Computer Science and Information Systems 22(3):755–781 https://doi.org/10.2298/CSIS240628016K

Image clustering using Zernike moments and self-organizing maps for gastrointestinal tract

Parminder Kaur[1], Avleen Malhi[2], and Husanbir Pannu[3]

[1] Durham University, UK
parminder.kaur@durham.ac.uk
[2] Aalto University, Finland
avleen.malhi@aalto.fi
[3] Thapar University, India
hspannu@thapar.edu

Abstract. Typically, the image features are compared to find the similarity among the images in a content-based image clustering system. However, images with high feature similarity may be different from each other in terms of semantics. Hence, this paper proposes a novel algorithm based on unsupervised neural classifier systems for in-vivo image clustering to address the semantic gap issue. The visual features are represented using Wavelet transform and Zernike moments, and a self-organizing map is utilized for the clustering of images. The algorithm-based prototype system is trained for categorizing gastral images in the respective clusters as per the similarity. The system can be used to segment images with automatic noise reduction and rotation invariances for given images. Experiments are performed on the real gastrointestinal images obtained from a known gastroenterologist, and the results using Daubechies Wavelet Transform + Zernike Moments on LUV color scheme yield 88.3% accuracy.

Keywords: Machine learning, Self-organizing maps, Zernike moments, Wavelet transforms, Gastroenterology.

1. Introduction

Automatic image analysis and segmentation is a skilled task carried out by experienced professionals. Features in an image are used to decompose and analyze the underlying anatomy by defining a mechanical and systematic procedure. Given the explosive growth of visual information, partly due to the expansion of the Web and partly due to the introduction of sophisticated and inexpensive image capture systems, there is an urgent need to develop programs that can learn to segment and annotate. Automatic segmentation and annotation systems are among the critical areas of research and development for the next decade and beyond, and machine learning will be a vital technology in developing such systems [54], [53]. Self-organizing maps (SOM) incorporated with extended fuzzy c-means clustering have been a popular method for image segmentation as studied in [3]. It has used a discrete wavelet transform for image description for edges and lines involved in contrast variation.

The objective of the proposed study is to analyze, segment, and cluster the endoscopy images such that the trained system can be helpful for gastroenterologists in problem diagnosis of the gastrointestinal tract. The *motivation* behind the current problem selection

is its complexity in terms of image feature distribution. An example of in-vivo gastral images has been shown in Fig. 1 in which an image has been analyzed using two segmentation algorithms:

1. *Region Growing* (It has been applied in [7] to segment 2D microscopy digital images of freshwater green microalgae. In this approach, the image is segmented into multiple disjoint regions (sub-regions), and then they are merged with their nearest neighboring seeded region (to grow regions) that satisfies a predefined homogeneity criterion.);
2. (b) *2D Otsu algorithm* [47] (which employs the gray level information of each pixel and its spatial correlation information within the neighborhood). The algorithm has failed to capture the region of interest in both the cases, which is bleeding and not the dark spot.

It can be observed that it is pretty challenging to accurately segment blood due to the obscure nature of the color distribution and irregular region boundary. The red and green boundaries have captured the wrong dark region instead of the red spot ROI (region of interest). Moreover, the underlying images are dynamic, involving continuous movements of the camera in the drifting capsule, body organs, insufficient light conditions to capture texture at the region of interest, and varying luminance and noise due to food particles and body fluid. In addition, complementary metal-oxide semiconductor (CMOS) image sensors involve noise, high compression ratio, and low resolution of 256×256. If a segmentation method can enhance the classification accuracy in this confounding case, then inherently, it would also contribute to other applications of image processing. This is the reason for the underlying case study about image segmentation for gastral images. Challenges involved in image retrieval have been discussed in Table 1.

Table 1. Summary of challenges of image representation and learning

Challenge	**Elaboration**
Image invariance	Yields same image, when rotated, scaled or moved.
Noise	The 'lens' of the camera is never perfect; surrounding environment may contribute to the noise, noise could be Gaussian or distributed differently.
Representation	In terms of the optical properties of the (individual) pixels of an image – mean intensity, x-tilt, y-tilt, focus astigmatism @ 0 degree & focus astigmatism @ 45 degrees, coma & x-tilt, coma & y-tilt, spherical & focus.
Learning	For recognizing the contents of a new image having "see" similar images before.

We have used wavelet resolution which helps to remove noise and makes images scale invariant. Zernike moments have been used for image vectorization and self-organizing maps based on unsupervised learning is used to cluster images for sick and healthy classes.

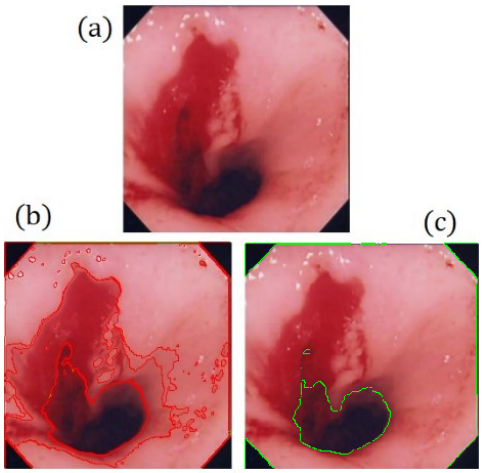


Fig. 1. (a) Original gastral image, (b) Region growing segmentation [7], and (c) 2D Otsu segmentation algorithm [47]. The red and green boundaries have captured the wrong (dark) spot instead of red region of interest (ROI) showing that problem is complex for image segmentation

Overview

The overview of the proposed research is as follows:

1. A novel algorithm for medical image clustering has been proposed which is based on unsupervised neural classifier systems.
2. The characteristic visual features are obtained from the images using Wavelet Transforms (WT), Zernike Moments (ZM), and Kohonen self-organizing feature map algorithm has been applied for clustering.
3. The proposed image clustering approach has been applied to the real capsule endoscopy images obtained from a known gastroenterologist and data distribution has been carefully studied using PCA and LDA plots to motivate the application of advanced machine learning techniques.
4. Performance analysis of the proof-of-concept model has been compared with both traditional and contemporary methods to support the belief.

This paper introduces an efficient image segmentation algorithm using Wavelet Transforms, Zernike Moments, and Linear Discriminant Analysis due to their characteristic visual feature extration and then unsupervised clustering algorithm – Kohonen self-organizing feature maps have been used to categorise the bleeding regions. For the performance comparison, ten different techniques have been executed on the dataset to justify the choice of the proposed technique.

The paper is organized as follows: Section 2 is about related works, Section 3 discusses Wavelet Transforms and Zernike moments for image vectorization, Section 4 explains single SOM, Section 5 presents the proposed method, Section 6 shows the experimental analysis, and Section 7 concludes this research study and talks about future work.

2. Literature review

This section summarizes the miscellaneous works by various researchers related to the proposed work. In [34], a comprehensive survey of computer vision techniques for wireless capsule endoscopy (WCE) has been studied. Information regarding various publicly available datasets of WCE has also been provided along with challenges and future scope. A survey has been presented in [45] for including deep learning to automate the process of WCE examination. Deep learning applications for WCE such as detecting polyps, bleeding, ulcers, hookworm, and celiac disease are discussed. A computer-aided diagnosis technique has been proposed in [10] for identifying and categorizing the abnormalities in vision-centered endoscopy detection. A novel deep sparse SVM feature selection model with group sparsity has also been incorporated, which assigns an appropriate weight to the feature dimensions and also removes the inadequate features from the feature pool. In [40], authors have utilized Zernike moments (ZM) to authenticate online signatures, and ZM represents the shape of the acceleration plot.

A novel recurrent framework has been proposed in [49] for joint unsupervised learning of deep representations and image clusters. The sequential tasks in the clustering algorithm are expressed as steps in the recurrent process, stacked on top of convolutional neural network (CNN) representations output. The research is inspired by the fact that good representation benefits image clustering, and clustering output gives supervisory indications to representation learning. Authors in [55] have proposed a Nonlinear Subspace Clustering (NSC) technique for image clustering that exposes the multi-cluster nonlinear structure of data instances using a nonlinear neural network. The technique introduced in [50] quantifies the clusterability of a dataset and is based on the probability density of a measure (S) of clusterability (in 1D) of projection of data onto a random line. After comparing the clusterability of image datasets with synthetically created clusters, it has been inferred that the structures we discover in image datasets do not fit the notion of clusters in the traditional sense. Moreover, the authors introduced a fast approach to hierarchically clustering high-dimensional data. In [8], the Deep Adaptive Clustering (DAC) approach has been proposed to represent the clustering problem as a binary pairwise classification framework for identifying whether pairs of images belong to the same cluster. The cosine distance metric has been utilized for calculating the similarities between label features of images produced by a deep convolutional network.

A novel technique, Robust learning for Unsupervised Clustering (RUC), has been introduced in [38] that is motivated by robust learning and overcomes the issues of faulty predictions and overconfident results in the case of unsupervised image clustering. This approach utilizes the pseudo-labels of existing image clustering models as noisy data that may comprise misclassified instances. In [41], the authors have proposed a two-stage deep density-based image clustering (DDC) framework to address the issue of selecting an appropriate number of clusters in advance. A pseudo-supervised joint approach has been proposed in [19] for image clustering, named Discriminative Pseudo Supervision Clustering (DPSC). Authors have resolved two significant issues in image clustering problems: appropriate image representation and lack of supervision. The main idea is to determine and use the pseudo supervision information for providing supervisory guidance for discriminative representation learning.

An improved version of ZM has been introduced in [21], which has been utilized for face recognition. In addition to the basic orthogonal and intrinsic characteristics, this ver-

sion is also invariant to noise, illumination, translation, in-plane rotation, and scaling. A hybrid similarity measure has also been proposed in this by integrating Jaccard similarity with L1 distance. Fractional-order Zernike moments, an improved version of ZM, have been presented in [23] for analyzing the grape leaf images. Multi support vector machine classifier is utilized to classify grape leaf diseases. In [25], Daubechies complex wavelet transform (DCxWT) and ZM have been used in combination for image representation. The multi-class support vector machine is used for object classification. To denoise image sequences using nonlocal means extended by ZMs, is proposed by [44]. It is found to be faster due to a reduction in weight computations, and block matching has been discounted. Similarity distance is found using photometric distance in consecutive images. A local ZM based spatio-temporal feature is proposed in [14] in the spatial domain exploiting motion change frequency for recognizing facial expressions. In [48], a study of modified principal component analysis has been performed to extract image features from the ORL face database and has been named image projection PCA (IMPCA). Sparse coded features are introduced for identifying bleeding in wireless capsule endoscopy images in [39]. These features are obtained after computing Scale-Invariant Feature Transform (SIFT) and uniform Local Binary Pattern features for WCE images. SVM is utilized for classifying the images. In [18], authors have proposed an automated system for detecting focal electroencephalogram (EEG) signals by using differencing and flexible analytic wavelet transform (FAWT) techniques. K-nearest neighbor and least squares support vector machine are applied as classifiers for automatic diagnosis.

In [4], automatic quality assessment of sperm quality (damaged or intact) has been predicted using ANN and KNN. Co-occurrence matrix and discrete wavelet transforms have been calculated from the underlying images for texture features and have been found to outperform moment-based descriptors in the study. A probability density function (PDF) based approach has been proposed in [29] for automatic detection of bleeding in WCE images. After determining the pixels of interest, local spatial features are extracted from the images by employing a linear separation scheme. In [30], an image retrieval system based upon semantic features has been studied. It uses ontological terms to define the image using multi-scale Reisz wavelets to analyze their annotation similarity. Liver lesions in CT images have been experimented with to validate the proof-of-concept. Normalized discounted cumulative gain (NDCG) score and AUC have been calculated and compared for the real-time decision-making capabilities of the model. For the robust representation of WCE images, the study given in [51] provides the assistance and discriminated definition for polyp images using a deep learning technique utilizing sparse auto-encoder. It uses a nearest neighbor graph to define inherent image manifold characteristics. A summary of the motivational literature review has been given in Table 2. In [32] a survey of large language models (LLM) have been studied for gastroentrology and semi-supervised variational models in [13].

Table 2. Summary of literature survey: pre-processing and noise removal, image representation, and learning

Sr	Method	Purpose	Outcome	Study
Pre-processing and noise removal				
1	DCxWT, ZM and multi-class SVM	Object classification	Better precision and accuracy values	Khare 2021 [25]
2	Nonlocal means extended by ZMs	Faster computation	Denoising and faster computation	Singh 2017 [44]
3	Differencing and FAWT	Automatic detection of focal EEG signals	94.41% accuracy	Gupta 2017 [18]
4	Riesz wavelets	image retrieval for a hemangioma, liver lesions	NDCG score = 0.92, AUC = 0.77	Kurtz 2014 [30]
Image representation				
5	Zernike moments	Online signature authentication	4% of False Rejection Rate, 2% of False Acceptance Rate	Radhika 2011 [40]
6	Local modified Zernike moment per unit mass	Face recognition	Higher recognition accuracies on two datasets	Kar 2020 [21]
7	Deep sparse SVM	Computer aided endoscopy diagnosis	New endoscopy dataset, Computation reduction and improved robustness	Cong 2015 [10]
8	Image principal component analysis	To analyse IMPCA is better than PCA, FDA	Better accuracy and reduced time	Yang 2002 [48]
Learning				
9	Local ZM, SVM	Facial expression recognition	Improved recognition rate	Fan 2017 [14]
10	Survey of computer vision methods for WCE	Determining major challenges of WCE and future scope	Comparative analysis	Muhammad 2020 [34]
11	Survey of deep learning for WCE	Systematic review and meta-analysis of deep learning methods for WCE	Comparative analysis	Soffer 2020 [45]
12	Sparse coded features, SVM	Detect bleeding in WCE	accuracy = 98.18%	Patel 2021 [39]
13	DWT, Invariant moments, ANN, KNN	veterinary field, spermatozoa healthy or sick	accuracy = 95%	Alegre 2012 [4]
14	Local spatial features, Rayleigh PDF model	Automatic bleeding detection in WCE images	Improved performance with less complexity	Kundu 2019 [29]
15	Stacked sparse autoencoder with image manifold constraint	polyp recognition	Overall accuracy = 98%	Yuan 2017 [51]

3. Image feature vectors

Image features involve color, texture, and shape metrics based upon the contrast-related discontinuities in the image. For this study, Wavelet Transforms [52] and Zernike moments [12] have been used due to their efficiency and power to capture the inherent characteristics.

3.1. Wavelet Transforms (WT)

These mathematical functions divide a signal (image) into different frequency components. The goal is to study each component with a resolution with a matching scale. WT

is better than Fourier Transforms (FT) or Short-Time Fourier Transform (STFT), which cannot analyze both frequency and time components [22]. Wavelet transform is composed of wavelet function $w(.)$, defined in finite time and normalized. The formula for WT is:

$$W_{f(\mu,\sigma)} = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{\sigma}} w\left(\frac{x-\mu}{\sigma}\right) dx \tag{1}$$

where (μ, σ) are translation and scaling parameters, respectively. To see lower frequency components of the signal, increase the value of σ for instance. Some prominent mother wavelets have been shown in Fig. 2. In our study, Daubechies 4 wavelet has been used (details in [11]). Spatial information comprises the image pixel positions (x, y) that act as the time axis and changes in pixel intensity $f(x, y)$ that serve as the frequency axis. Thus, edges have a higher frequency as compared to smooth areas. For Discrete WT (DWT), an image is decomposed into four components: approximation, horizontal, vertical, and diagonal. As shown in Fig. 3, the image is decomposed into one level using DWT (3a) with an example of a face image.

In our study, the image has been decomposed on three levels using WT, as shown in Fig. 4. It explains about horizontal, vertical and diagonal edges being detected in the original image. Ten components have been calculated as $\{(H_i, V_i, D_i, A_i) \mid i = 1, 2, 3$ for H, V, D and $i = 4$ for $A\}$. In expanded form, we get $H_1, V_1, D_1, H_2, V_2, D_2, H_3, V_3, D_3$, and A_3. Then for these 10 components, 12 Zernike moments have been calculated for $n = m = 5$ which are listed as $Z_{00}, Z_{11}, Z_{20}, Z_{22}, Z_{31}, Z_{33}, Z_{40}, Z_{42}, Z_{44}, Z_{51}, Z_{53}$, and Z_{55} or in the set notation $\{Z_{ij} \mid i \geq j$ and $i - |j|$ is even$\}$. After compiling all that information from LUV channels, the image feature vector has $12 \times 3 = 36$ dimensions. For example, Fig. 5 shows the results from a sample picture's approximation, horizontal, vertical, and diagonal edge detection decompositions.

Wavelet Transformation (WT) is quite useful for noise removal, image compression [52], and zooming capabilities for local characteristics of an image. It is also an efficient technique for texture characterization while preserving local and global spatial/spectral information. For instance, the noise removal feature of WT is shown in Fig. 5 with four decompositions levels, and image denoising has been illustrated in Fig. 6 for an image with a considerable amount of Gaussian noise.

3.2. Zernike moments (ZM)

Image moments are the weighted average of the intensity values of the image pixel (or a similar image function) to get the scalar quantities for image interpretation. Moments of different order yield varying information about the image, such as area, center of mass, and orientation. Zernike Moments (ZM) [16] of an image are similar to Discrete Cosine Transform (DCT) coefficients in their derivation and properties. ZM are projections of an image function along the real and imaginary axes (x-axis and y-axis), which are convolved by an orthogonal function. They represent an image in various frequency components which are referred to as the orders (along the radial) and repetitions (along the angular direction). Thus, Z_{00} represents the average intensity, Z_{11} represents the first-order moment, Z_{20} is similar to variance, and so on. Zernike polynomials are orthogonal functions that generate an orthogonal set over the unit circle in a complex plane. The center of the image stays the same as the center of the circle. Hence, a square image can be

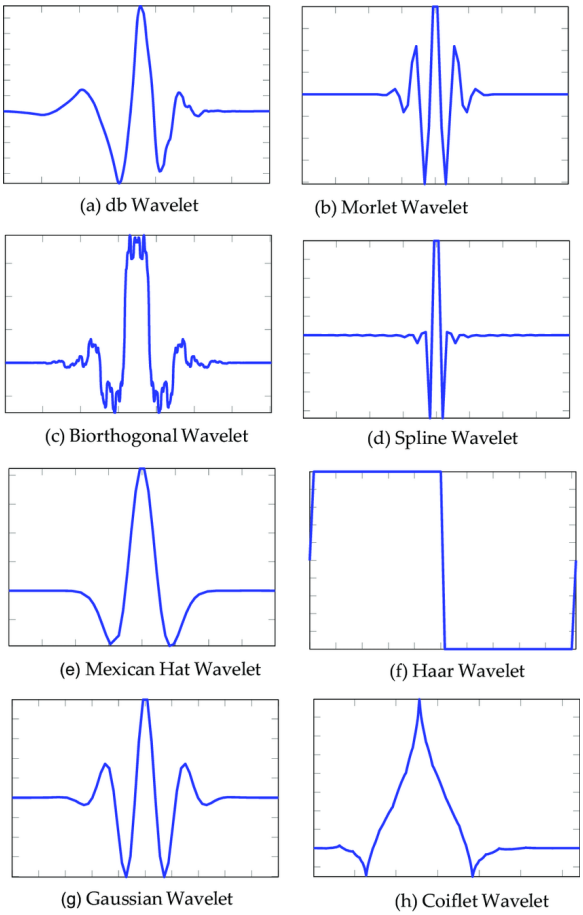
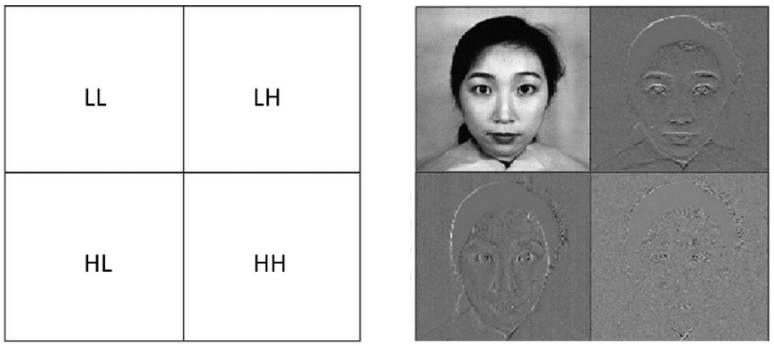


Fig. 2. A few popular mother wavelet functions [15] $w(.)$. Daubechies 4 wavelet have been utilized in the experimentation

mapped inside or outside an image [1]. In the case of inner mapping, the pixels which fall outside the unit disc must be discarded. So, to avoid the information loss from the edges, we have utilized outer mapping for our experimentation which is shown in Fig. 7.

Formula for Zernike polynomials is $V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta}$. Here n, m are whole numbers such that $n - |m| = \text{even}$, $n \geq 0$, $0 \leq \mid m \mid \leq n$, $\theta = arctan(y/x)$ and $j = \sqrt{-1}$. (ρ, θ) are radius and angle of the pixel from origin which simply means the polar coordinate of a pixel at (x, y). Formula for radial polynomial $R_{nm}(\rho)$ is given as follows:

$$R_{nm}(\rho) = \sum_{k=0}^{(n-|m|)/2} (-1)^k \times \frac{(n-k)!}{k!(\frac{n+|m|}{2}-k)!(\frac{n-|m|}{2}-k)!}\rho^{n-2k} \tag{2}$$



(a) DWT decomposition (b) DWT decomposition of face image

Fig. 3. Image decomposition using DWT with an example of face image [20]

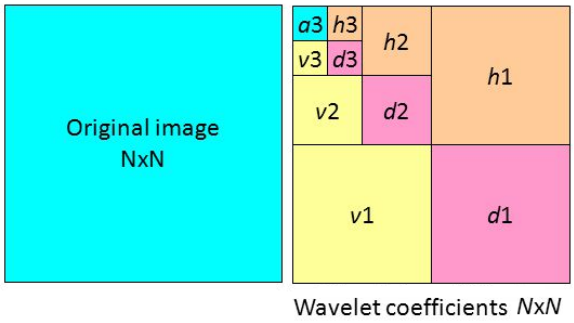


Fig. 4. Daubechies wavelet transformations are used in the experiments. a, v, h, d stands for approximation, vertical, horizontal, and diagonal details. Diagonal (low/low), horizontal (high/low), vertical (low/high), approximation (high/high)

$$Z_{nm} = \frac{n+1}{\pi} \sum_{x=1}^{N-1} \sum_{y=1}^{N-1} f(x,y) R_{nm}(\rho) e^{jm\theta} \tag{3}$$

Z_{nmx} and Z_{nmy} are cosine and sine values of Z_{nm} (Zernike moments). The corresponding value if ZM can be calculated as $Z_{nm} = \sqrt{Z_{nmx}^2 + Z_{nmy}^2}$. Rotational and scale invariance can be obtained in ZM by normalizing the image using Cartesian moments before the ZM calculation [26]. Moreover, if the center of mass of image is moved to origin then translation invariance can also be achieved.

4. Single SOM

Our method involves definitions for creating a set that associates the most active neuron for the set of the output layer of SOM, with a set of input vectors presented to the input layer of SOM as defined in Equations 6 and 7. It applies to a single SOM or can be

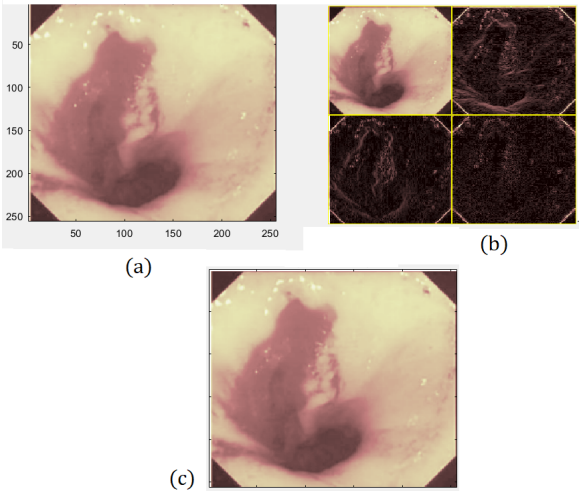


Fig. 5. The four decompositions explained with example: approximation, horizontal, vertical, and diagonal details to detect the corresponding edges. Fig. (a) is original image, (b) is the view of four decompositions, and (c) is denoised image

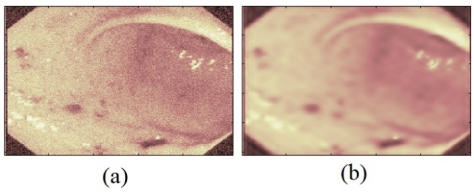


Fig. 6. (a) Noisy image and (b) denoised image with Daubechies wavelets (DB-4)

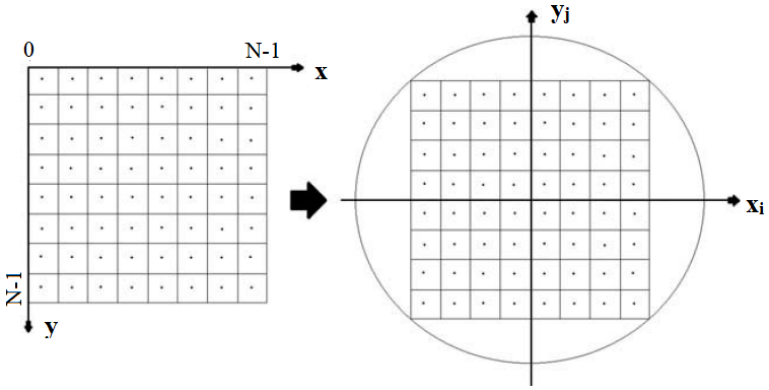


Fig. 7. Outer mapping which maps a given image inside the unit disc

extended as the collateral SOM for hybrid SOMs. Follow Algorithm 1 for creating single modal information systems for image clustering.

Algorithm 1 Algorithm for retrieving information from a single SOM

1: Identify the best match node $\vec{w}_k$.
2: Form a totally ordered set of the n nodes in the SOM, such that:

$$(W, \leq) = \left\{ \begin{array}{c} \vec{w}_k, k = 1..n \ \mid \ \vec{w}_i \leq \vec{w}_j \Leftrightarrow \\ \|\vec{x}_k - \vec{w}_i\| \leq \|\vec{x}_k - \vec{w}_j\| \end{array} \right\} \tag{4}$$

where $\vec{w}_i, \vec{w}_j \in W, 1 \leq i, j \leq n$ and $i \neq j$
3: Retrieve a totally ordered set R, of all p pre-stored items used in training, in response to the input vector $\vec{x}$

$$R_{single} = \{\vec{x}_l, l = 1..p \ \mid \ \exists \ \vec{w}_k \in W : (\vec{x}_l, \vec{w}_k) \in P_{single}\} \tag{5}$$

A Self-organizing Map (SOM) [28], also called a Kohonen Map, associates a multidimensional input space, comprising a set of feature vectors, onto a 2-dimensional surface (output map). The end of training leads to an association between an input vector $\vec{x}$ and a specific output node that 'wins' the input, known as the Best Matching Unit (BMU) for that input vector. If $\vec{w}$ represents the weight vector of an output node, then BMU for input vector $\vec{x}$ can be calculated as:

$$\|\vec{x} - \vec{w}_m\| = min\{\|\vec{x} - \vec{w}_m\|\} \tag{6}$$

where m depicts the index of SOM output node which is a BMU. One node may 'win' over more than one input forming a set. Let P_{single} be the pair set of q input vectors and the corresponding winning node is $\vec{w}_m$, then P_{single} is defined as:

$$P_{single} = \left\{ \begin{array}{c} (\vec{x}_k, \vec{w}_m), k = 1..q \\ \|\vec{x}_k - \vec{w}_m\| = \min\limits_{i=1}^{n}\{\|\vec{x}_k - \vec{w}_i\|\} \end{array} \right\} \tag{7}$$

Information retrieval from a SOM involves the presentation to the trained SOM of a set W. The mapping of the input vector from higher dimensional nodes in the output layer forming a space to the winning node in 2-D neuron space has shown in Fig. 8. The length of input vector X_i and neuron weight vector W_i must be the same. The retrieving information from a SOM has been depicted in the Algorithm 1. The following section explains image vector creation using Zernike Moments and Wavelet transformation for denoising.

During the initial stages of the SOM training, the weight vectors are initialized with random weights and then, together with the input vectors, are normalized. The learning and neighborhood rates are reduced exponentially during training following established practice in the SOM literature. Our testing regimen relies on the notion of best matching unit(s): the node(s) in the output layer that responds with the highest activation value to a given input vector. Note that if one or more neurons can be activated in response to the input vector, then the activated neurons are ordered according to their activation levels (Algorithm 1). If the category of the input vector matches the most activated neuron in the output layer, then we have a best-matching unit (BMU). If there are multiple activated nodes for a specific input then we are considering the two highly activated nodes only.

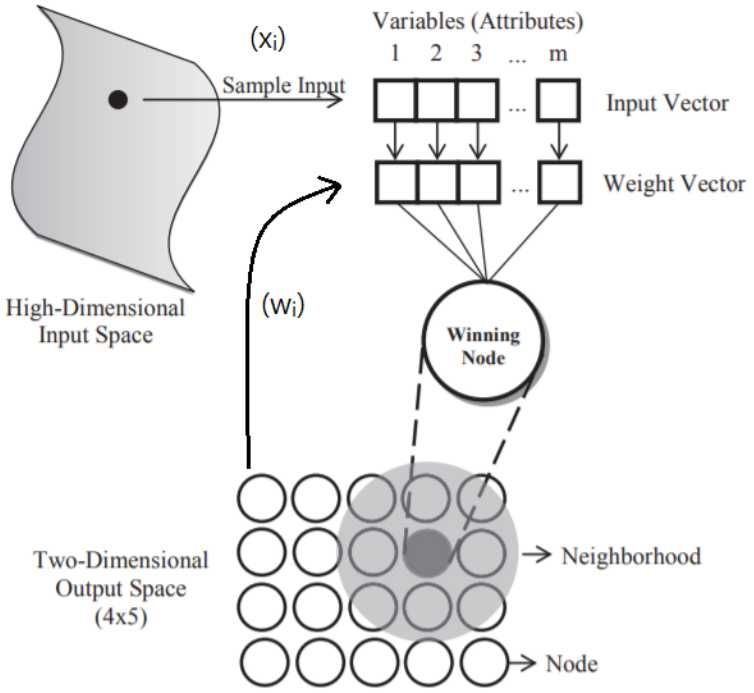


Fig. 8. Overview of single self-organizing map (SOM) model. X_i are input vectors with same length as weight vectors W_i. Each X_i is connected to every (winning) neuron

A matching matrix was created to analyze how an input vector may activate neurons that were trained to respond to one or more categories of keywords or images. If the winner or BMU in the output layer has the same category as the stimulus, and the stimulus did not excite any other neurons, then the match will be perfect. However, if a given stimulus activates neurons of various other categories, the match will be minimal. We define accuracy as the number of correctly clustered items (based upon the majority of similar items in the cluster as the test instance) divided by the total number of items in the category.

5. Proposed Methodology

The proposed research aims to effectively cluster the in-vivo gastrointestinal images based upon their similarity by carefully considering the image semantics. Let I be the training set of images that is an input to the proposed algorithm. The expected output is the trained self-organizing map and the image cluster sets (C_i) constructed as per the image similarity. The first step is to denoise the images using Daubechies wavelets with four decomposition levels: approximation, horizontal, vertical, and diagonal. The next step is the conversion of RGB to LUV channels. Wavelet transforms implementation details are

given in *Section 3.1*. Subsequently, 12 Zernike moments are calculated for each of the L, U, and V channel with $n = m = 5$, creating a total of $12 \times 3 = 36$ image vector dimensions. The ZM calculation steps and equations are mentioned in detail in *Section 3.2*. In the end, 4×4 SOM is trained using image vectors and constructs the image clusters. Algorithm 2 shows the steps for segmentation and clustering of the images using SOM.

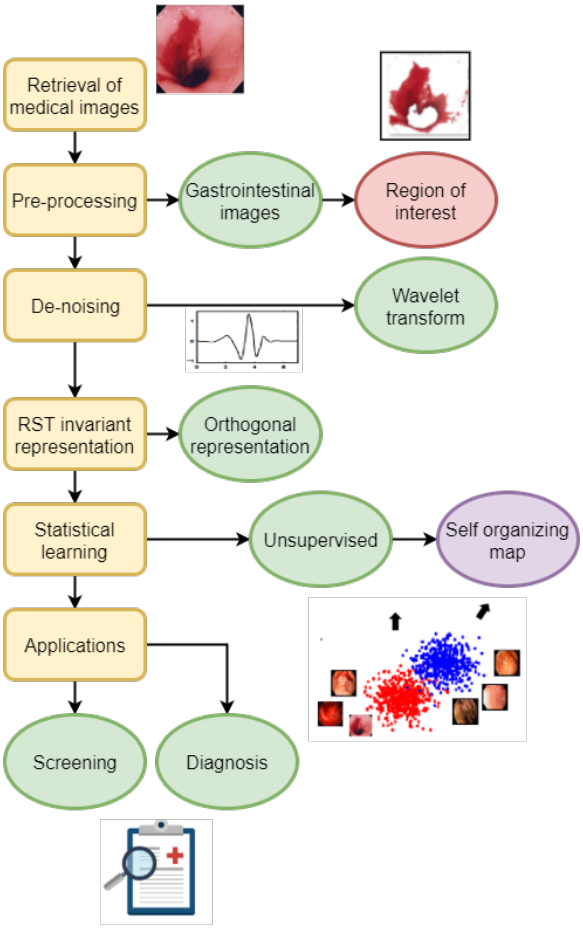


Fig. 9. Pipeline diagram for the proposed methodology

As per [37], the total number of multiplications required for computing a radial polynomial ($R_{nm}(\rho)$) using Equation 2, is almost $(n/2 + 1) \times (2n - 3) \times (n - 1)$. So, computational complexity of calculating a single $R_{nm}(\rho)$ value of order n and repetition m is $O(n^3)$. If the image dataset size is D, then the total complexity of ZM calculation becomes $O(Dn^3)$. The processing time of SOM is $O(D^2)$ [42]. So, the computational complexity of the proposed algorithm is $O(Dn^3 + D^2)$.

Algorithm 2 SOM image clustering

INPUT: Training set of images I
OUTPUT: Image cluster sets (C_i where i = number of SOM clusters and $C_i \subseteq I$)

procedure SOM TRAINING FOR ENDOSCOPY IMAGES
  Image denoising - Daubechies wavelets ($DB-4$) using four decomposition levels (a, v, h, d)
  RGB to LUV transformation
  Calculate ZM for each of L, U, and V channels with n=m=5
  (i) Calculate radial polynomial $R_{nm}(\rho)$ using eq. 2
  (ii) $V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta}$
  (iii) Z_{nmx} and Z_{nmy} are real and imaginary values of Z_{nm}
  (iv) $Z_{nm} = \sqrt{Z_{nmx}^2 + Z_{nmy}^2}$
  (v) Calculate 12 Z_{nm} for each L, U, and V channel, so total 36 elements in image vector
  Train SOM with 4×4 grid size using the obtained image vectors
  Required image clusters (C_i) are obtained after SOM training
end procedure

Fig. 9 is the pictorial representation of all the steps involved in implementing the proposed approach. Initially, we have a set of 300 raw endoscopy images. The images are pre-processed and the region of interest is identified. Afterward, the denoising of the images is performed using wavelet transforms and the RGB images are converted to LUV format. Subsequently, ZM features are extracted from the images, which are rotation, scaling, and translation invariant. The unsupervised self-organizing map is trained using the extracted image features, and the image clusters are formed based on the similarity. Now the trained system can be utilized by gastroenterologists for screening and diagnosis purposes for endoscopy images.

6. Experiments

This section includes the information regarding dataset, its analysis and results obtained using proposed approach. The configuration of the system used for experiments is: Desktop System is Dell Inc. with Model XPS 8930 with Windows 10 ProVersion 10.0.17763, Intel(R) Core (TM) i7-8700 CPU @ 3.20GHz, 3192 Mhz, 6 Core(s), 12 Logical Processor(s), with 16GB RAM.

6.1. Dataset

The dataset comprises 300 real gastrointestinal images obtained from a known gastroenterologist with a ratio of 180:120 for healthy and sick cases. All the images are of size 256×256 and are from both upper (esophagus and stomach) and lower (small bowel and colon) gastrointestinal tract. The sample images from the dataset are demonstrated in Fig. 10 and Table 3 provides the information regarding the dataset.

6.2. Data distribution analysis

The data distributions of healthy and sick image vectors have been examined from various aspects (to analyze an appropriate learning model), which are as follows:

Fig. 10. Sample gastral images for bleeding detection. Total 300 in number with ratio of 180:120 for healthy and sick. Image size is 256×256 pixels

Table 3. Description of images in dataset

Category	**Healthy**	**Sick**
Image ratio	180	120
Size	256×256	256×256
Redness	Overall	Spots or saturation

- Relative red intensity in healthy/sick images.
- Distribution of RGB intensities in all images.
- Thresholding to crop the red color (for example, $R > 100, G < 60, B < 50$).

In Fig. 11 the average red color in the sick and healthy classes has been sorted and plotted. Although all gastral images are reddish brown in color, the sick images are more saturated with redness. The intensity plots of all the images have been illustrated in Fig. 12: (a) shows that the left half has more dispersion, especially in red color and R values are relatively higher. The other three plots (b-d) show the R versus G, R versus B, and B versus G plots. There is tremendous overlap, so a simple linear regression may not be sufficient for the bleeding analysis. Thus, there is a need for a non-linear learning system (such as SOM). The red color segmentation has been experimented with using MATLAB to further analyze the problem complexity, which is illustrated in Fig. 13. The threshold values $R > 100, G < 60, B < 50$ have been chosen for best human eye subjective red color cognition through the experiments. Again, it seems quite difficult to distinguish the healthy red versus the sick red spots for confounding cases. The middle two images are healthy in these four images, and the left/right extremes are bleeding cases. Therefore, simple thresholding is also insufficient to spot the bleeding even with various threshold values.

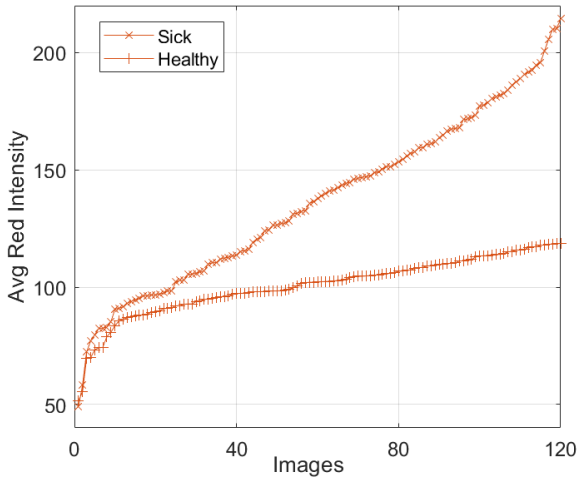


Fig. 11. Sorted average red pixel intensities for normal and abnormal images. The upper line is redness in sick images which is relatively higher as compared to healthy images

6.3. ZM extraction and SOM application

For each L, U, and V component, 12 ZM have been calculated, making a total of 36 ZMs to extract the luminance and color attributes as shown in Fig. 14. Zernike moments are rotation and noise invariant as studied in [26], which can be seen in the Figures 15 and 16. Furthermore, feature transformation techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied to these 300×36 image vectors as shown in Fig. 17. Both of these transformations are used for dimension reduction, but PCA focuses on maximizing the variance among mutually orthogonal transformed dimensions, and LDA focuses on the separability of the data concerning the labels [33]. Linear separability in the case of PCA has been found to be 74% and 84% in case of LDA. Therefore, linear separability becomes possible after extracting ZM on LUV image components. Self-Organizing Maps have been involved further to analyze the accuracy with the hope of improvement.

Fig. 18 shows the results when the model was trained using only healthy (180) points and tested for 120 sick images. It can be observed that there is some overlap of the tested sick images with healthy images due to the obscure nature of the images. But still, there is a complementary saturation between these two images showing that sick images have different data distribution on a broader scale.

Wavelet transforms applied to all the images before extracting their Zernike moments (ZM) for noise removal. In Table 4, the results of accuracy for ZM versus WT+ZM on 300×36 image vectors have been compared. Table 5 shows the difference between WT+ZM and ZM accuracy is positive on the average of 5 trials, confirming the advantage of WT application before ZM extraction. Fig. 19 is the visual illustration of Table 5. Finally, Table 6 presents the confusion matrix obtained after experiments and the best accuracy obtained using WT+ZM and Kohonen self-organized feature maps has been found to be approximately 88.3%. In the table, P_Healthy, P_Sick and A_Healthy, A_Sick are

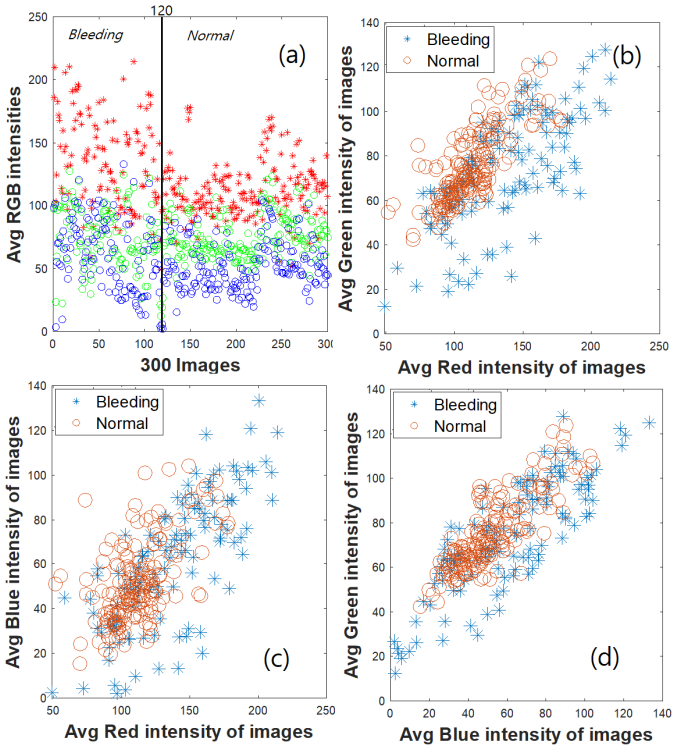
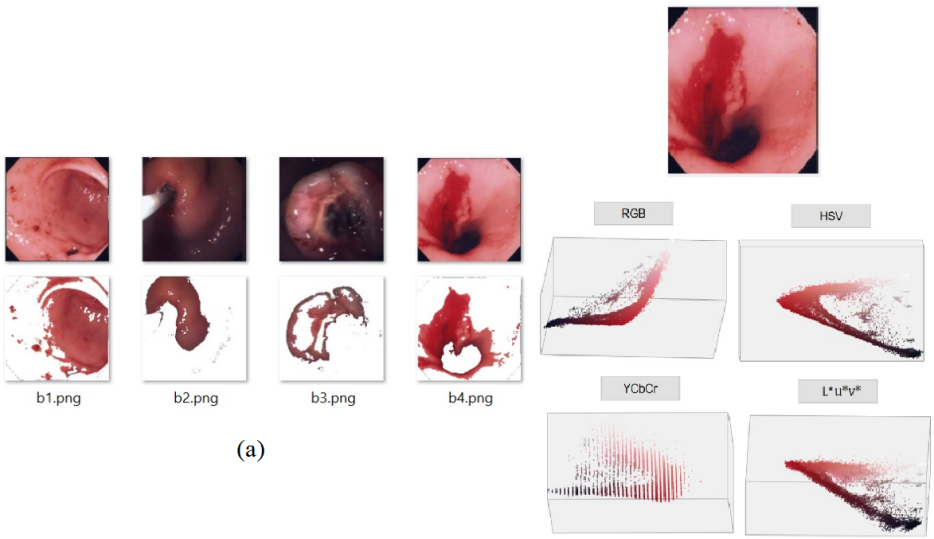


Fig. 12. (a) RGB intensity of 300 images (180 : 120 for healthy:sick cases) (b) Red vs Green average intensities (c) Red vs Blue avg. intensities (d) Blue vs Green intensities plotted for all the images. It is clear from (b-d) that none of the color intensities are easily separable for healthy and sick cases

Table 4. Accuracy given by ZM versus WT+ZM on 300 × 36 images. TR = training data, TS = testing data, VAL = validation data, and AVG = average.

	ZM				**WT+ZM**			
Trials	**TR**	**VAL**	**TS**	**All**	**TR**	**VAL**	**TS**	**All**
1	82.4	78.3	71.7	78.7	83.8	86.7	75.6	83
2	79.5	84.4	77.2	80	81.4	86.7	77.8	81.7
3	77.6	82.2	77.8	78.3	82.4	81	76.9	81
4	75.2	80.2	77.8	76.7	82.4	80	80	81.7
5	80.5	82.2	80.2	81	83.3	88.9	79.6	83
AVG	79	81.5	76.9	78.9	82.9	84.5	79.8	82.4



(a)

(b)

Fig. 13. (a) Red color cropping using subjective threshold RGB values (using MATLAB) $R > 100, G < 60, B < 50$. Hence, in order to classify the images, robust features are required to extract such as ZM. Firstly, RGB values have been converted to LUV color representation to separate the luminance component from the color composition as shown in Fig. (b)

	1	2	3	4	5	6	...	34	35	36
n1	512.4	165.4	204.8	6.6	144.3	36.6		34.5	17.8	0.2
n2	533.6	165.3	243.5	4.6	157.9	31.0		10.0	5.7	0.9
n3	457.5	141.7	217.3	1.3	143.5	19.6		31.7	11.1	1.1
n4	444.7	139.3	204.2	1.2	136.0	21.8		25.7	12.3	0.0
n5	520.9	157.2	239.8	11.5	154.9	30.0		30.1	16.6	0.6
n6	518.1	158.5	252.4	3.0	162.1	21.2		39.3	18.4	0.3
n7	476.4	150.6	218.2	5.1	146.0	22.5		30.1	8.3	0.1
n8	555.1	170.9	263.1	0.9	171.3	25.1		19.5	8.1	0.3
n9	470.4	143.6	215.9	3.8	131.2	28.2		22.5	9.8	1.1

Fig. 14. Snapshot of ZM for 9 healthy images. Rows corresponds to the image vectors and 36 columns are the Zernike moments of image. n means normal/healthy.

predicted and actual values of healthy and sick classes respectively. Table 7 shows the comparison of the other approaches with the proposed technique based on obtained accuracy. The comparison is performed with both traditional methods such as PCA and LDA (used for linear separability of data), and contemporary methods such as deep learning based approaches.

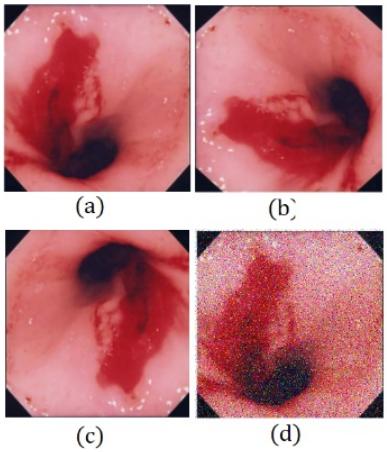


Fig. 15. (a)-(d) are original image, rotations of 90 and 180 degrees, Gaussian noise introduced

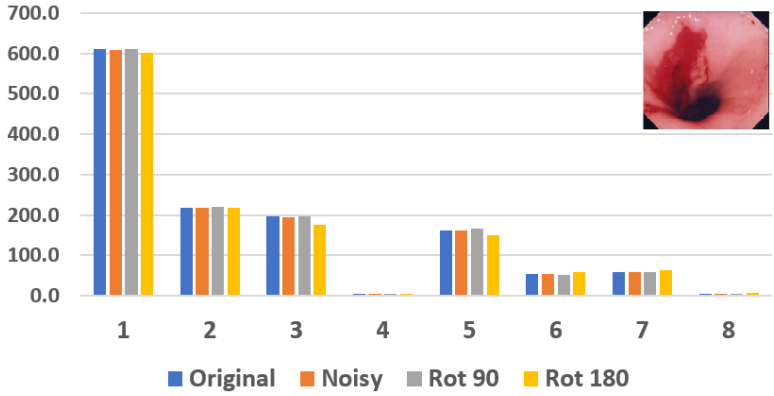


Fig. 16. ZM for all images in Fig. 15 are nearly same. Only 8 out of 36 ZM have been shown to simplify the demonstration. Slight differences are due to 3 types of errors involved in ZM calculation as studied in [1]

6.4. Incorporating Image Captions for Multi-modal Learning

In this subsection, the freely available descriptions with the gastral images are incorporated into the system to improve learning accuracy. In real-world medical diagnosis, image features are just not enough to yield the required information. For example, in Fig. 20, two confounding images have been considered which look identical; however, Fig. (a) involves bleeding, and Fig. (b) has an air bubble in case of a healthy image. Therefore, linguistic cues (provided by experts) can be associated with images to handle these kinds of cases.

Information from multiple modalities, such as images and collateral text, can be utilized simultaneously for different tasks, including classification, clustering, or object de-

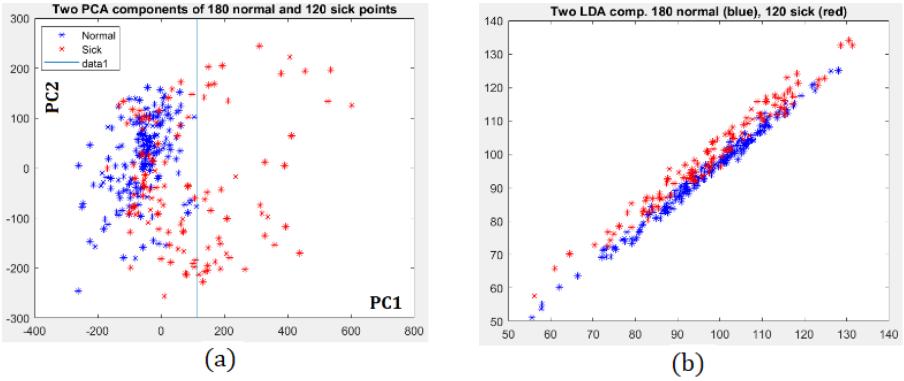


Fig. 17. PCA and LDA draws of (300) image vectors. PCA yields 74% and LDA yields 84% linear separability of 120 bleeding and 180 normal image vectors. *Blue* for healthy and *Red* for sick.

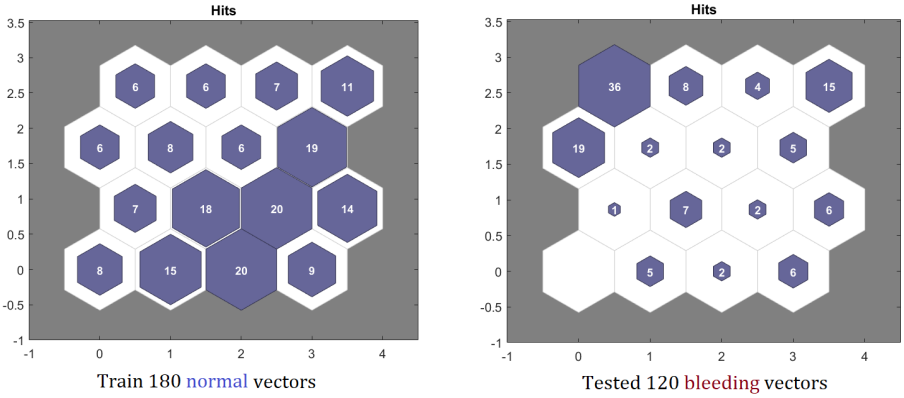


Fig. 18. SOM maps obtained upon training with 180 normal points and testing with 120 bleeding images using MATLAB. It can be observed that the mapping is different and thus distribution of 180×36 and 120×36 image vectors are different.

tection, which is known as multi-modal learning [24]. Sometimes, the clusters constructed by SOM are pretty small, which acts as the outliers. These clusters can be merged with similar bigger clusters using related textual information. The endoscopy images are accompanied by corresponding labels or small text that provides some description of them. A SOM is trained with this linguistic information similar to the SOM trained with image data. The raw text is pre-processed to remove the noise, and then it is represented as Bag-of-Words (BoW) for vectorization before training SOM. The small image clusters can now be merged based on the corresponding textual features associated with these images. The new accuracy obtained with this technique is 90.12% which is better than the previously achieved accuracy. From the results, it can be inferred that the system performance has improved with the inclusion of the second modality. The performance of the approach

Table 5. Difference in the accuracy values of after and before WT while extracting ZM. Positive values in the last row signifies the benefit of WT application prior to ZM.

	Accuracy of WTZM - ZM			
Trials	**TR**	**VAL**	**TS**	**All**
1	1.4	8.4	3.9	4.3
2	1.9	2.3	0.6	1.7
3	4.8	-1.2	-0.9	2.7
4	7.2	-0.2	2.2	5
5	2.8	6.7	-0.6	2
AVG	3.8	3	2.9	3.4

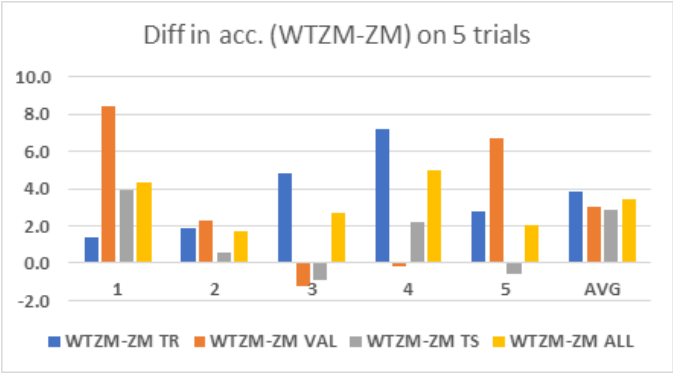


Fig. 19. Difference in the accuracy of results by using WT+ZM versus only ZM as in Table 5.

Table 6. Average test results for 5-trials with the proposed method on 300 image vectors. P_Healthy, P_Sick and A_Healthy, A_Sick are predicted and actual values of healthy and sick classes respectively.

	A_Healthy	**A_Sick**	**Total**
P_Healthy	161	16	177
P_Sick	19	104	123
Total	180	120	300
Accuracy	89.4%	86.7%	88.3%

can be further boosted if the quality of image captions is good and they are noise-free, having no stop words and more meaningful information about the corresponding image.

Table 7. Comparative analysis of techniques on the underlying dataset.

Sr.	Reference	Technique	Accuracy
1	[48]	Principal component analysis	74%
2	[49]	Agglomerative clustering and CNN	80.5%
3	[55]	Nonlinear subspace clustering	83.3%
4	[33]	Linear discriminant analysis	84%
5	[38]	Robust learning for unsupervised clustering	85.7%
6	[41]	Deep density-based image clustering	86.8%
7	[50]	Hierarchical clustering using 1D random projections	87.1%
8	[8]	Deep adaptive image clustering	87.6%
9	[19]	Discriminative pseudo supervision clustering	87.9%
10	**Proposed**	**WT+ZM on LUV**	**88.3%**

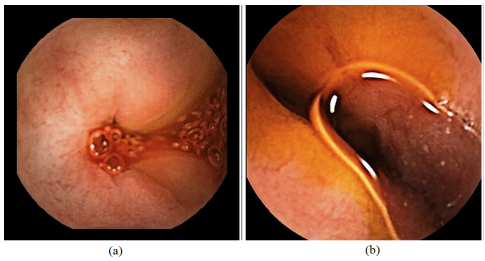


Fig. 20. (a) Active bleeding in small bowel (b) False positive (air bubble), images from [6]

6.5. Discussion

The primary goal of the proposed research is to present the importance of understanding and analyzing the data to find the appropriate methods for its processing as per the essence of the data and the underlying problem. The final values chosen for all the tuning parameters for experimentation have been decided based on multiple trials of experiments. We have also tested the proposed technique by increasing the size of the endoscopy data to 1200 and observed that the performance and accuracy are almost similar to the smaller data. The proposed approach outperforms the traditional as well as contemporary image clustering approaches due to following reasons:

1. An appropriate medical image representation is an important task for which the combination of Wavelet transform and Zernike moments have been utilized to retrieve noise-free, least redundant, and rotation, scaling, and translational invariant features. ZM captures the global features of an image and also effectively describe the shape characteristics [2].

2. Self-organizing map provides the robust medical image clustering as it works similar to the brain neurons [27]. In addition, SOM has been quite effective in diverse recent applications such as mental stress detection [46], coronary heart disease diagnosis [35], speech recognition [31], and in feature extraction as an add-on for better network intrusion detection [9].
3. SOM provides easy data interpretation, and understanding [27]. It provides potent data visualization and has the capability of clustering even complex datasets [28].
4. Deep learning requires a colossal dataset to perform well [5], which is unavailable in the proposed research; this may cause the over-fitting issue [43].
5. Generally, a deep learning approach (such as CNN) cannot directly outperform a machine learning approach as its performance mostly depends upon the design comprising training strategies, layer depth, and size [17].
6. To use transfer learning and retraining the deep learning model on a new dataset requires understanding various model parameters and the layer modifications, which is computationally expensive [36].

7. Conclusion

This paper introduced new ways of intelligently segmenting and analyzing image collections by training neural computing systems with images having obscure color and texture contrasts. The characteristic visual features of the image collection are derived from Wavelet Transforms, Zernike Moments, and Linear Discriminant Analysis. The images are categorized using an unsupervised clustering algorithm – Kohonen self-organizing feature maps. The proposed system can classify sick and healthy in-vivo images effectively without the labeled data, which is hard to get in reality, specifically medical data. It is often expensive to manually label the data by an expert in the related field. The system is beneficial in clustering vague color distribution, asymmetrical region boundary, and noisy image data. It is rotation, scaling, and translation invariant due to the use of ZM for image representation. The system efficacy improved by incorporating the second modality, i.e., free text with the gastral images in the experiments.

7.1. Limitations

There are three types of errors involved in the calculation of ZM: (a) Geometric error due to mapping of a digital image into a unit circle with pixels, (b) Discretization error due to the computer's digital representation of continuous variables, and (c) Numerical integration error due to the calculation of double integration through double summations while the center of a grid is used to calculate the basis function. The size of the data considered for experimentation in this study is small, which is a drawback.

7.2. Future scope

Grid size for SOM is a parameter for subjective tuning. The overall accuracy of the proposed system is encouraging, although image semantics need to be considered more carefully to improve the automatic learning system. Future experimentation can be performed

on ample data from diverse fields, and miscellaneous noise removal and appropriate feature extractors can be considered (as per the underlying data), which can further improve the accuracy. Various ways of integrating multi-modal data can be explored to extend this work further and improve the clustering process.

Acknowledgments. The authors are thankful to (a) Gastroenterologists Dr. Sunil Arya at Leela Bhawan Patiala & Dr. G.S. Sidhu at Max Hospital Mohali, India, for the dataset and technical feedback, and (b) Professor Khurshid Ahmad, Trinity College Dublin, Ireland for his valuable advice.

References

1. Aggarwal, A., Singh, C.: Zernike moments-based gurumukhi character recognition. Applied Artificial Intelligence 30(5), 429–444 (2016)
2. Aggarwal, H., Vishwakarma, D.K.: Covariate conscious approach for gait recognition based upon zernike moment invariants. IEEE Transactions on Cognitive and Developmental Systems 10(2), 397–407 (2017)
3. Aghajari, E., Chandrashekhar, G.D.: Self-organizing map based extended fuzzy c-means (seefc) algorithm for image segmentation. Applied Soft Computing 54, 347–363 (2017)
4. Alegre, E., González-Castro, V., Alaiz-Rodríguez, R., García-Ordás, M.T.: Texture and moments-based classification of the acrosome integrity of boar spermatozoa images. Computer Methods and Programs in Biomedicine 108(2), 873–881 (2012)
5. Bekhouche, S., Dornaika, F., Benlamoudi, A., Ouafi, A., Taleb-Ahmed, A.: A comparative study of human facial age estimation: handcrafted features vs. deep features. Multimedia Tools and Applications 79(35), 26605–26622 (2020)
6. Boal Carvalho, P., Magalhães, J., Dias de Castro, F., Monteiro, S., Rosa, B., Moreira, M.J., Cotter, J.: Suspected blood indicator in capsule endoscopy: a valuable tool for gastrointestinal bleeding diagnosis. Arquivos de gastroenterologia 54, 16–20 (2017)
7. Borges, V.R., de Oliveira, M.C.F., Silva, T.G., Vieira, A.A.H., Hamann, B.: Region growing for segmenting green microalgae images. IEEE/ACM transactions on computational biology and bioinformatics 15(1), 257–270 (2016)
8. Chang, J., Wang, L., Meng, G., Xiang, S., Pan, C.: Deep adaptive image clustering. In: Proceedings of the IEEE international conference on computer vision. pp. 5879–5887 (2017)
9. Chen, Y., Ashizawa, N., Yeo, C.K., Yanai, N., Yean, S.: Multi-scale self-organizing map assisted deep autoencoding gaussian mixture model for unsupervised intrusion detection. Knowledge-Based Systems 224, 107086 (2021)
10. Cong, Y., Wang, S., Liu, J., Cao, J., Yang, Y., Luo, J.: Deep sparse feature selection for computer aided endoscopy diagnosis. Pattern Recognition 48(3), 907–917 (2015)
11. Daubechies, I.: Different perspectives on wavelets, vol. 47. American Mathematical Soc. (2016)
12. Deng, A.W., Wei, C.H., Gwo, C.Y.: Stable, fast computation of high-order zernike moments using a recursive method. Pattern Recognition 56, 16–25 (2016)
13. Du, W., Rao, N., Yong, J., Wang, Y., Hu, D., Gan, T., Zhu, L., Zeng, B.: Improving the classification performance of esophageal disease on small dataset by semi-supervised efficient contrastive learning. Journal of Medical Systems 46, 1–13 (2022)
14. Fan, X., Tjahjadi, T.: A dynamic framework based on local zernike moment and motion history image for facial expression recognition. Pattern recognition 64, 399–406 (2017)
15. Faust, O., Acharya, U.R., Adeli, H., Adeli, A.: Wavelet-based eeg processing for computer-aided seizure detection and epilepsy diagnosis. Seizure 26, 56–64 (2015)

16. Fredo, A.J., Abilash, R., Femi, R., Mythili, A., Kumar, C.S.: Classification of damages in composite images using zernike moments and support vector machines. Composites Part B: Engineering 168, 77–86 (2019)
17. Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J.: Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing 11(2), 196 (2019)
18. Gupta, V., Priya, T., Yadav, A.K., Pachori, R.B., Acharya, U.R.: Automated detection of focal eeg signals using features extracted from flexible analytic wavelet transform. Pattern Recognition Letters 94, 180–188 (2017)
19. Hu, W., Chen, C., Ye, F., Zheng, Z., Du, Y.: Learning deep discriminative representations with pseudo supervision for image clustering. Information Sciences 568, 199–215 (2021)
20. Indolia, S., Nigam, S., Singh, R.: A self-attention-based fusion framework for facial expression recognition in wavelet domain. The Visual Computer pp. 1–17 (2023)
21. Kar, A., Pramanik, S., Chakraborty, A., Bhattacharjee, D., Ho, E.S., Shum, H.P.: Lmzmpm: local modified zernike moment per-unit mass for robust human face recognition. IEEE Transactions on Information Forensics and Security 16, 495–509 (2020)
22. Karim, S.A.A., Kamarudin, M.H., Karim, B.A., Hasan, M.K., Sulaiman, J.: Wavelet transform and fast fourier transform for signal compression: A comparative study. In: 2011 International Conference on Electronic Devices, Systems and Applications (ICEDSA). pp. 280–285. IEEE (2011)
23. Kaur, P., Pannu, H.S., Malhi, A.K.: Plant disease recognition using fractional-order zernike moments and svm classifier. Neural Computing and Applications 31(12), 8749–8768 (2019)
24. Kaur, P., Pannu, H.S., Malhi, A.K.: Comparative analysis on cross-modal information retrieval: a review. Computer Science Review 39, 100336 (2021)
25. Khare, M., Khare, A.: Integration of complex wavelet transform and zernike moment for multiclass classification. Evolutionary Intelligence 14(2), 1151–1162 (2021)
26. Khotanzad, A., Hong, Y.H.: Invariant image recognition by zernike moments. IEEE Transactions on pattern analysis and machine intelligence 12(5), 489–497 (1990)
27. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biological cybernetics 43(1), 59–69 (1982)
28. Kohonen, T.: Essentials of the self-organizing map. Neural networks 37, 52–65 (2013)
29. Kundu, A.K., Fattah, S.A.: Probability density function based modeling of spatial feature variation in capsule endoscopy data for automatic bleeding detection. Computers in Biology and Medicine 115, 103478 (2019)
30. Kurtz, C., Depeursinge, A., Napel, S., Beaulieu, C.F., Rubin, D.L.: On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. Medical image analysis 18(7), 1082–1100 (2014)
31. Lokesh, S., Kumar, P.M., Devi, M.R., Parthasarathy, P., Gokulnath, C.: An automatic tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map. Neural Computing and Applications 31(5), 1521–1531 (2019)
32. Maida, M., Celsa, C., Lau, L.H., Ligresti, D., Baraldo, S., Ramai, D., Di Maria, G., Cannemi, M., Facciorusso, A., Cammà, C.: The application of large language models in gastroenterology: A review of the literature. Cancers 16(19), 3328 (2024)
33. Martinez, A.M., Kak, A.C.: Pca versus lda. IEEE transactions on pattern analysis and machine intelligence 23(2), 228–233 (2001)
34. Muhammad, K., Khan, S., Kumar, N., Del Ser, J., Mirjalili, S.: Vision-based personalized wireless capsule endoscopy for smart healthcare: Taxonomy, literature review, opportunities and challenges. Future Generation Computer Systems 113, 266–280 (2020)
35. Nilashi, M., Ahmadi, H., Manaf, A.A., Rashid, T.A., Samad, S., Shahmoradi, L., Aljojo, N., Akbari, E.: Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. International Journal of Fuzzy Systems 22(4), 1376–1388 (2020)

36. Pannu, H.S., Ahuja, S., Dang, N., Soni, S., Malhi, A.K.: Deep learning based image classification for intestinal hemorrhage. Multimedia Tools and Applications 79, 21941–21966 (2020)
37. Papakostas, G., Boutalis, Y., Karras, D., Mertzios, B.: Efficient computation of zernike and pseudo-zernike moments for pattern classification applications. Pattern Recognition and Image Analysis 20(1), 56–64 (2010)
38. Park, S., Han, S., Kim, S., Kim, D., Park, S., Hong, S., Cha, M.: Improving unsupervised image clustering with robust learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12278–12287 (2021)
39. Patel, A., Rani, K., Kumar, S., Figueiredo, I.N., Figueiredo, P.N.: Automated bleeding detection in wireless capsule endoscopy images based on sparse coding. Multimedia Tools and Applications 80(20), 30353–30366 (2021)
40. Radhika, K., Venkatesha, M., Sekhar, G.: An approach for on-line signature authentication using zernike moments. Pattern Recognition Letters 32(5), 749–760 (2011)
41. Ren, Y., Wang, N., Li, M., Xu, Z.: Deep density-based image clustering. Knowledge-Based Systems 197, 105841 (2020)
42. Roussinov, D., Chen, H.: A scalable self-organizing map algorithm for textual classification: A neural network approach to thesaurus generation. Communication Cognition and Artificial Intelligence 15(1-2), 81–111 (1998)
43. Siddiqui, S.A., Salman, A., Malik, M.I., Shafait, F., Mian, A., Shortis, M.R., Harvey, E.S.: Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES Journal of Marine Science 75(1), 374–389 (2018)
44. Singh, C., Aggarwal, A.: An efficient approach for image sequence denoising using zernike moments-based nonlocal means approach. Computers & Electrical Engineering 62, 330–344 (2017)
45. Soffer, S., Klang, E., Shimon, O., Nachmias, N., Eliakim, R., Ben-Horin, S., Kopylov, U., Barash, Y.: Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. Gastrointestinal endoscopy 92(4), 831–839 (2020)
46. Tervonen, J., Puttonen, S., Sillanpää, M.J., Hopsu, L., Homorodi, Z., Keränen, J., Pajukanta, J., Tolonen, A., Lämsä, A., Mäntyjärvi, J.: Personalized mental stress detection with self-organizing map: From laboratory to the field. Computers in Biology and Medicine 124, 103935 (2020)
47. Xue-guang, C., et al.: An improved image segmentation algorithm based on two-dimensional otsu method. Information Sciences Letters 1(3), 2 (2012)
48. Yang, J., Yang, J.y.: From image vector to matrix: A straightforward image projection technique—impca vs. pca. Pattern Recognition 35(9), 1997–1999 (2002)
49. Yang, J., Parikh, D., Batra, D.: Joint unsupervised learning of deep representations and image clusters. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5147–5156 (2016)
50. Yellamraju, T., Boutin, M.: Clusterability and clustering of images and other "real" high-dimensional data. IEEE Transactions on Image Processing 27(4), 1927–1938 (2018)
51. Yuan, Y., Meng, M.Q.H.: Deep learning for polyp recognition in wireless capsule endoscopy images. Medical physics 44(4), 1379–1389 (2017)
52. Zhang, D.: Wavelet transform. In: Fundamentals of Image Data Mining, pp. 35–44. Springer (2019)
53. Zhang, J., Wu, Q., Zhang, J., Shen, C., Lu, J., Wu, Q.: Heritage image annotation via collective knowledge. Pattern Recognition 93, 204–214 (2019)
54. Zhou, J., Fu, X., Zhou, S., Zhou, J., Ye, H., Nguyen, H.T.: Automated segmentation of soybean plants from 3d point cloud using machine learning. Computers and Electronics in Agriculture 162, 143–153 (2019)
55. Zhu, W., Lu, J., Zhou, J.: Nonlinear subspace clustering for image clustering. Pattern Recognition Letters 107, 131–136 (2018)

Parminder Kaur is a postdoc fellow at Durham University UK. She did her PhD from Thapar Institute Patiala India. Her research interests are machine learning, cross-modal learning using image and text, and image segmentation.
https://www.durham.ac.uk/staff/parminder-kaur/
https://scholar.google.co.in/citations?user=qvk7yvAAAAAJ&hl=en

Avleen Malhi is faculty at WMG, University of Warwick UK. She was a postdoc fellow at Aalto University Finland and did her PhD from Thapar Institute Patiala India. Her research interests are machine learning, explainable AI, wireless ad hoc networks.
https://ieeexplore.ieee.org/author/37085517161
https://scholar.google.co.in/citations?user=bMA1WcMAAAAJ&hl=en

Husanbir Singh Pannu is an assistant professor at Thapar Institute Patiala India. He was a postdoc in University of Liverpool UK and Trinity College Dublin Ireland. He did his PhD from University of North Texas USA. His research interests are machine learning and data analysis.
https://scholar.google.co.in/citations?user=DNPuK98AAAAJ&hl=enoi=ao

Received: June 28, 2024; Accepted: December 17, 2024.