. Moreover, feature learning methods like whole DCNN require hardware and a huge amount of labeled data which is an issue in the field of gastric abnormality detection. In this work, we have shown how a DCNN can be efficiently used for a small dataset (e.g., CH dataset with 176 images). Also, we tested this method on a large data set [56].

Deeper layers of DCNN have more parameters than shallow layers,

that's why these layers have a larger dimension of feature vector than shallow layers. Smaller filter sizes in the convolutional layers directly impact the dimension of the feature vector. First issue is the number of sub-divisions of images before the extraction of features. If the number of blocks increased, the complexity and feature dimensionality are the bottleneck for this method. However, this problem can be resolved with efficient feature selection methods (e.g., [21] used the Genetic algorithm for important feature selection). The second issue is with L-GLCM, i.e., the lack of illumination invariance. These features only preserve texture information, however, color information can be useful in the case of chromoendoscopy and VE images. In the DeepGLCM method, only layers of the Alexnet model are selected for experimentation where other deep learning models can be used in the same way as this model is utilized. In this work, only four statistical features are extracted from the GLCM of shallow convolutional layers and other statistical features are not tested in this work.

4.4.2. Future Directions

In this study, we have seen the impact of locally computed texture features by dividing images into sub-images on the classification of abnormal CH images. However, these texture features do not preserve any color information, which is also a crucial descriptor for classification. In the case of chromoendoscopy, where color dyes are used to highlight mucosa, the importance of colors also increased. This paper focuses on classifying CH images with gastric cancer and CLE with Celiac and Barret diseases based on texture descriptors. These features can also be used for the classification of images of other endoscopy modalities like wireless capsule endoscopy or narrow-band imaging frames. The classification of endoscopy images can be extended by analyzing images in various stages of cancer. This work can also be extended by using these features in segmentation. We have used a basic DCNN model, the AlexNet. One future direction is to test out the latest models like DenseNet, Inception V3, and MobNet.

5. Conclusions

Early detection of gastric cancer via an endoscopic screening procedure is crucial for both gastroenterologists and patients. The endoscopy helps the gastroenterologist in visualizing the gastric mucosa. It is a tedious task to carefully inspect each frame. A DeepGLCM texture extraction scheme is proposed for the analysis of video endoscopy frames in their respective class. The experimental outcomes reveal that the suggested method is viable for texture extraction and provides $\approx 92\%$ accuracy on chromoendoscopy dataset, $\approx 85\%$ accuracy on CLE-Barret dataset, $\approx 90\%$ on CLE-Celiac dataset and $\approx 85\%$ accuracy on video endoscopy dataset with 8 classes and $\approx 87\%$ lesion-level and $\approx 84\%$ image-level accuracy. The classification of endoscopy frames into a respective category is also performed based on locally computed gray-level co-occurrence matrix (L-GLCM) texture representation. The texture of the CH, VE, and CLE images are represented by dividing images into multiple blocks of the same sizes. These experiments are looped over one hundred iterations to get more concrete results. Thus, these methods can be used for providing aid to medical experts with good accuracy and it can also be used for compiling abnormal frames from an endoscopic session.

CRedit authorship contribution statement

Hussam Ali: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation. **Muhammad Sharif:** Supervision. **Mussarat Yasmin:** Validation. **Mubashir Husain Rehmani:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank Dr. Farhan Riaz, Assistant Professor, CE&ME, National University of Sciences and Technology (NUST) Pakistan. Dr. Farhan Riaz helped us for obtaining data from INSTITUTO DE TELECOMUNICAÇÕES - IT, Non-Profit Association, headquartered in the University of Aveiro, University Campus of Santiago, Aveiro.

References

- [1] S.E. Ridge, K.R. Shetty, D.J. Lee, Current trends and applications in endoscopy for otology and neurotology, *World Journal of Otorhinolaryngology - Head and Neck Surgery* 7 (Apr 2021) 101–108.
- [2] A. Paderno, F.C. Holsinger, C. Piazza, Videomics: bringing deep learning to diagnostic endoscopy, *Current opinion in otolaryngology & head and neck surgery* 29 (Apr 2021) 143–148.
- [3] S. Beg, K. Ragunath, Image-enhanced endoscopy technology in the gastrointestinal tract: What is available? Best Practice and Research: Clinical Gastroenterology 29 (4) (2015) 627–638.
- [4] M. Song, T.L. Ang, Early detection of early gastric cancer using image-enhanced endoscopy: Current trends, *Gastrointestinal Intervention* 3 (1) (2014) 1–7.
- [5] M. Liedlgruber, A. Uhl, Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review, *IEEE Reviews in Biomedical Engineering* 4 (2011) 73–88.
- [6] F. Riaz, M.D. Ribeiro, and M.T. Coimbra, "Quantitative comparison of segmentation methods for in-body images," in 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC, pp. 5785–5788, 2009.
- [7] F. Riaz, F.B. Silva, M.D. Ribeiro, M.T. Coimbra, Invariant gabor texture descriptors for classification of gastroenterology images, *IEEE Transactions on Biomedical Engineering* 59 (10) (2012) 2893–2904.
- [8] H. Ali, M. Sharif, M. Yasmin, and M.H. Rehmani, "Color-based template selection for detection of gastric abnormalities in video endoscopy," *Biomedical Signal Processing and Control*, vol. 56, p. 101668, 2020.
- [9] M. Pietikäinen, T. Ojala, Z. Xu, Rotation-invariant texture classification using feature distributions, *Pattern Recognition* 33 (1) (2000) 43–52.
- [10] F. Riaz, F.B. Silva, M.D. Ribeiro, M.T. Coimbra, Impact of visual features on the segmentation of gastroenterology images using normalized cuts, *IEEE Transactions on Biomedical Engineering* 60 (5) (2013) 1191–1201.
- [11] F. Riaz, A. Hassan, R. Nisar, M. Dinis-Ribeiro, M. Coimbra, Content-adaptive region-based color texture descriptors for medical images, *IEEE journal of biomedical and health informatics* 21 (1) (2017) 162.
- [12] R. Nawarathna, J. Oh, J. Muthukudage, W. Tavanapong, J. Wong, P.C. de Groen, S. J. Tang, Abnormal image detection in endoscopy videos using a filter bank and local binary patterns, *Neurocomputing* 144 (2014) 70–91.
- [13] M.T. Coimbra, J.S. Cunha, MPEG-7 visual descriptors contributions for automated feature extraction in capsule endoscopy, *IEEE Transactions on Circuits and Systems For Video Technology* 16 (5) (2006) 628–637.
- [14] C. Lima, D. Barbosa, J. Ramos, A. Tavares, L. Monteiro, L. Carvalho, Classification of endoscopic capsule images by using color wavelet features, higher order statistics and radial basis functions, in: 30th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, EMBS, 2008, pp. 1242–1245.
- [15] R.M. Haralick, K. Shanmugam, et al., Textural features for image classification, *IEEE Transactions on systems, man, and cybernetics* 3 (6) (1973) 610–621.
- [16] H. Ali, M. Sharif, M. Yasmin, M.H. Rehmani, F. Riaz, A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract, *Artificial Intelligence Review* 53 (4) (2020) 2635–2707.
- [17] G. Liu, J. Hua, Z. Wu, T. Meng, M. Sun, P. Huang, X. He, W. Sun, X. Li, and Y. Chen, "Automatic classification of esophageal lesions in endoscopic images using a convolutional neural network," *Annals of translational medicine*, vol. 8, no. 7, 2020.
- [18] X. Wei, X. Yu, B. Liu, L. Zhi, Convolutional neural networks and local binary patterns for hyperspectral image classification, *European Journal of Remote Sensing* 52 (1) (2019) 448–462.
- [19] J. Tan, Y. Gao, W. Cao, M. Pomeroy, S. Zhang, Y. Huo, L. Li, Z. Liang, Glcm-cnn: gray level co-occurrence matrix based cnn model for polyp diagnosis, in: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019, pp. 1–4.
- [20] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.
- [21] H. Ali, M. Sharif, M. Yasmin, M.H. Rehmani, Computer-based classification of chromoendoscopy images using homogeneous texture descriptors, *Computers in biology and medicine* 88 (2017) 84–92.
- [22] H. Ali, M. Yasmin, M. Sharif, M.H. Rehmani, Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images, *Computer methods and programs in biomedicine* 157 (2018) 39–47.

- [23] S.S. Sarwar, P. Panda, K. Roy, Gabor filter assisted energy efficient fast learning convolutional neural networks, in: 2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), 2017, pp. 1–6.
- [24] H. Nosato, H. Sakanashi, E. Takahashi, M. Murakawa, Method of retrieving multi-scale objects from optical colonoscopy images based on image-recognition techniques, in: IEEE Biomedical Circuits and Systems Conference: Engineering for Healthy Minds and Able Bodies, BioCAS, 2015, pp. 1–4.
- [25] S.E. Martinez-Herrera, Y. Benezeth, M. Boffety, J.F. Emile, F. Marzani, D. Lamarque, F. Goudail, Identification of precancerous lesions by multispectral gastroendoscopy, *Signal, Image and Video Processing* 10 (3) (2016) 455–462.
- [26] A. Vécsei, T. Fuhrmann, A. Uhl, Towards automated diagnosis of celiac disease by computer-assisted classification of duodenal imagery, in: 4th IET International Conference on Advances in Medical, Signal and Information Processing, IET, MEDSIP, 2008, pp. 1–4.
- [27] G.D. Magoulas, V.P. Plagianakos, M.N. Vrahatis, Neural network-based colonoscopic diagnosis using on-line learning and differential evolution, *Applied Soft Computing* 4 (4) (2004) 369–379.
- [28] B. Li, M.Q.H. Meng, Tumor CE image classification using SVM-based feature selection, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2010, pp. 1322–1327.
- [29] B. Li, M.Q.-H. Meng, Small Bowel Tumor Detection for Wireless Capsule Endoscopy Images Using Textural Features and Support Vector Machine, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2009, pp. 498–503.
- [30] B. Li, M.Q.-H. Meng, Texture analysis for ulcer detection in capsule endoscopy images, *Image and Vision Computing* 27 (9) (2009) 1336–1342.
- [31] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, A. Vécsei, Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images, *Pattern Analysis and Applications* 12 (4) (2009) 407–413.
- [32] Z. Zhang, L. Bai, P. Ren, E.R. Hancock, High-order graph matching kernel for early carcinoma eus image classification, *Multimedia Tools and Applications* 75 (7) (2016) 3993–4012.
- [33] Z. Sobri, H. Amylia, M. Sakim, Texture Color Fusion Based Features Extraction for Endoscopic Gastritis Images Classification, *International Journal of Computer and Electrical Engineering* 4 (5) (2012) 674–678.
- [34] B. André, T. Vercauteren, A. Perchant, A.M. Buchner, M.B. Wallace, N. Ayache, "Introducing space and time in local feature-based endomicroscopic image retrieval," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 5853 LNC (2010) 18–30.
- [35] B. Giritharan, X. Yuan, J. Liu, B. Buckles, J. Oh, S.J. Tang, Bleeding detection from capsule endoscopy videos, in: 30th Annual International Conference of the IEEE in Engineering in Medicine and Biology Society, EMBS, 2008, pp. 4780–4783.
- [36] S. Zhang, W. Yang, Y.L. Wu, R. Yao, S.D. Cheng, Abnormal region detection in gastroscopic images by combining classifiers on neighboring patches, *International Conference on Machine Learning and Cybernetics* 4 (2009) 2374–2379.
- [37] S. Gross, T. Stehle, A. Behrens, R. Auer, T. Aach, R. Winograd, C. Trautwein, J. Tischendorf, A comparison of blood vessel features and local binary patterns for colorectal polyp classification, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2009, 72602Q–72602Q.
- [38] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," *Bildverarbeitung für die Medizin*, pp. 346–350, 2009.
- [39] P. Spyridonos, F. Vilarino, J. Vitrià, F. Azpiroz, P. Radeva, Anisotropic feature extraction from endoluminal images for detection of intestinal contractions, *International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI* 9 (2006) 161–168.
- [40] D.K. Iakovidis, D.E. Maroulis, S.A. Karkanis, A. Brokos, A comparative study of texture features for the discrimination of gastric polyps in endoscopic video, in: IEEE Symposium on Computer-Based Medical Systems, 2005, pp. 575–580.
- [41] P. Sasmal, M. Bhuyan, Y. Iwahori, K. Kasugai, Colonoscopic polyp classification using local shape and texture features, *IEEE Access* 9 (2021) 92629–92639.
- [42] F. Rustam, M.A. Siddique, H.U.R. Siddiqui, S. Ullah, A. Mehmood, I. Ashraf, G. S. Choi, Wireless capsule endoscopy bleeding images classification using cnn based model, *IEEE Access* 9 (2021) 33675–33688.
- [43] A. Caroppo, A. Leone, P. Siciliano, Deep transfer learning approaches for bleeding detection in endoscopy images, *Computerized Medical Imaging and Graphics* 88 (2021), 101852.
- [44] Z. Wu, R. Ge, M. Wen, G. Liu, Y. Chen, P. Zhang, X. He, J. Hua, L. Luo, S. Li, Elnet: Automatic classification and segmentation for esophageal lesions using convolutional neural network, *Medical Image Analysis* 67 (2021), 101838.
- [45] L. Zhu, D. Ji, S. Zhu, W. Gan, W. Wu, and J. Yan, "Learning statistical texture for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12537–12546, June 2021.
- [46] S. Devulapalli, A. Potti, R. Krishnan, M.S. Khan, Experimental evaluation of unsupervised image retrieval application using hybrid feature extraction by integrating deep learning and handcrafted techniques, *Materials Today: Proceedings* (2021).
- [47] S. Mirniaharikandehi, M. Heidari, G. Danala, S. Lakshmivarahan, B. Zheng, Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using ct images, *Computer Methods and Programs in Biomedicine* 200 (2021), 105937.
- [48] W. Ao, G. Cheng, B. Lin, R. Yang, X. Liu, S. Zhou, W. Wang, Z. Fang, F. Tian, G. Yang, et al., A novel ct-based radiomic nomogram for predicting the recurrence and metastasis of gastric stromal tumors, *American journal of cancer research* 11 (6) (2021) 3123.
- [49] B.T. Varghese and Akhil, "Upper Aerodigestive Tract Endoscopy During COVID-19," *Indian Journal of Surgical Oncology*, vol. 12, no. December, pp. 306–307, 2021.
- [50] Y.H. Chen, Y.K. Wang, Y.S. Chuang, W.H. Hsu, C.H. Kuo, C.W. Wu, L.P. Chan, M. T. Wu, I.C. Wu, Endoscopic surveillance for metachronous esophageal squamous cell neoplasms among head and neck cancer patients, *Cancers* 12 (12) (2020) 1–3.
- [51] P.F. Lai, X. Wu, S.H. Lan, B. Tang, H.Y. Huang, T. Hong, Anatomical study of a surgical approach through the neck to the jugular foramen under endoscopy, *Surgical and Radiologic Anatomy* 43 (Feb 2021) 251–260.
- [52] L.F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F.M. Sánchez-Margallo, and J.B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," *Artificial Intelligence in Medicine*, vol. 108, p. 101923, Aug 2020.
- [53] V. Raut, R. Gunjan, Transfer learning based video summarization in wireless capsule endoscopy, *International Journal of Information Technology* 2022 (Feb 2022) 1–8.
- [54] X. Luo, J. Zhang, Z. Li, R. Yang, Diagnosis of ulcerative colitis from endoscopic images based on deep learning, *Biomedical Signal Processing and Control* 73 (Mar 2022), 103443.
- [55] D.-R.M.A.M. Sousa, André and M. Coimbra, "Identifying cancer regions in vital-stained magnification endoscopy images using adapted color histograms," in 16th IEEE International Conference on Image Processing (ICIP), pp. 681–684, 2009.
- [56] K. Pogorelov, K.R. Randel, C. Griwodz, S.L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, P.T. Schmidt, et al., "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, pp. 164–169, ACM, 2017.
- [57] "Kvasir, <https://datasets.simula.no/kvasir>."
- [58] "Cleceliachy, [url:https://aidasub-cleceliachy.grand-challenge.org/description/](https://aidasub-cleceliachy.grand-challenge.org/description/)."
- [59] "Clebarrett, [url:https://aidasub-clebarrett.grand-challenge.org/](https://aidasub-clebarrett.grand-challenge.org/)."
- [60] J.A. Swets, Roc analysis applied to the evaluation of medical imaging techniques, *Investigative radiology* 14 (2) (1979) 109–121.
- [61] T. Fawcett, An introduction to roc analysis, *Pattern recognition letters* 27 (8) (2006) 861–874.
- [62] D. Alemayehu, K.H. Zou, Applications of roc analysis in medical research: recent developments and future directions, *Academic radiology* 19 (12) (2012) 1457–1464.
- [63] S. Lee, H. Ye, D. Chittajallu, U. Kruger, T. Boyko, J.K. Lukan, A. Enquobahrie, J. Norfleet, S. De, et al., Real-time burn classification using ultrasound imaging, *Scientific reports* 10 (1) (2020) 1–13.
- [64] S. Charfi, M. El Ansari, A locally based feature descriptor for abnormalities detection, *Soft Computing* 24 (6) (2020) 4469–4481.
- [65] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [66] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.