

Article

A Hybrid Linear Iterative Clustering and Bayes Classification-Based GrabCut Segmentation Scheme for Dynamic Detection of Cervical Cancer

Anousouya Devi Magaraja ¹, Ezhilarasie Rajapackiyam ², Vaitheki Kanagaraj ³, Suresh Joseph Kanagaraj ³, Ketan Kotecha ^{4,*} , Subramaniaswamy Vairavasundaram ², Mayuri Mehta ^{5,*}  and Vasile Palade ^{6,*} 

¹ Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry 605107, India

² School of Computing, SASTRA Deemed University, Thanjavur 613401, India

³ Department of Computer Science, Pondicherry University, Karaikal Campus, Puducherry 609605, India

⁴ Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International, Deemed University, Pune 412115, India

⁵ Department of Computer Engineering, Sarvajanic College of Engineering and Technology, Surat 39500, India

⁶ Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 2TU, UK

* Correspondence: head@scaai.siu.edu.in (K.K.); mayuri.mehta@scet.ac.in (M.M.); vasile.palade@coventry.ac.uk (V.P.)



Citation: Magaraja, A.D.; Rajapackiyam, E.; Kanagaraj, V.; Kanagaraj, S.J.; Kotecha, K.; Vairavasundaram, S.; Mehta, M.; Palade, V. A Hybrid Linear Iterative Clustering and Bayes Classification-Based GrabCut Segmentation Scheme for Dynamic Detection of Cervical Cancer. *Appl. Sci.* **2022**, *12*, 10522. <https://doi.org/10.3390/app122010522>

Academic Editor: Nikolaos Dikaos

Received: 13 September 2022

Accepted: 12 October 2022

Published: 18 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Cervical cancer earlier detection remains indispensable for enhancing the survival rate probability among women patients worldwide. The early detection of cervical cancer is done relatively by using the Pap Smear cell Test. This method of detection is challenged by the degradation phenomenon within the image segmentation task that arises when the superpixel count is minimized. This paper introduces a Hybrid Linear Iterative Clustering and Bayes classification-based GrabCut Segmentation Technique (HLC-BC-GCST) for the dynamic detection of Cervical cancer. In this proposed HLC-BC-GCST approach, the Linear Iterative Clustering process is employed to cluster the potential features of the preprocessed image, which is then combined with GrabCut to prevent the issues that arise when the number of superpixels is minimized. In addition, the proposed HLC-BC-GCST scheme benefits of the advantages of the Gaussian mixture model (GMM) on the extracted features from the iterative clustering method, based on which the mapping is performed to describe the energy function. Then, Bayes classification is used for reconstructing the graph cut model from the extracted energy function derived from the GMM model-based Linear Iterative Clustering features for better computation and implementation. Finally, the boundary optimization method is utilized to considerably minimize the roughness of cervical cells, which contains the cytoplasm and nuclei regions, using the GrabCut algorithm to facilitate improved segmentation accuracy. The results of the proposed HLC-BC-GCST scheme are 6% better than the results obtained by other standard detection approaches of cervical cancer using graph cuts.

Keywords: Bayes classification; cervical cancer; GrabCut; linear iterative clustering; Pap smear cell test

1. Introduction

CERVICAL cancer is an abnormal development in the cervix's tissues that connect the uterus and vaginal region of the women's reproductive system. This cervical cancer is generally caused by an initial infection caused by the intrusion of the high-risk enabled human papillomavirus or any other cause that could not be warranted [1]. This cervical cancer is a slow and steady developing cancer that does not exhibit any symptoms. Hence, this category of cancer needs to be detected through periodic Pap smear cell tests. In this Pap smear cell test, a specific collection of cervix cells is scraped and screened under a microscope for predominant cancer cell detection.

In contrast to the other cancer infectious categories, precursor lesions must be identified and treated by examining Pap smear cell tests [2]. Further, several automated cervical cancer detection schemes contributed to error minimization and productivity improvement during cervical cancer screening [3]. The considered automated cervical cancer detection schemes need to automatically differentiate between the normal and abnormal cells for a higher screening rate. The cytoplasm and nuclei morphological features are essential in distinguishing between normal and abnormal cells for a higher cervical cancer detection rate [4].

The most recent contributions in the relevant literature focus on the potential role of the nucleus in the process of cancer detection. Specifically, cervical cytoplasm investigation is vital in detecting all cervical cancer categories corresponding to squamous cell carcinoma, low-grade squamous cell intra-epithelial carcinoma, and accompanied nucleus abnormality categories [5]. Hence, reliable and automated detection of cervical cytoplasm and nuclei is essential for accurate screening of nucleus abnormalities categories of cervical cancer. This automated scheme also suffers from the limitations of the poor contrasting quality of cells and inappropriate cervical cell staining process that makes the extraction step of the cervical cells a complex issue in the detection phases. In addition, the Pap smear cells used for examination vary from thousands to ten thousand, leading to a complex investigation task.

Moreover, most of the contributed automatic cervical cancer detection schemes focused on enhancing the classification accuracy rate, which in turn depends on the potential segmentation approach incorporated in the cervical cancer cell screening process [6]. Specifically, the segmentation of cervical Pap smear cells using a graph cut approach was identified to be remarkable in obtaining a good classification rate in the detection process of cervical cancer [7–9]. However, the Pap smear cells of cervical cancer are segmented using graph cut such as normalized graph cut and min-maximum cut set are determined to be incapable of resolving the problems which emerge because of the unstable staining, substandard quality and features that are contrasting of Pap smear cell images [10]. Hence, a significant detection of cervical cancer cells using the graph cut-based segmentation approach needs to be formulated to ensure an optimal classification accuracy rate.

The proposed technique also helps identify the pre-cancerous cells so that the mortality rate in detecting cervical cancer can be reduced. The screening of cervical cancer is processed through the Pap Smear test or liquid cytology-based lesions. The features of the cells, such as texture, color, size and nucleus cytoplasm (n/c) ratio, are extracted and classified with the severity of the stage in cancer so that it can be treated at the earliest.

In this paper, a Hybrid Linear Iterative Clustering and Bayes Classification-based GrabCut Segmentation Technique (HLC-BC-GCST) is proposed for significant extraction of cytoplasm and nuclei of Pap smear cells by deriving the benefits from Simple Linear Iterative Clustering Process (SLICP), GrabCut-based graph cut model, Bayes classification [11] and minimum flow algorithms in a systematic manner to facilitate an effective segmentation in the process of cervical cancer detection. In this proposed HLC-BC-GCST approach, first, SLICP is used to ensure rapid clustering and sustain the pixel block count during segmenting Pap smear cells. Then the GrabCut is used for formulating a graph cut model based on which the error is minimized to a significant degree using an integrated Bayes classifier. Further, iterative enforcement of the GrabCut-based graph cut model and the second level of SLICP are employed to improve the quality of clustered pixel points considered from the cervical cancer Pap smear cells. Finally, two potential Gaussian Mixture Models (GMM) [12] are determined to achieve the maximum degree of segmentation by applying the minimum graph cut algorithm that determines the optimal boundaries from the Pap smear cell test. The experimental investigation of the proposed HLC-BC-GCST approach achieves better results when compared with the standard metrics.

The mathematical model developed for implementing the proposed method HLC-BC-GCST was derived from published or other sources to enable the proposed model for economic incorporation and be feasible for clinical trial implementation. The cost ratio is identified as incremental with the effectiveness of cost, and the year life cost

gained per year estimates the overall cost-effectiveness of the model with the impact of the proposed method.

The paper is structured as follows: The existing segmentation schemes using graph cut algorithms are summarized with their pros and cons in Section 2. The sequential steps involved in implementing the proposed HLC-BC-GCST approach with their significance are presented in Section 3. Section 4 highlights the results of the experiments conducted to quantify the potential of the proposed HLC-BC-GCST approach over the considered baseline detection schemes used to detect cervical cancer. Section 5 depicts concluding remarks on the potential contribution of the proposed HLC-BC-GCST approach with its feasible plan of future research focus.

2. Previous Related Work

This section discusses some of the most recent cervical cancer detection schemes in the literature with their merits and limitations. In 2008, Marinakis [13] developed a Particle Swarm Optimization-based meta-heuristic approach for accurate Pap smear cell classification (PSO-CCD). This PSO-CCD scheme was integrated with the nearest neighbor-oriented classifier for the appropriate extraction of cytoplasm and nuclei regions of the cervical cells. The classification accuracy of this PSO-CCD scheme was confirmed to be 97.64%, which is comparatively greater than the cervical cancer detection approaches propounded during the years 2001 to 2007. Marinakis et al. proposed another cervical cancer detection in 2009 for detecting cervical cancer using an optimization process of the neighbor classifiers [14]. This cervical cancer detection scheme utilized an intelligent nature-inspired approach to effectively and appropriately determine the Pap smear cell boundaries. The nature-inspired meta-heuristic approaches, such as particle swarm optimization, genetic algorithms, tabu search and ant colony algorithms, are used to achieve predominant classification accuracy. In [15], an automation-assisted cervical cancer detection was proposed using Fuzzy C-Means to segment the Pap smear cells into their background, cytoplasm and nucleus. This Fuzzy C-Means scheme was proposed to handle the issue of color-stained Pap smear cells that result in inappropriate detection of a nucleus and cytoplasmic boundaries. This Fuzzy C-Means algorithm with an Artificial Neural Network (ANN) was determined to be 96.20% and 97.23% accurate during the enforcement in resolving the issues of two classes and four classes considered from the Herlev dataset [15]. An integrated Global and Local Graphs segmentation scheme using graph cut (GLGC-CCDT) for dynamic detection of cervical cancer diagnosis was introduced to classify abnormal cells from healthy cells in [16]. This GLGC-CCDT approach enabled a global multi-perspective graph cut using a* channel base for accurate cytoplasmic segmentation.

In addition, a Neutrosophic Adaptive Mean Shift Cut-based Cervical Cancer Detection Technique (NAMSC-CCDT) was contributed in [17] to facilitate an effective segmentation of Pap smear cells. In this NAMSC-CCDT, the image is first transformed into a neutrosophic set represented through true, false and indeterminacy membership values. Then the indeterminacy membership value and neighborhood characteristics are employed to minimize the indeterminacy value of the Pap smear cells [18] considered for investigation. Finally, a mean-shift clustering scheme was used in [19] to improve the potential of the derived neutrosophic set to classify the pixels adaptively. The classification accuracy, specificity [20] and sensitivity of this NAMSC-CCDT were concluded to be 98.32%, 97.84% and 98.03%, respectively. A Dual-Level Cascade Classification Approach (DLCCA) was proposed in [21] for the segmentation of Pap smear cells to aid an effective cervical cancer detection. This DLCCA approach utilized 20 morphological features, eight textures-based features and 28-dimensional characteristics to ensure an effective segmentation process. The DLCCA approach utilized two levels of a cascading process for the exact determination of nucleus and cytoplasmic boundaries. This DLCCA approach also used the benefits of a rapid C4.5 classifier for better and precise classification between healthy and abnormal cells. The classification accuracy, specificity and sensitivity of this NAMSC-CCDT were identified as 98.55%, 97.21%, and 98.09%, respectively.

3. Proposed Hybrid Linear Iterative Clustering and Bayes Classification-Based GrabCut Segmentation Technique

In this proposed HLC-BC-GCST scheme, an effective segmentation of the cervical Pap smear cells is done through the following six steps: (i) Employing a Simple Linear Iterative Clustering Process (SLICP) for effective image clustering, (ii) Construction of the optimized graph cut model derived from a mean value of each clustered pixel block, (iii) Classification of hyper pixels in the optimized graph cut model using a Bayes Classifier, (iv) Second level enforcement of SLICP for repeated image clustering process, (v) Estimation of Gaussian mixture model (GMM) over clustered images and (vi) Implementation of the minimum graph cut algorithm for optimal cervical Pap smear image segmentation. These steps are shown in Algorithm 1.

Algorithm 1. Proposed HLC-BC-GCS Scheme

Input-Cervical smear image

C_f -Compact factor, S_{PC} -Super pixel count. P_{DI} -Individual pixel distance, I -Number of iterations, $G(l, m)$ -Gradient of the image

U_r -Unknown region, B_r -Background region, E_f -Energy function

Output: Regions of Nuclei, Cytoplasm

Stage 0: Conversion from RGB to CIELAB space

Stage 1: Initial preprocessing with SLICP

Stage 2: Simple linear iterative clustering process (SLICP)

Step 1: Calculate the C_f and S_{PC}

Step 2: Calculate the inter-pixel individual distance of P_{DI}

Step 3: The seed point's generation has to be transited with the new shortest distance

Step 4: Recalculate the new seed point generation using $G(l, m)$

Stage 3: GrabCut-based Bayes classifier

Step 1: Construct a (S, t) graph

Step 2: Select the unknown U_r and background region B_r

Step 3: Select the potential clustered region

Step 4: Use the Bayes Classifier for the generation of maximum outliers

Step 5: If P_r is not the maximum value, then the clustering process needs to be enhanced

Repeat Second SLICP

From Stage 1 step 1 to step 4

Until P_r is maximum

Stage 4: Energy minimization

Input parameters ()

Step 1: Estimate the GMM parameters

Step 2: Map the boundary value and priority factor with the energy function E_f

Step 3: If B_k is not equal to B_n

Step 4: then the condition is satisfied

Repeat from

Step 1 to step 3

Break

Until Convergence

3.1. Simple Linear Iterative Clustering Process (SLICP) for Effective Image Clustering

In this initial phase of the proposed HPLC-BC-GCST scheme, the process of iterative and effective clustering is enabled by the SLICP mechanism. This SLICP is the effective clustering process that is enhanced based on the benefits of the K-means algorithm. This incorporated SLICP can facilitate rapid clustering speed, maintaining several pixel blocks while segmenting Pap smear cells, automating clustering of Pap smear images and locating smooth precision edges of cytoplasm and nuclei in the Pap smear cells. This utilized SLICP is one of the widely used effective preprocessing techniques in image engineering.

In SLICP, two significant parameters, i.e., the compact factor C_f and the superpixel count SP_C , are used for clustering images. The utilization of the superpixel count SP_C aids in controlling the number of partitioned image area pixel blocks, and the compact factor

C_F helps formulate the fit degree, which needs to be dynamically updated for framing the pixel area. The increase in the value of the compact factor C_F leads to enhanced quality of the pixel blocks. In the initial part of implementing SLICP, the input color cervical Pap smear image is transformed into its lab space that represents the lightness (l_p), green-red (c), and blue-yellow (d) parameters into a numerical value. Further, the input colour image is partitioned into superpixels based on the assignment of superpixel count SP_C . The size of the superpixel resulting after the process of partitioning is defined through (1)

$$SP(Size) = \left(\frac{N_A}{SP_C} * \frac{S_U}{2}\right) \tag{1}$$

where N_A and S_U highlight the complete area of the input cervical Pap smear cell and the number of seed points that could be derived during this partitioning process. In this context, the seed points represent the position information (l, m) with lab space-related parameters of l_p, c , and d , respectively. In turn, the value of each superpixel depends on the number of seed points in each row of the image, the rank of each pixel in the image and the size of the derived pixel block. Furthermore, the classification intensity is assigned to -1 with the complete number of pixel distances $P_{d(i)}$ set to ∞ respectively. Thus, the pixels in the $2SP(Size) * 2SP(Size)$ region are estimated to be the constituent of the background or target based on each pixel distance $P_{d(i)}$ and represent the location of the seed points. This essential individual pixel distance $P_{d(i)}$ in the process of SLICP is computed based on (2).

$$P_{d(i)} = \sqrt{(Ip_{(d)})^2 + \frac{(Ip_{(d)})^2}{E_{SL}} * C_F} \tag{2}$$

where $Ip_{(d)}$ and E_{SL} represent the inter-pixel distance and the expected side measure of each pixel considered for investigation. Then the inter-pixel distance $Ip_{(d)}$ and difference in the value of the lab space $Ip_{(ls)}$ are determined based on (3) and (4), respectively.

$$Ip_{(d)} = \sqrt{(Ip_{(j)} - Ip_{(i)})^2 + (c_j - c_i)^2 + (d_j - d_i)^2} \tag{3}$$

$$Ip_{(ls)} = \sqrt{(l_j - l_i)^2 + (m_j + m_i)^2} \tag{4}$$

However, the location of the seed points needs to be recalculated, and new seed points need to be transited towards the shortest distance approaching the gradient to prevent clustering error over the edges of the cytoplasm nuclei in the cervical Pap smear cells. Thus, the location of the seed points is recalculated based on the gradient of the image $G(l, m)$ described in (5).

$$G(l, m) = \|(I_p(l + 1, m) - I_p(l - 1, m))\|^2 + \|(I_p(l, m + 1) - I_p(l, m - 1))\|^2 \tag{5}$$

3.2. Construction of the Optimized Graph Cut Model Derived from the Mean Value of Each Clustered Pixel Block

In this phase, a graph cut model called GrabCut is used as the interactive segmentation algorithm that maps the clustered pixel points of the input cervical Pap smear image into a network diagram. This GrabCut-based network consists of nodes and edges derived from the input image’s pixel points. In this GrabCut-based graph cut network, two nodes, such as the source and sink (s, d), are designated for determining the path to select a particular path feature for detecting cervical cancer during the employment of the minimum graph cut algorithm. The network’s vertices represent the pixel points, and the edge highlights the associate degree existing between the clustered pixel points. This GrabCut-based network is utilized for converting it into a function in which each vertex $G_V(s, d)$ is mapped onto the value of $\{0,1\}$.

Further, a potential region needs to be selected in the clustered pixel points. The region inside the focused area is Unknown Region (U_R), and the area outside the concentrated

region is Background Region (B_R). Furthermore, GMM features of the Unknown Region (U_R) and Background Region (B_R) are initialized based on the computed association between pixel points. In addition, the Unknown Region (U_R) is also divided into target and background regions, which are demerged based on the application of minimum flow and maximum flow graph cut methods. In this context, the GMM features of the Unknown Region (U_R) and Background Region (B_R) are updated periodically. The GrabCut-based graph cut network is partitioned, and its related GMM factors are determined during the convergence of the algorithms. This GrabCut-based graph cut model is mainly used after clustering pixel points to minimize the overhead incurred in the segmentation process with improved precision.

3.3. Bayes Classification of Hyper Pixels in the Optimized Graph Cuts

In this phase, a traditional classifier like the k-NN classifier is utilized for partitioning each component of the image histogram into two categories of GMM features, such that the key significance of the formulated GrabCut-based graph cut network is represented in the foreground [18]. In this scenario, the component of the image histogram is inferred to be mutually independent with equal probabilistic values only when each pixel value corresponding to background and foreground is equally probable and mutually independent in nature. The components related to the class that could not be enforced during partitioning in the data samples are represented using a k-dimensional feature vector $F_{V(k)} = [F_{V(1)}, F_{V(2)}, \dots, F_{V(k)}]$ that consists of k attributes $[A_1, A_2, \dots, A_k]$ in them. The number of unknown data samples deployed in the background and foreground of the image histogram is derived based on (6).

$$P(R_{(i)}|F_{V(k)}) = \text{Max}_{j \in \{0,1\}} (P(R_{(j)}|F_{V(k)})) \quad (6)$$

where the values of i and j are determined to be 0 or 1, respectively, if the value of i and j is equal to 0, then the estimated unknown data sample is distributed in the background, and in contrast, the estimated unknown data sample is distributed in the foreground when the value ij is equal to 1. The components of the same partition are highlighted with a similar token value, and the tokens possessed by the same components are determined to possess the same value. In this phase, significant performance is ensured since (i) It is capable of generating the maximum number of outliers, (ii) It decreases the discrepancy association existing between regions of the image histogram, and (iii) It prevents the process of stripping off the target and the background region. However, the error in the process of classification results in a reduced number of hyper pixels; hence, the Bayes classifier scheme is employed to minimize the error estimated during the classification process. This Bayes classification scheme permits the possibility of reclassifying a huge number of isolated pixel points of the image to enhance the degree of discrepancy between regions of investigation. The Bayes classifier improves the degree of discrepancy between regions for facilitating the prominent foreground of the simplified image. Thus, using the Bayes classification scheme improves the effective and efficient segmentation rate. Furthermore, the segmented image after the process of clustering, GrabCut-based graph cut representation and Bayes classification-based error minimization might also be low due to the poor performance of the incorporation of the GrabCut-based graph cut in the proposed HLC-BC-GCST schemes. Hence, the process of GrabCut-based graph cut needs to be improved again based on the SLICP process.

3.4. Second Level Enforcement of SLICP within the Repeated Image Clustering Process

In this phase, the second level of enforcement of the SLICP process is enabled to decrease the time incurred in the segmentation process and ensure an enhanced rate of segmentation accuracy. This second level of SLICP enforcement is also essential for improving the segmentation quality that completely depends on the estimation of the GMM features of the hyper pixel points. In the first phase, SLICP is first utilized for processing

the input image by considering specific points of significance rather than the hyper pixel block derived from the input Pap smear cell image. In this phase, the SLICP process is again used for the second time to enhance the segmentation quality by reducing the error. This possible degree that could be estimated during segmentation is mainly due to the reduced number of reborn hyper pixels of the images generated during the process of segmentation. To improve the number of reborn hyper pixels, the process of SLICP is used for estimating the mean values of the pixel points in each pixel block such that the value of each pixel point and the value of the isolated points are further reduced. The iterative use of GrabCut-based graph cut again introduces the possibility of appropriate estimation of GMM parameters. Incorporating the iterative GrabCut-based graph-cut approach also increases the possibility of decreasing the time and computational complexity involved in the segmentation process.

3.5. Estimation of Gaussian Mixture Model Parameters (GMM) over Clustered Images

After the iterative enforcement of the GrabCut-based graph cut scheme and second-level application of SLICP, the GMM parameters need to be estimated. In this proposed scheme, two GMM parameters related to the foreground and background consisting of ‘m’ Gaussian models are determined using kn factors pertaining to each pixel. Further, the energy function minimization is essential during the process of image segmentation when the utilized GrabCut-based graph cut scheme is employed over the clustered images as defined in (7)

$$E_n(F) = \alpha * G(F) + H(F) \tag{7}$$

where α is the factor of balance with $G(F)$ and $H(F)$ corresponding to the term of region and boundary.

Then the Gibbs energy function [19–21] is employed in the proposed HLC-BC-GCST scheme using Equation (8)

$$E_n(\beta, \delta, \theta, M) = V(\beta, \delta, \theta, M) + W(\beta, \delta, \theta, M) \tag{8}$$

where $V(\beta, \delta, \theta, M)$ and $W(\beta, \delta, \theta, M)$ relate to the region and boundary of the Pap smear cell considered for investigation. Furthermore, the region term $V(\beta, \delta, \theta, M)$ used in the proposed HLC-BC-GCST scheme is expressed in (9)

$$V(\beta, \delta, \theta, M) = \sum_m C(\beta_m, \delta_m, \theta, M_m) \tag{9}$$

where the value of $C(\beta_n, \delta_n, \theta, M_n)$ is determined based on (10)

$$C(\beta_n, \delta_n, \theta, M_n) = -\log \pi(\beta_m, \delta_m) + \left(\frac{\log \det \Sigma(\beta_m, \delta_m)}{2}\right) + \frac{(M_m - \omega(\beta_m, \delta_m))^T}{2} + \sum (\beta_m, \delta_m)^{-1} [M_m - \omega(\beta_m, \delta_m)] \tag{10}$$

where $\pi(\beta_m, \delta_m)$, $\omega(\beta_m, \delta_m)$ and $\Sigma(\beta_m, \delta_m)$ represents the mixed priority factor, an average of the Gaussian Model and covariance matrix with θ derived using (11)

$$\theta = [\pi(\beta_m, \delta_m), \omega(\beta_m, \delta_m), \Sigma(\beta_m, \delta_m)] \tag{11}$$

In addition, the boundary value $W(\beta, \delta, \theta, M)$ related to the Pap smear cell considered for investigation is derived based on (12)

$$W(\beta, \delta, \theta, M) = \psi \sum_{k,n \in C} [\beta_k \neq \beta_n] \exp -\lambda \|M_k - M_n\| \tag{12}$$

where M is the considered neighbourhood pixel pair with $\beta = 50$ and $\lambda = (2\|M_k - M_n\|)^{-1}$ such that $\beta_k \neq \beta_n$ condition is satisfied.

3.6. Implementation of the Minimum Graph Cut Algorithm for Optimal Cervical Pap Smear Image Segmentation

Finally, the minimum graph cut algorithm [20] is applied after estimating GMM parameters for optimal cervical Pap smear image segmentation. This minimum graph cut algorithm is implemented over the GrabCut modelled network to estimate the maximum flow between the source and destination (s, d). In this minimum graph cut algorithm, all the possible numbers of minimum (s, d) cut sets are determined between the source and destination pairs to optimally identify the minimum cut set. Thus, the minimum graph cut algorithm is utilized for the optimal segmentation of cancer cells from the cervical pap smear cells.

4. Experimental Results

In this section, the proposed HLC-BC-GCST scheme towards an effective segmentation process of cytoplasm and nuclei boundaries in cervical cancer detection is evaluated and compared with the existing baseline techniques GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT using the Herlev dataset. The Herlev dataset is mainly used in investigating the proposed HLC-BC-GCST scheme because it is the predominant dataset used in most automated schemes proposed for effectively detecting cervical cancer. This popular Herlev dataset consists of approximately 917 cervical Pap smear cells gathered and stored in the Herlev University Hospital over the years. This Herlev dataset contains seven types of cervical Pap smear cells corresponding to severe dysplasia, mild dysplasia, moderate dysplasia, intermediate epithelium, columnar epithelium, and the superficial and columnar intensity of infected cells. More specifically, approximately 146, 150, 182 and 192 cervical cell images out of the complete 917 cervical moderate dysplasia and carcinoma in nature. Matlab 2013 is used to investigate the proposed HLC-BC-GCST scheme over the benchmarked GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques for evaluating its performance. Cell images are referred to as severe dysplasia and mild dysplasia using the Herlev dataset (Figures 1 and 2).

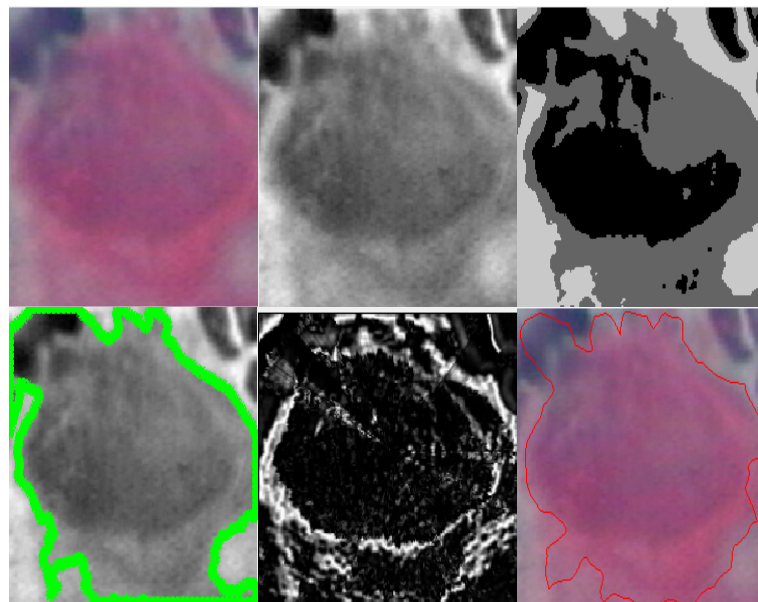


Figure 1. From left to right images represents the carcinoma cell from the Herlev dataset where the input image is given in CIE Lab colour space, and where initially the linear iterative clustering is applied to separate the regions of nucleus and cytoplasm, the optimized graph cut is generating the hyper pixels, and a Bayesian classifier is used to classify the boundaries of nucleus and cytoplasm where the high intensity darker pixels are regions of a nucleus and the red boundary marked is the region of the cytoplasm.

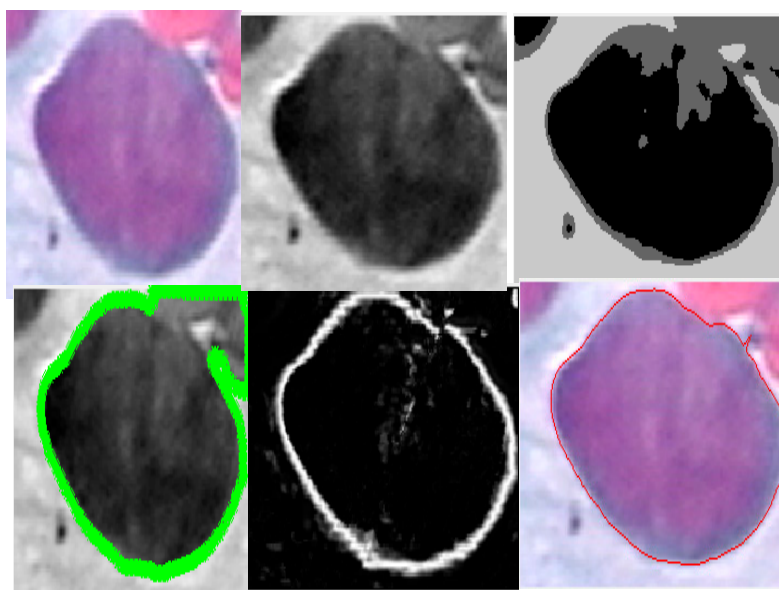


Figure 2. From left to right images represents the moderate dysplasia cell from the Herlev database, where the segmentation of nucleus and cytoplasm regions is performed using the simple iterative clustering initially, where the regions are separated, an optimized GrabCut is used for the hyper pixels generation with a Gaussian mixture model, and finally, the Bayes classifier is used to identify the nucleus and cytoplasm regions, where the dark intensity regions are represented as a nucleus and the boundary marked in red is identified as cytoplasm.

Figures 1 and 2 explain the potential of the HLC-BC-GCST scheme evaluated using classification accuracy and sensitivity under varying rounds in the implementation process. The classification accuracy of the proposed HLC-BC-GCST scheme is determined to be 99.51%. This classification accuracy of 99.51% facilitated by the proposed HLC-BC-GCST scheme is determined to be 3.56% excellent compared to the existing benchmarked GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques used for comparison. This improved classification accuracy rate was made feasible by using the Bayes classifier for error minimization and multiple applications of the SLICP process during the segmentation process. Similarly, the HLC-BC-GCS scheme obtains a sensitivity that is determined to be 98.81%. This sensitivity rate of 98.81% enabled by the proposed HLC-BC-GCST scheme is determined to be 4.32% predominant baseline GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques are analyzed for comparison. The enhanced sensitivity rate in the proposed HLC-BC-GCST scheme is mainly due to the incorporation of multiple GrabCut-based graph-cut approaches and potential estimation of GMM parameters during the process of segmenting the cytoplasm and nucleus boundaries that aids in cervical cancer detection. The representation of their cells in the graph cut is presented in Figure 3. Figures 4 and 5 presents the performance of the proposed HLC-BC-GCST scheme evaluated using classification accuracy and sensitivity under a varying number of rounds in the implementation process.

Figures 6 and 7 depict the performance of the proposed HLC-BC-GCST scheme evaluated using specificity and mean processing time incurred per image under various rounds in the implementation process. The specificity of the proposed HLC-BC-GCST scheme is determined to be 99.11%. This specificity of the proposed HLC-BC-GCST scheme is 4.13% excellent compared to the existing benchmarked GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques used for comparison. The mean processing time incurred by the proposed image of the HLC-BC-GCST scheme is 2.26 milliseconds. This mean processing time of the proposed HLC-BC-GCST is obtained by 3.78%, remarkably compared to the GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques used for comparison. This improved rate of specificity and mean processing rate

achieved by the proposed HLC-BC-GCST scheme is mainly due to the enforcement of the iterative SLICP, Bayes classifier and efficient GMM parameter estimation during the segmentation process.

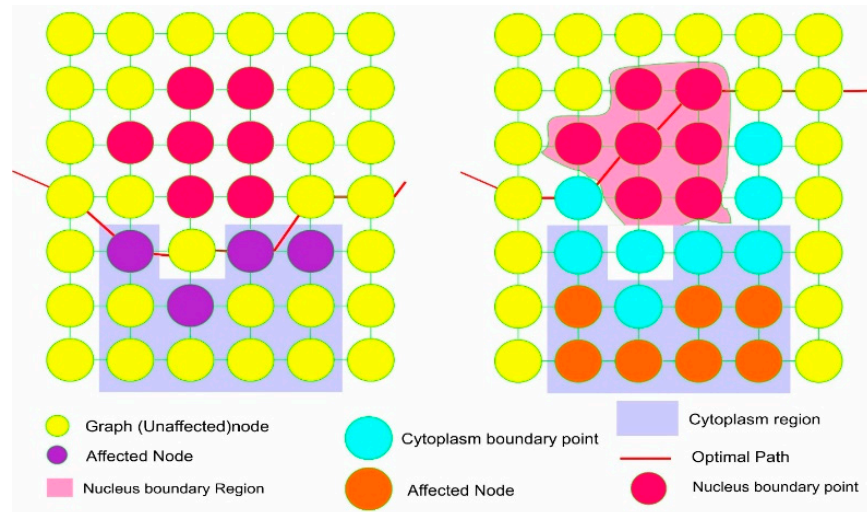


Figure 3. Representation of graph cut.

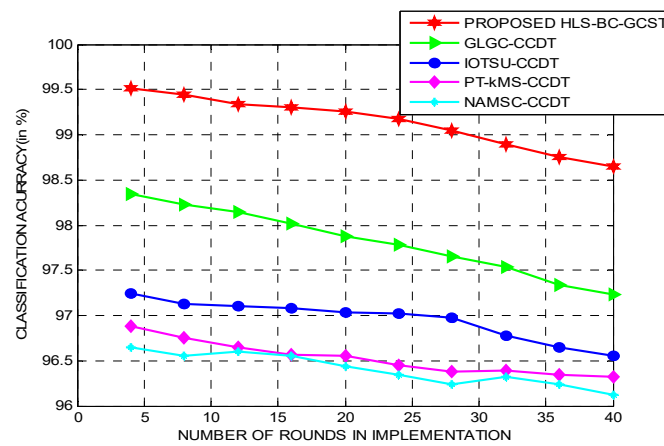


Figure 4. Comparative performance of the proposed HLC-BC-GCST scheme (using classification accuracy).

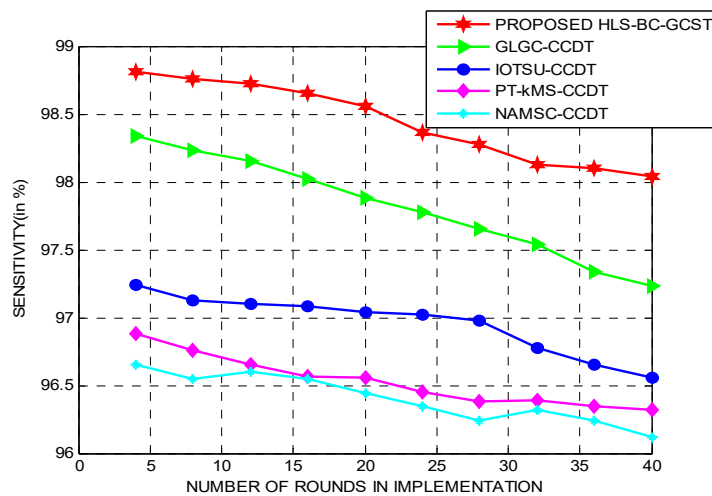


Figure 5. Comparative performance of the proposed HLC-BC-GCST scheme (using sensitivity).

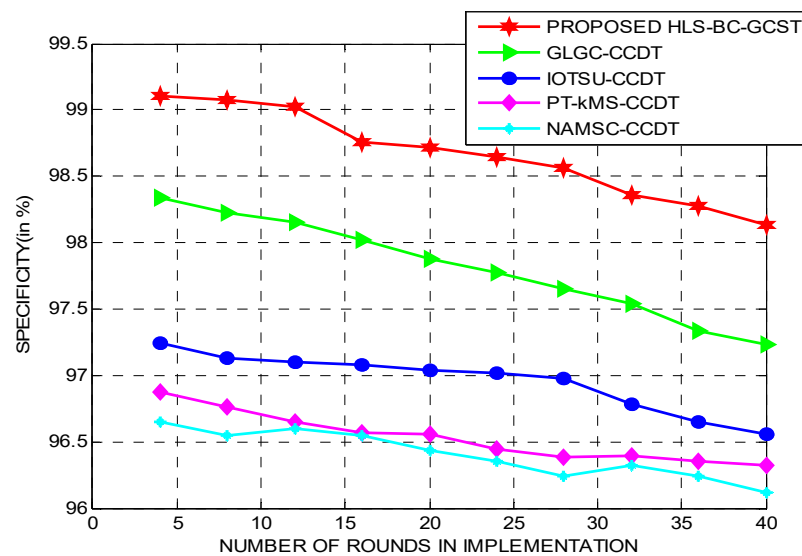


Figure 6. The proposed HLC-BC-GCST scheme evaluated using specificity.

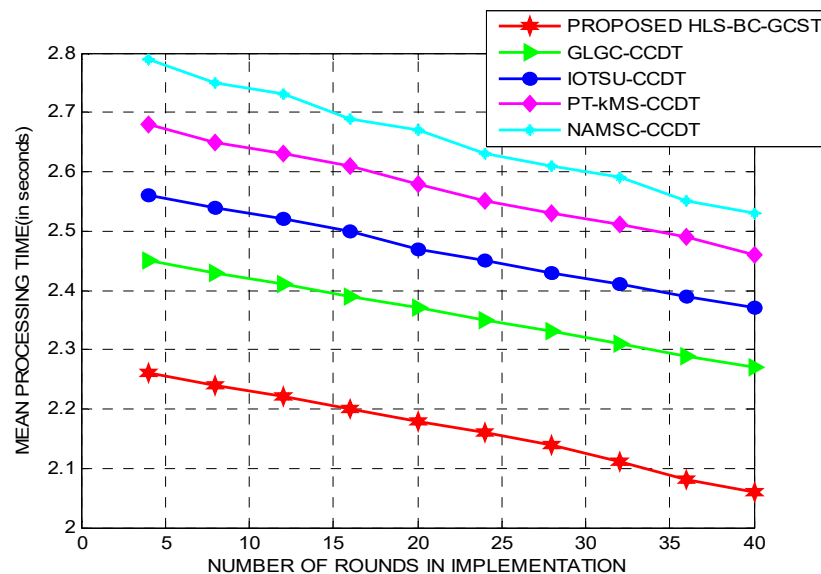


Figure 7. The proposed HLC-BC-GCST scheme evaluated using mean processing time.

Figures 8 and 9 highlight the significance of the proposed HLC-BC-GCST scheme evaluated using precision and recall values under the varying number of rounds in the implementation process. The precision value of this proposed HLC-BC-GCST scheme is estimated as 0.96 ± 0.19 , which is enhanced to a maximum level of 4.97%, remarkable to the baseline GLGC-CCDT IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques. In particular, the precision values of the proposed HLC-BC-GCST scheme are determined to be 0.96 ± 0.03 , 0.95 ± 0.12 , 0.94 ± 0.08 and 0.94 ± 0.18 under cervical images of severe dysplasia, moderate dysplasia, mild dysplasia and carcinoma used for analysis. Likewise, the recall value in the HLC-BC-GCST scheme is estimated as 0.97 ± 0.11 , which is enhanced to a maximum level of 5.12% predominant to the baseline GLGC-CCDT, IOTSU-CCDT, PT-KMS-CCDT and NAMSC-CCDT techniques. In particular, the recall value in HLC-BC-GCST is determined by 0.97 ± 0.15 , 0.96 ± 0.13 , 0.95 ± 0.05 and 0.95 ± 0.17 under cervical images of severe dysplasia, moderate dysplasia, mild dysplasia and carcinoma are analyzed for investigation.

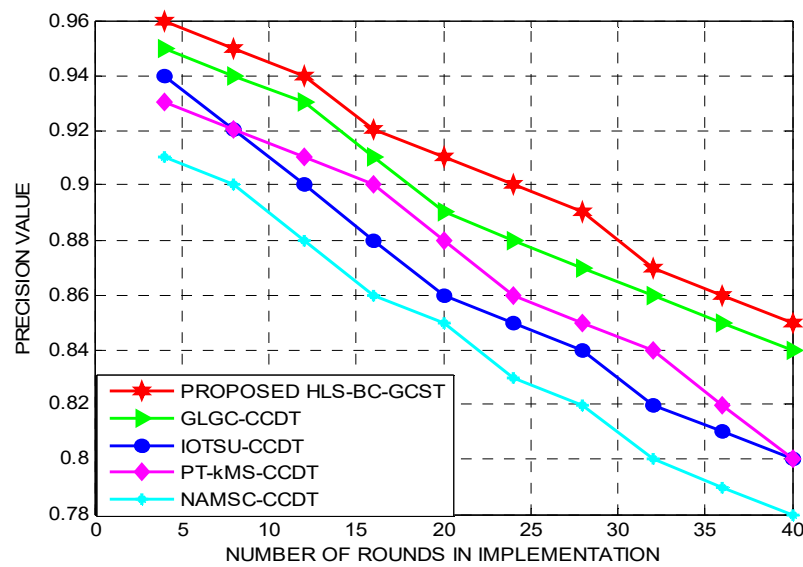


Figure 8. The proposed HLC-BC-GCST scheme evaluated using a precision value.

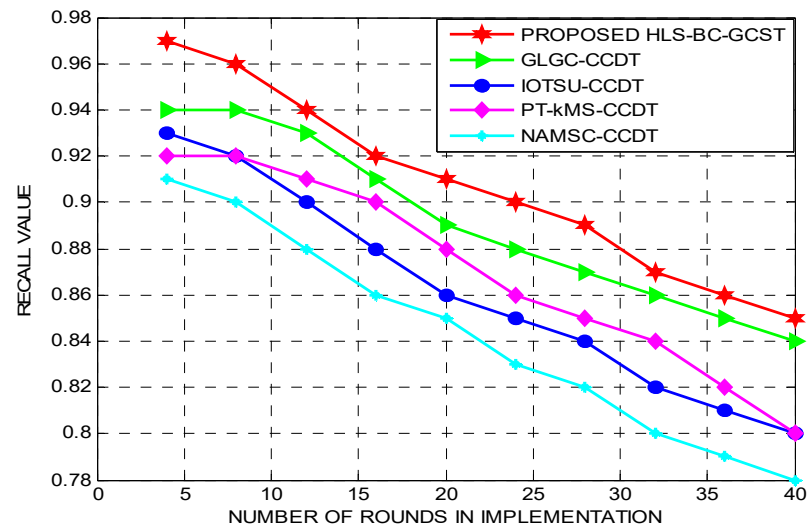


Figure 9. Proposed HLC-BC-GCST scheme evaluated using recall value.

Tables 1–3 highlight the performance of the proposed HLC-BC-GCST scheme, where different metrics are used to compare and evaluate the existing systems’ performance over the proposed system. The results obtained by the proposed system show an increase in the overall performance accuracy of the detection of cervical cancer.

Table 1. Comparative Performance: Accuracy and Specificity.

First Author and Year	Classification Accuracy (%)	Specificity (%)
<i>Proposed HLC-BC-GCST</i>	99.51	99.11
Marinakis, 2008 [13]	98.56	98.13
Marinakis, 2009 [14]	98.73	98.24
Chankong, 2014 [15]	98.92	97.86
Zhang, 2014 [16]	98.43	97.83
Guo, 2018 [21]	98.32	97.84
Ray, 2018 [21]	98.55	97.21

Table 2. Comparative performance: sensitivity and mean processing time.

First author and Year	Sensitivity (in %)	Meantime for Processing (in a s)
<i>Proposed HLC-BC-GCST</i>	98.84	2.26
Marinakis, 2008 [13]	98.12	3.42
Marinakis, 2009 [14]	97.86	4.32
Chankong, 2014 [15]	97.43	3.78
Zhang, 2014 [16]	97.23	3.65
Guo, 2018 [21]	98.03	2.96
Ray, 2018 [17]	98.09	2.92

Table 3. Comparative Performance: Precision and Recall Value.

First Author and Year	Precision	Recall Value
<i>Proposed HLC-BC-GCST</i>	0.96 ± 0.19	0.97 ± 0.12
Marinakis, 2008 [13]	0.92 ± 0.11	0.94 ± 0.09
Marinakis, 2009 [14]	0.92 ± 0.19	0.94 ± 0.13
Chankong, 2014 [15]	0.92 ± 0.21	0.92 ± 0.11
Zhang, 2014 [16]	0.94 ± 0.13	0.93 ± 0.12
Guo, 2018 [21]	0.95 ± 0.15	0.94 ± 0.16
Ray, 2018 [17]	0.94 ± 0.19	0.94 ± 0.14

The summarized results in Tables 1–3 confirm that the classification accuracy, specificity, sensitivity, mean processing time, precision, and recall value of the proposed HLC-BC-GCST scheme are better in mean rate of 3.21%, 4.52%, 3.84%, 0.82 s, 3.42% and 4.12% that are excellent compared to the cervical cancer detection approaches detailed in the related work section.

5. Conclusions

The HLC-BC-GCST scheme was presented in this paper as a high-performing cervical cancer detection approach that utilizes the key benefits of the GrabCut-based graph cut set for concentrating on the segmentation process that aids in finding accurate cytoplasm and nucleus boundaries. The iterative construction of GrabCut-based graph cut and the dual-level enforcement of SLICP were employed to improve the quality of pixel points in the Pap smear cells considered for the analysis. The Bayes classifier used in this HLC-BC-GCST scheme mainly focused on minimizing errors incurred during the derivation of cytoplasm and nucleus boundaries. Identifying GMM parameters and applying the minimum graph cut algorithm confirmed an optimal segmentation in detecting cervical cancer cells. The results of the proposed HLC-BC-GCST scheme were also confirmed to enhance the classification accuracy, specificity, sensitivity and processing time with an average of 23%, 21%, 18% and 25%, compared to the benchmarked schemes considered for investigation. As a part of future work, an enhanced Boykov's Graph cut-based segmentation scheme is planned to be investigated for an optimal segmentation of Pap smear cells that play a vital role in cervical cancer detection.

Author Contributions: Conceptualization, A.D.M., E.R., V.K., S.J.K., K.K., S.V., M.M. and V.P.; methodology, A.D.M., E.R., V.K., S.J.K., K.K., S.V., M.M. and V.P.; investigation, A.D.M., E.R., V.K., S.J.K., K.K., S.V., M.M. and V.P.; writing—original draft preparation, A.D.M., E.R., V.K., S.J.K., K.K., S.V., M.M. and V.P.; writing—review and editing, A.D.M., E.R., V.K., S.J.K., K.K., S.V., M.M. and V.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Mishra, G.A.; Pimple, S.A.; Shastri, S.S. An overview of prevention and early detection of cervical cancers. *Indian J. Med. Paediatr. Oncol.* **2011**, *32*, 125. [[CrossRef](#)] [[PubMed](#)]
2. Mbagi, A.H.; ZhiJun, P. Pap Smear Images Classification for Early Detection of Cervical Cancer. *Int. J. Comput. Appl.* **2015**, *118*, 10–16.
3. Sajeena, T.A.; Jereesh, A.S. Automated cervical cancer detection through RGVF segmentation and SVM classification. In Proceedings of the 2015 International Conference on Computing and Network Communications (CoCoNet), Trivandrum, Kerala, India, 16–19 December 2015; Volume 1, pp. 67–75.
4. Zhang, L.; Kong, H.; Liu, S.; Wang, T.; Chen, S.; Sonka, M. Graph-based Segmentation of abnormal nuclei in cervical cytology. *Comput. Med. Imaging Graph.* **2017**, *56*, 38–48. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, L.; Liu, S.; Wang, T.; Chen, S.; Sonka, M. Improved Segmentation of abnormal cervical nuclei using a graph-search based approach. In *Medical Imaging 2015: Digital Pathology*; SPIE: Orlando, FL, USA, 2015; Volume 9420, pp. 221–227. [[CrossRef](#)]
6. Garg, S.; Urooj, S.; Vijay, R. A Software-Based Novel Approach: Integrated Segmentation & Nuclei Extraction of Overlapped Cervical Cell in High-Resolution MRI Images. *Int. J. Eng. Trends Technol.* **2017**, *49*, 206–211.
7. Islam, Z.; Haque, M.A. Multi-step level set method for Segmentation of overlapping cervical cells. In Proceedings of the 2015 IEEE International Conference on Telecommunications and Photonics (ICTP), Dhaka, Bangladesh, 26–28 December 2015; Volume 1, pp. 78–85.
8. Zhang, L.; Sonka, M.; Lu, L.; Summers, R.M.; Yao, J. Combining fully convolutional networks and graph-based approach for automated Segmentation of cervical cell nuclei. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; Volume 1, pp. 98–112.
9. Song, Y.; Zhang, L.; Chen, S.; Ni, D.; Lei, B.; Wang, T. Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning. In Proceedings of the IEEE Transactions on Biomedical Engineering, Beirut, Lebanon, 16–18 September 2015; Volume 62, pp. 2421–2433.
10. Ozlem, A.; Umit, I. Comparative analysis of cervical cytology screening methods and staining protocols for detection rate and accurate interpretation of ASC-H: Data from a high-volume laboratory in Turkey. *Diagn. Cytopathol.* **2015**, *43*, 863–869. [[CrossRef](#)]
11. Kharya, S.; Soni, S. Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection. *Int. J. Comput. Appl.* **2016**, *133*, 32–37. [[CrossRef](#)]
12. Ragothaman, S.; Narasimhan, S.; Basavaraj, M.G.; Dewar, R. Unsupervised Segmentation of Cervical Cell Images Using Gaussian Mixture Model. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; Volume 1, pp. 67–75.
13. Marinakis, Y.; Marinaki, M.; Dounias, G. Particle swarm optimization for pap-smear diagnosis. *Expert Syst. Appl.* **2008**, *35*, 1645–1656. [[CrossRef](#)]
14. Marinakis, Y.; Marinaki, M.; Dounias, G.; Jantzen, J.; Bjerregaard, B. Intelligent and nature-inspired optimization methods in medicine: The Pap smear cell classification problem. *Expert Syst.* **2009**, *26*, 433–457. [[CrossRef](#)]
15. Chankong, T.; Theera-Umpon, N.; Auephanwiriyakul, S. Automatic cervical cell segmentation and classification in Pap smears. *Comput. Methods Programs Biomed.* **2014**, *113*, 539–556. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, L.; Kong, H.; Chin, C.T.; Liu, S.; Chen, Z.; Wang, T.; Chen, S. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. *Comput. Med. Imaging Graph.* **2014**, *38*, 369–380. [[CrossRef](#)]
17. Ray, A.; Maitra, I.K.; Bhattacharyya, D. Detection of Cervical Cancer at an Early Stage Using Hybrid Segmentation Techniques from PAP Smear Images. *Int. J. Adv. Sci. Technol.* **2018**, *112*, 23–32. [[CrossRef](#)]
18. Akturk, S.M.; Aykut, M. An improvement on GrabCut interactive segmentation method based on input colour spaces. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; Volume 1, pp. 45–55.
19. Escolano, F.; Lozano, M. Optimal Image Segmentation Methods Based on Energy Minimization. *Adv. Image Video Segm.* **2017**, *1*, 78–85.
20. Shi, C.; Lin, W.; Li, X.; Wen, J. Polarized characteristics image segmentation based on minimum cut. *J. Comput. Appl.* **2018**, *30*, 1587–1589. [[CrossRef](#)]
21. Guo, Y.; Şengür, A.; Akbulut, Y.; Shipley, A. An effective colour image segmentation approach using neutrosophic adaptive mean shift clustering. *Measurement* **2018**, *119*, 28–40. [[CrossRef](#)]