

Quantum Statistical Information Grid Clustering for Early Esophageal Adenocarcinoma Detection

G. Vani¹; Dr.A. Hema²

¹Computer Science, Kongunadu Arts and Science College, Coimbatore, India.

¹gvanimca82@gmail.com

²Computer Science, Kongunadu Arts and Science College, Coimbatore, India.

²hemasrgm@yahoo.com

Abstract

Recent advances in the field of digital endoscopy have recognized as a primary key knowledge for medical screening of presence of diseases at its earlier stages and minimal invasive surgery. Many research works are done on automatic analyzing and detection of early esophageal adenocarcinoma which is not easy to detect at its primary stage. Because the esophageal images are naturally indeterministic and it is uncertain to detect its earliest appearance more precisely at its right stage of diagnosis. Thus, in this present research work the statistical information-based grid clustering is developed by empowering its clustering capability using Quantum mechanism. The conventional Statistical Information Grid Clustering (STING) reduces the computation complexity of the clustering process, but outlier detection and uncertainty handling are very challenging, because it partitions the Esophageal image in the layer by layer levels. Each layer is divided into cells and they are known as child cells. The relevant child cells alone are used to determine the statistical information about the entire image and based on the spatial information the similarity measure among two images are determined. The distance measure and the neighborhood pixel information are computing using quantum theory to overcome the uncertainty and parallel processing is done for concurrency based early esophageal adenocarcinoma prediction. The Barret's Esophageal images are used for clustering the cancerous and noncancerous images. The simulation results proved that the proposed Quantum Statistical Information Grid Clustering produce better result in detection of Early Esophageal Adenocarcinoma detection compared to other standard clustering models.

Key-words: Early Esophageal Adenocarcinoma Detection, Barret's Esophageal Images, Uncertainty, Statistical Information Grid Clustering, Quantum Theory.

1. Introduction

The prevalence of Esophageal Adenocarcinoma (EAC) is rapidly increasing and it is one of the common gastrointestinal tracts of malignancies. EAC is the seventh common cancer in the world with 3.2% and ranks 6th cancer mortality which is 5.3% [1]. In India according to the World Health Organization, Globocon report 2018, with the ratio of 2:4:1 as the male to female ratio [2].

When cancer cells developed in the tube-like structure known as esophagus that runs from throat to stomach is known as Esophageal cancer. The cancer begins its growth from the inner layer of the esophagus and spreads all over other layers of esophagus and results in metastasis [3]. There are two major categories of esophageal cancer they are squamous cell carcinoma and adenocarcinoma. In the former type, cancer developed in the squamous cell and spreads along the whole esophagus, where the latter one develops from the gland cells. The squamous cells which normally line will be replaced by gland cells in esophagus when adenocarcinoma is developing.

The dataset used in this work is collected from the Early Barret's Cancer detection, which comprised of 50 examples of early cancerous images and 50 examples of non-cancerous images. It is very complex to identify the cancerous tissue during endoscopic investigation of the esophagus. But detecting its presence at the earlier stage can help to surgically remove it using endoscopic mucosal resection. Hence with the advancement of image processing in the medical images are greatly desirable to diagnose the presence of early esophageal adenocarcinoma, which may improve the survival rate of the victims.

This present research work concentrates on developing an uncertain scheme-based principle for effective clustering of early esophageal adenocarcinoma presence more conspicuously.

2. Related Work

Fang Yang et al [4] developed computer aided esophageal adenocarcinoma diagnostic model by performing classification using support vector machine and K nearest neighbor. It performs preprocessing, extraction of features, classification of images to improve the quality of cancer diagnosis.

Yu et al [5] developed a multi-image features which improves classification accuracy but it results in overfitting risk. The robustness of the system is decreased if image features are not well balanced, thus, it essential to use limited features of images during classification.

Gladis et al [6] in their work the type of brain MR image is classified by developing integrated model of support vector machine with principle component analysis for performing recognition of brain type.

Li et al [7] designed a novel feature selection using sequential forward selection method. This work focusses on nonunique probe selection issue, the model is compared with greedy algorithm.

Fu et al [8] stated that datamining techniques and artificial intelligence are used for predicting abnormalities in medical images to predict specific diseases.

Papadopoulos et al [9] devised a statistical theory based on artificial neural network and support vector machine which characterizes digitized mammograms using micro classifications clusters.

Katsuyoshi et al [10] in their work developed a K-Nearest Neighbor based breast cancer detection. The number of nearest neighbor selection greatly affects the process of classification accuracy. The sequential backward selection and forward selection models are used to remove the redundant information.

Sharma and Khanna [11] designed a Support vector machine-based breast cancer detection, they used gray level co-occurrence, Zernike moment and Discrete Cosine Transformation (DCT) extracts the features and that information are used for diagnosis process.

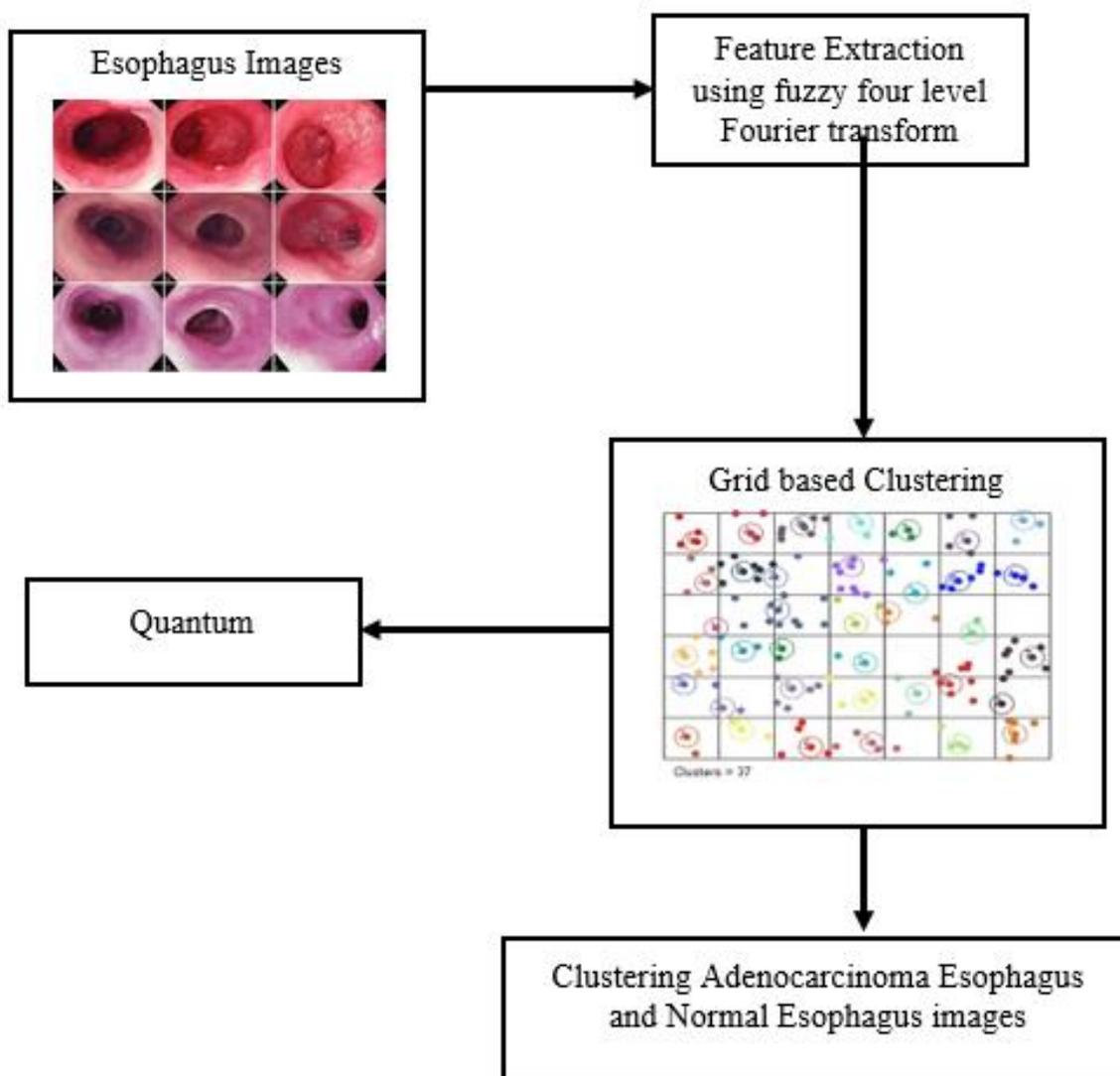
Fons van et al [12] designed a promising tool to detect cancer in Barrett's Esophagus using Volumetric Laser Endomicroscopy. The clinical knowledge-based features are used in this work to improve the accuracy of diagnosis.

Yoshimasa et al [13] developed a convolutional neural network for predicting the esophageal cancer by collecting 8428 images from the cancer institute hospital, Japan. The testing image comprised of 47 subjects with esophageal cancers and 50 patients with normal health to evaluate the accuracy of the prediction model.

Ilango and Mohan [14] performed a detailed survey on various types of clustering, specifically about the grid-based clustering, which quantize the data space into finite cells and performs needed operation on the quantized space. They stated that grid clustering fastens the computation of clustering depending on the cell dimension.

Romy et al [15] presented a novel diagnoses model using Dutch pathology registry. They diagnose the patients as BE or EAC or esophageal cancer without histological validation. The information of familial status is also considered as the additional information for esophageal cancer detection.

3. Methodology Detection of Adenocarcinoma Esophagus Using Enhanced Chaotic Grid based Clustering



This phase performs clustering of images as normal esophagus or abnormal esophagus. This work uses Barrett's esophagus image collected from Endoscopic Vision Challenge MICCAI 2015, consisting of 100 subjects [18]. Initially, the raw image is preprocessed by noise removal and contrast enhancement. The significant features of esophagus are extracted using fuzzy four-level Fourier transform. This work adapts Statistical Information Grid based Clustering (STING) to determine the similar pattern of images and quantum mechanism is used to overcome indeterminacy and uncertainty in centroid selection and presence of outliers during the process of clustering of esophagus images. The detailed description of the proposed model is explained in the following subsections.

4. Esophagus Image Preprocessing

The raw dataset collected from Endoscopic Vision Challenge MICCAI 2015 of both male and female patients with the age group between 45-80 [18]. The input images consist of noise information, so noise removal is done using Adaptive Wiener Filter. It splits the esophagus image into equal size of 8 x 8 blocks. Intensity histogram is calculated on each contextual region of image, clip limits are used for clipping the histogram with the predefined threshold value, so that the local contrast of the Esophagus image is improved without losing its information. The histogram is equalized using Joint Histogram Equalization (JHE) method. While using JHE it calculates both pixel intensity correlation and its mean of its neighborhood pixel to improve and equalize the contrast enhancement

The feature extraction is done using fuzzy four level Fourier transform, which investigates the features of the Esophageal image by deep investigation of lower level to higher level segmentation [19].

5. Grid based Clustering

The clustering algorithm's computation complexity is directly proportional to the dataset size. To overcome this issue in adenocarcinoma disease detection, grid-based clustering [16] is used to potentially decrease the computation complexity even when the dataset size is larger. This is because it considers the value space which surrounds the data points, not with the data. Typically, the grid-based clustering model comprised of five main steps they are

- Partition the space of data into predetermined number of cells, it is in the grid structure.
- For each cell compute the cells densities.
- According to their densities, sort the cells.
- Discover the cluster centers.
- Navigation to neighboring cells.

There are two types of Grid based Clustering they are:

- Statistical Information Grid based Clustering (STING): This algorithm divides the spatial region is split into rectangular cells. Depending on the resolution of an image, there are several levels of cells. The statistical information of each cell is computed and used for further processing.

- Wave Cluster: This algorithm uses wavelet transform over the feature space and it belong to the multi-resolution clustering model. It converts the image into signals and divides them as sub band for further processing
- Clustering In QUEst (CLIQUE): This algorithm integrates both the density and grid-based clustering concepts. It discovers the high dimensional data space and their spaces automatically and better clusters than the original space.

In this research work Statistical Information Grid based clustering is adapted for clustering the esophagus images into normal and adenocarcinoma disease. The STING clustering it is easy to parallelize, but its boundaries are diagonal, so to overcome this disadvantage, this present research work induced the concept of quantum mechanism, which overcomes the uncertainty and outliers by focusing on the mean fitness values of the centroids and the images are clustered in an optimized manner.

6. Statistical Information Grid based Clustering

The spatial information plays a vital role in image process especially when handing the medical images, it extracts the implied information, spatial relation about the neighboring pixels and interesting features about the image are discovered in addition with patterns which are not mentioned in the database. STING is a kind of grid model which performs multi-resolution clustering where the spatial zone of image is divided into rectangular cells and it employs a hierarchical structure. Depending on the levels of resolution of esophagus image, the levels of rectangular cells will also vary.

It divides the spatial space by starting at the higher level and it partitioned to generate child cells at lower level. For instance, i^{th} cell related to the union of its children at the $i+1^{\text{th}}$ level. It is formed in such a way that each cell comprised of four children and each child related to one quadrant of its parent cell.

The statistical information of each grid cell, that is their attributes like mean, minimum and maximum values, standard deviation are precomputed and stored. Using the parameters of the lower cell units the higher-level cells can be computed easily. There are both independent and dependent parameters attributes are gathered from each cell.

- Independent parameter Attribute - count
- Dependent parameter Attribute – MV mean value, MN- minimum value of the cell attribute, MX – maximum value of the cell attribute, SD- Standard deviation of all the cell values, DT- Type of distribution either normal, exponentially uniform and node.

- The MV, MX, MN, SD, count is directly computed for the bottom cells of an image which is loaded
- The relevancy of each cell is calculated
- The distribution is obtained by using chi square test.

The layer is determined from the beginning of query processing, the relevancy alone is considered and irrelevant cells are removed from further processing.

7. Quantum Statistical Information Grid based Clustering

The statistical information Grid based clustering is modified in this research work by introducing the quantum mechanism. The conventional STING model performs clustering by dividing the image into layers of cells and expand it with the cells when it goes in deeper investigation, it frames the hierarchical structure based on the levels and the information of each cell is computed to determine the interesting patterns.

In this research work, the Esophageal images region of interest is extracted using the fuzzy four level fourier transformation, ROI region is used for performing further processing. Using sting index begin at root of the ROI and proceed to the next lower levels. Using the statistical information compute the likelihood of a cell to determine the relevancy with the query image similarity confidence level. The child cells which are relevant to the querying image are further explored in deep. Repeat the process until it reaches the bottom layer. If the matching compact ability of the querying image and the image under investigation are high then they are clustered together and considered to be more similar than the other clusters.

But the conventional STING during clustering it can examine the pixel information in the rectangular format only, it cannot able to examine the diagonal parts thus handling outliers and uncertainty are very challenging. Thus, to overcome this issue quantum mechanism which performs the information gathering of relevant cells in a concurrent manner by determining the m dimensional space for each layer. Quantum computing has its unique method for computation which is entirely different from the standard computation [17]. When there is a continuous increase of esophagus image dataset set, then using standard STING clustering will considerably increase the time and space complexity. To reduce the space and time complexity, the quantum computing will greatly reduce them by representing the pixels in the space and distance in the quantum representation. The main different with the conventional clustering and the proposed QSTING clustering is representation of pixels and

the distance calculation in the space. They data points in space are signified as vectors. The pixel of an image \hat{Y}_i is represented in the quantum state as

$$\hat{Y}_i = |Y_i\rangle|Y_i\rangle$$

Where $|Y_i\rangle$ denotes qbits. If the vector is a four dimension then $|Y_i\rangle$ is denoted as

$$|Y_i\rangle = \beta_0|00\rangle + \beta_1|01\rangle + \beta_2|10\rangle + \beta_3|11\rangle$$

Where calculation basis is signified as $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, $\beta_{i=0,1,2,3}$ refers to the probability of corresponding computed basis and they are satisfied with the condition of summing the values of the probability equal to 1.

Likewise, the distance between the two cells are determined using the quantum expression as follows

$$DT = |\hat{Y}_i - \hat{Y}_j| = \sqrt{|\hat{Y}_i - \hat{Y}_j|} \\ = \sqrt{(|Y_i\rangle\langle Y_i| - |Y_j\rangle\langle Y_j|)(|Y_i\rangle|Y_i\rangle - |Y_j\rangle|Y_j\rangle)}$$

The major idea of this proposed QSTING is to adapt the coherence state of quantum, so that it influences the result of clustering with the satisfied accuracy. The quantum states spatial non-localization, probability and its coherence greatly improve the probability of accurate clustering. The quantum mechanism entangle state gets the relation among the Euclidean distance, probability and its inner product.

The major aspect of QSTING which makes it superior than the standard STING is listed as follows:

- The diverse nature of searching the relevant cells are handled well in the large spatial area as it's an uncertain scheme, so that it can discover its better child cells for gathering the spatial information of an image for clustering to the concern group.
- The standard STING clustering faces the premature convergence due to local optima and it produce cluster result with less accuracy. While using QSTING it finds the mean best of the child cells which are relevant to frame the hierarchical structure to avoid the earlier convergence.
- The parameters used in QSTING very less and the computation complexity is highly reduced because of its parallel computation to gather the statistical information.

8. Experimental Results and Discussions

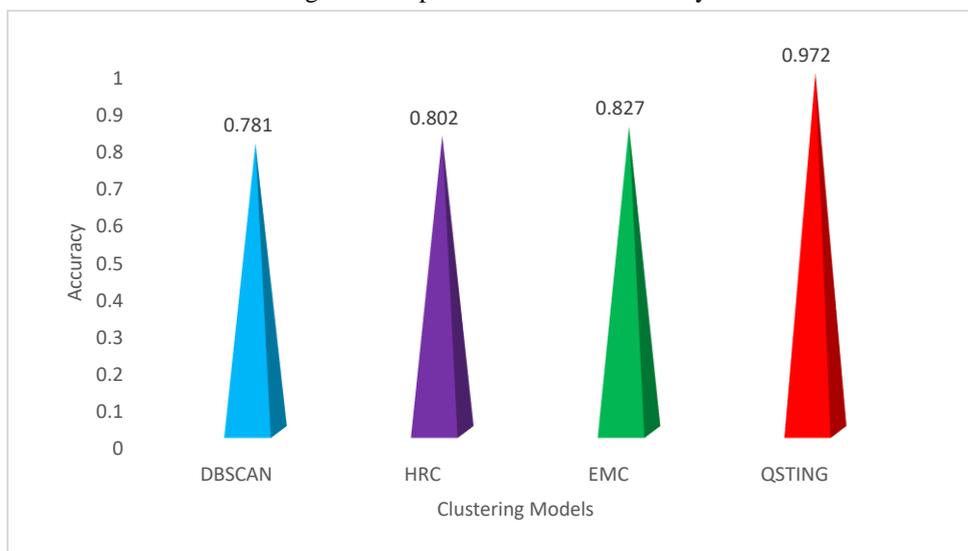
The proposed Quantum Statistical Information Grid based clustering (QSTING) is designed and deployed using python code. The dataset is collected from Endoscopic Vision Challenge MICCAI 2015 [] with 1689 Esophagus images. The proposed work QSTING is compared with three conventional clustering models DBSCAN, Hierarchical Clustering (HC), Expectation Maximization (EM). The detailed evaluation analysis is explained as follows.

Table: Performance Evaluation of Four Clustering Models

	Accuracy	Precision	Recall	Error Rate
DBSCAN	0.781	0.775	0.771	0.0352
HRC	0.802	0.786	0.783	0.0218
EMC	0.827	0.815	0.806	0.0139
QSTING	0.972	0.968	0.962	0.0052

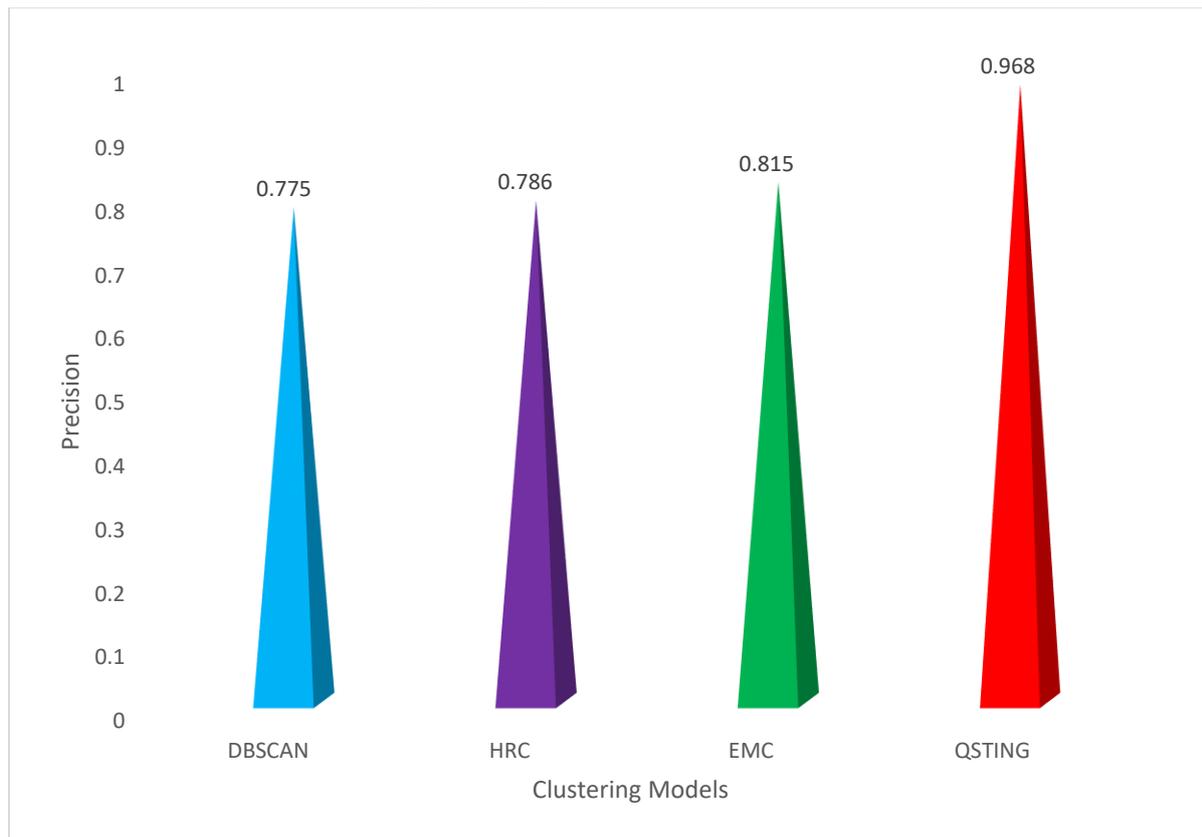
The table shows the results produced by four different clustering models for prediction of Adenocarcinoma in Esophagus. It is observed from the results that Quantum Statistical Information Grid based clustering (QSTING) produced highest accuracy, precision and recall because DBSCAN, Hierarchical and Expectation Maximization clustering are directly proportional to the computation complexity and clustering is done based on the value of data. They work feasibility with small size of dataset. In QSTING it depends on value space which surrounds the data instead of data value and it has the ability to potentially handle computation complexity without considering the size of dataset. Hence, QSTING produce less error rate when comparing with other three models.

Figure: Comparison based on Accuracy



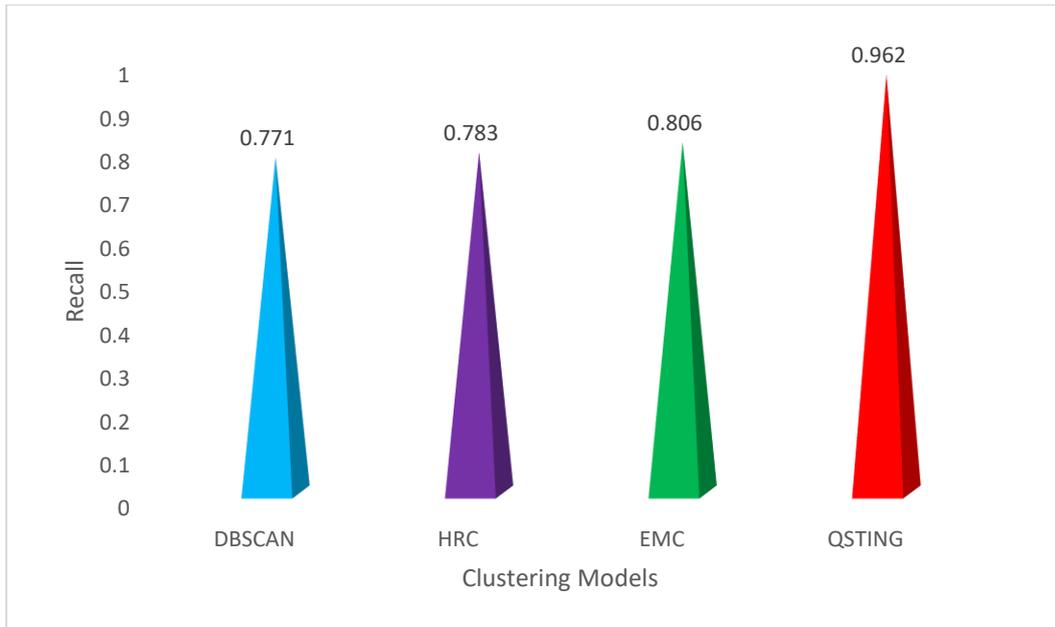
The figure shows the performance comparison of four different clustering models involved in prediction of Adenocarcinoma Esophagus based on accuracy. The result explores that proposed QSTING produced highest accuracy of clustering normal and Adenocarcinoma in Esophagus images. The reason is, it quantizes the space of Esophagus image and captures statistical information related to the spatial information to determine the similarity among other images in the dataset. While DBSCAN, Hierarchical and Expectation Maximization clustering produced less accuracy because they concentrate on distance-based similarity alone.

Figure: Comparison based on Precision



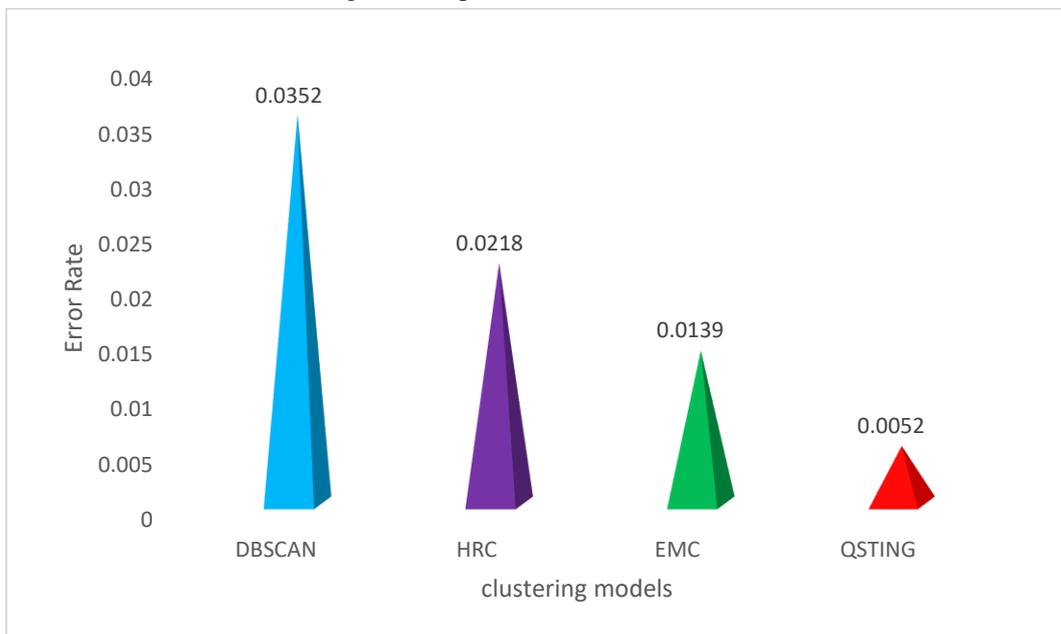
The precision value produced by four different clustering models QSTING, DBSCAN, Hierarchical and Expectation Maximization clustering is displayed in the figure. The QSTING produced best precision rate compared to other three clustering models because it uses Quantum mechanism which handles uncertainty principles and works in concurrent manner to reduce computation complexity.

Figure Comparison based on Recall



The figure illustrates the performance of four diverse clustering algorithms using recall metric for detection of Adenocarcinoma Esophagus. Proposed QSTING produced highest fraction of relevant clusters that were retrieved because the statistical information of Esophagus images partitioned in a grid manner and by applying quantum mechanism it overcomes the problem of premature convergence and it avoids local optima. Thus, the proposed QSTING competes the other three clustering models because their centroids are initially selected in a random manner.

Figure: Comparison based on Error Rate



The figure shows the error rate obtained by DBSCAN, HRC, EMC and proposed QSTING while predicting the adenocarcinoma in Esophagus image dataset. The conventional clustering models initially selects the centroids in a arbitrary fashion and clustering is done based on the distance metric for determining the similarity among the images. But the proposed QSTING uses the statistical information extracted from the spatial information of each cells of the concern Esophagus images and the quantum mechanism is used to determine the mean optimum clustering spaces for achieving optimized clustering by avoiding local maxima in their premature stage. Hence, the error rate of the QSTING is least compared to the other conventional clustering approaches.

9. Conclusion

This present research work aims to develop a spatial grid-based clustering which has the ability to handle the indeterminacy and uncertainty in identifying the early esophageal adenocarcinoma detection in Barret's Esophageal images. This proposed work QSTING improves the clustering of cancerous and non-cancerous images by adapting the quantum states spatial non-localization, probability and its coherence greatly improve the probability of accurate clustering. The quantum mechanism entangle state gets the relation among the Euclidean distance, probability and its inner product. The simulation results also proved that statistical information of Esophagus images partitioned in a grid manner and applying quantum mechanism it overcomes the problem of premature convergence and it avoids local optima. Thus, the performance analysis proved that the proposed QSTING produced best accuracy of clustering with reduced error rate compared to DBSCAN, Hierarchical Clustering and Expectation Maximization.

References

- Choksi D, Kolhe KM, Ingle M, Rathi C, Khairnar H, Chauhan SG, Chaudhary V, Shukla A, Pandey V. Esophageal carcinoma: An epidemiological analysis and study of the time trends over the last 20 years from a single center in India. *J Family Med Prim Care* 2020; 9: 1695-9.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.*, 2018; 68 :394-424.
- Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; 127: 2893-917.
- Fang Yang, Murat Hamit, Chuan B. Yan, Juan Yao, Abdugheni Kutluk, Xi M. Kong, and Sui X. Zhang "Feature Extraction and Classification on Esophageal X-Ray Images of Xinjiang Kazak Nationality" *Hindawi Journal of Healthcare Engineering* Volume 2017, Article ID 4620732, pages 1 to 11.

L. Yu and H. Liu, "Efficient feature selection via analysis of relevant and redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205–1224, 2004.

V. P. Gladis and S. Palani, "A novel approach for feature extraction and selection on MRI images for brain tumor classification," *Computer Science & Information Technology*, vol. 10, no. 5, pp. 225–234, 2012.

L. W. Li, N. Alioune, and R. Luis, "Sequential forward selection approach to the non-unique oligonucleotide robe selection problem," *Lecture Notes in Computer Science*, vol. 5265, pp. 262–275, 2008.

J. C. Fu, S. K. Lee, S. T. Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics*, vol. 29, no. 6, pp. 419–429, 2005.

A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered micro-calcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 141–150, 2005.

O. Katsuyoshi and P. P. Alberto, "A detailed description of the use of the KNN method for breast cancer diagnosis," *In the 2014 7th international conference on biomedical engineering and informatics*, pp. 606–610, Dalian, China, 2014.

S. Sharma and P. Khanna, "Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM," *Journal of Digital Imaging*, vol. 28, no. 1, pp. 77–90, 2015.

Fons van der Sommen, Sander R. Klomp, Anne-Fré Swager, Svitlana Zinger, Wouter L. Curvers, Jacques J.G.H.M. Bergman, Erik J. Schoon, Peter H.N. De With, Predictive features for early cancer detection in Barrett's esophagus using Volumetric Laser Endomicroscopy, *Computerized Medical Imaging and Graphics*, Volume 67, 2018, Pages 9-20,

Yoshimasa Horie, Toshiyuki Yoshio, Kazuharu Aoyama, Shoichi Yoshimizu, Yusuke Horiuchi, Akiyoshi Ishiyama, Toshiaki Hirasawa, Tomohiro Tsuchida, Tsuyoshi Ozawa, Soichiro Ishihara, Youichi Kumagai, Mitsuhiro Fujishiro, Iruru Maetani, Junko Fujisaki, Tomohiro Tada, Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks, *Gastrointestinal Endoscopy*, Volume 89, Issue 1, 2019, Pages 25-32,

Ilango, V Mohan "A Survey of Grid Based Clustering Algorithms. *International Journal of Engineering Science and Technology*, Vol. 2(8), 2010, 3441-3446

Romy E. Verbeek, Lisanne F. Spittuler, Anique Peute, Martijn G. H. van Oijen, Fiebo J. ten Kate, Jacob R. Vermeijden, Ardi Oberndorff, Jantine W. P. M. van Baal, and Peter D. Siersema Familial Clustering of Barrett's Esophagus and Esophageal Adenocarcinoma in a European Cohort *Clinical Gastroenterology and Hepatology* 2014;12:1656–1663

Wang, W., J. Yang and R. Muntz. "STING: A Statistical Information Grid Approach to Spatial Data Mining." *VLDB* (1997).

N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum nearest-neighbor algorithms for machine learning," *Quantum Information and Computation*, vol. 15, 2015.

<https://endovissub-barrett.grand-challenge.org>

G. Vani, Dr.A. Hema," Energy Consumed Multi Level Feature Extraction on Esophageal X-Ray Images Based on Fuzzy Four Level Fast Fourier Transform", Volume-10, Issue-11, November 2020.