

# Automated Segmentation of Nucleus and Cytoplasm of Cervical Cells from Pap-smear Images using A Quadtree Decomposition Approach

Wasswa William (✉ [wwasswa@must.ac.ug](mailto:wwasswa@must.ac.ug))

Mbarara University of Science and Technology

Andrew Ware

University of South Wales

Annabella Habinka Basaza-Ejiri

St. Augustine International University

Johnes Obungoloch

Mbarara University of Science and Technology

---

## Research Article

**Keywords:** Cell Segmentation, Quadtree, Pap-smear, Cervical cancer

**Posted Date:** November 8th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-955958/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **Automated Segmentation of Nucleus and Cytoplasm of Cervical Cells from Pap-smear Images using A Quadtree Decomposition Approach.**

Wasswa William<sup>1</sup>, Andrew Ware (PhD)<sup>2</sup>, Annabella Habinka Basaza-Ejiri (PhD)<sup>3</sup>, Johnes Obungoloch (PhD)<sup>1</sup>

<sup>1</sup>*Department of Biomedical Sciences and Engineering, Mbarara University of Science and Technology,  
Mbarara, 1410, Uganda*

*Tel: +256775046515, Email: [wwasswa@must.ac.ug](mailto:wwasswa@must.ac.ug), [jobungoloch@must.ac.ug](mailto:jobungoloch@must.ac.ug)*

<sup>2</sup>*Faculty of Computing, Engineering and Science, University of South Wales, Prifysgol, UK*

*Tel: 01443 4 82650, Email: [andrew.ware@southwales.ac.uk](mailto:andrew.ware@southwales.ac.uk)*

<sup>3</sup>*College of Computing and Engineering, St. Augustine International University, Kampala, Uganda*

*Tel: +256772571444, Email: [aejiri@saiu.ac.ug](mailto:aejiri@saiu.ac.ug)*

*Corresponding author*

*(Wasswa William, Mbarara University of Science and Technology, Tel: +256775046515, Email: [wwasswa@must.ac.ug](mailto:wwasswa@must.ac.ug))*

## **Abstract**

**Background:** Digital pathology and microscopy image analysis is widely used for comprehensive studies of cell morphology especially for cervical cancer screening from pap-smears. Manual assessment of pap-smears is labour intensive and prone to interobserver variations. Computer-aided methods, which can significantly improve the objectivity and reproducibility, have attracted a great deal of interest in recent literature. A critical prerequisite in automated analysis of pap-smears is nucleus and cytoplasm segmentation, which is the basis of cervical cancer screening. This paper articulates a potent approach to the segmentation of cervical cells into nucleus and cytoplasm using a quadtree decomposition approach with statistical measures.

**Results:** Choosing an appropriate quadtree decomposition strategy was a great challenge and a novel task in the proposed approach. The image is pre-processed using an enhanced median filter and is decomposed based on the mean, maximum entropy and the variance statistical measures of the pixels in the subtree. As a result, highly efficient and segmentations of acceptable performance were obtained. Comparison of the segmented nucleus and cytoplasm with the ground truth nucleus and cytoplasm segmentations resulted into a Zijdenbos similarity index of greater than 0.9034 and 0.9498 for nucleus and cytoplasm segmentation respectively.

**Conclusion:** Given the accuracy of the classifier in segmenting the nucleus which plays an important role in cervical cancer diagnosis and classification, the classifier can be adapted for automated systems for cervical cancer diagnosis and classification. The method serves as a basis for first level segmentation of cervical cells for diagnosis and classification of cervical cancer from pap-smears.

**Keywords:** *Cell Segmentation, Quadtree, Pap-smear, Cervical cancer,*

## 1.0 Background

Cervical cancer is one of the most deadly and common forms of cancer among women in the world [1]. About 85% of the cases occur in developing countries [2]. Cervical cancer is preventable through a regular simple pap-smear screening test and there have been numerous attempts to automate the analysis of pap-smears since its introduction more than 70 years ago [3]. However, in many of the low middle-income countries, a pap-smear analysis is a manual process. This manual analysis is carried out by experienced cytotechnicians who are also very few in low middle-income countries. The manual visual examination of the pap-smears is time-consuming, labour intensive, subjective and error prone [4]. Due to the limitations of the manual pap-smear analysis, it is beneficial to develop a computer-assisted diagnosis system to make the pap-smear test more accurate and reliable.

Cell segmentation is one of the most important stages of such an automated system. However, segmentation of the cervical cells in a pap-smear digital image is an ill-posed problem. Very few of the current methods can achieve complete nucleus and cytoplasm segmentation due to a number of challenges involved in delineating individual nucleus and cytoplasm with severe overlap and poor contrast [5].

High-tech light and electron microscopes are capable of acquiring quality pap-smear images, but quantitatively evaluating these images often involves manually annotating structures of interest in the cell. This process is error-prone, time-consuming and is becoming the main bottleneck in the automated pap-smear analysis [6]. Cell nuclei segmentation from pap-smear images is an important image processing task necessary for automated cervical cancer screening due to the fundamentally important role of nuclei in cervical cancer cell [7]. The success of automated cervical cancer classification is often a direct consequence of the accuracy of the nucleus and cytoplasm segmentation. A cell segmentation paradigm involves separating a cell into regions or contours corresponding to different objects in the cell [8]. This is usually achieved by identifying common properties in an image or identifying differences between regions (edges) in the cell.

As the Biomedical Engineering field is growing, new imaging modalities and staining techniques are being developed, hence many existing methods specifically designed for current imaging modalities and staining techniques may not work well. Therefore, considerable resources have to be spent to modify existing methods or develop entirely new nuclei and cytoplasm segmentation methods to better suit the new applications. This study proposes a cervical cell nucleus and cytoplasm segmentation algorithm based on a quadtree region segmentation and statistical measures. The goal is to provide a method that can be used effectively for segmenting nuclei and cytoplasm from a cervical cell for automated diagnosis and classification of cervical cancer from pap-smears.

### **1.1 Review of some of the segmentation approaches applied to cervical cells.**

Image segmentation techniques can be classified as edge based, region growing, region split, watershed, thresholding, deformable model fitting and morphology-based [9]. A number of surveys have been carried out on the different image segmentation approaches. Shaoo et al. [10] surveyed only segmentation algorithms based on thresholding and attempted to evaluate the performance of some thresholding algorithms using some uniformity and shape measures. Eri et al. [11] presented a survey of cell segmentation techniques 50 years down the road. He showed that threshold based methods were the commonest form of segmentation techniques used for cell segmentation between 1960 and 2015. William et al. [12] presented a survey of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images.

The simplest property that pixels in an image can share is intensity. So many approaches [6-8] segment the image through thresholding which involves separation of light and dark regions. Wong et al [9] proposed an entropy based thresholding method. However, the results obtained by this approach were found to be biased. Ortiz et al. [10] presented an edge-based deformable model for cell segmentation. The approach utilized gradient information to capture nuclei and cytoplasm surfaces. However, the object boundaries were very blurred susceptible to noise. Zhang et al. [15] presented a nuclei segmentation algorithm where HSV colour space was used to enhance the contrast between nuclei and cytoplasm and nuclei segmentation was achieved using a concave point based algorithm. Plissiti et al. [16] presented an approach for cell nuclei segmentation where the detection of the candidate nuclei was based on a morphological operations and the segmentation was accomplished with the application of the watershed transform. Lin et al. [17] presented a novel nucleus and cytoplasm segmentation technique where image enhancement was achieved using a two-group object enhancement method.

Plissiti et al. [18] presented a fuzzy C-means algorithm for cervical cell segmentation and clustering. In a paper by Yang et al. [19], the nucleus and cytoplasm were segmented using an edge enhancement contour detector. Kale et al. [20] determined the nucleus optimal segmentation threshold based on the stability of the cell perimeter. Bergmeir et al. [21] developed an algorithm that localizes cell nuclei based on a dynamic voting scheme. Pai et al. [7] developed a method based on maximal grey level gradient difference for accurate segmentation of the nucleus and cytoplasm in pap-smear images. Muhimmah et al. [22] implemented a method based on morphology and watershed transformation and Li et al. [23] extracted nucleus and cytoplasm from pap-smear images using a novel radiating-gradient-vector-flow snake algorithm.

## 1.2 Region Quadtree Segmentation

The quadtree is a data structure concept that refers to a hierarchical collection of maximal blocks that partition a region. It was first introduced by Hanan et.al [24]. The item to be partitioned is the root quadtree which is recursively partitioned according to predefined criteria [24]. Each step of decomposition produces four new quadtrees of the same size that are hierarchically associated with their parent quadtree based on a predefined partitioning criteria. Decomposition finishes whenever there are no more quadtrees to be partitioned or when the quadtrees have reached their minimum size. Quadtree-based approaches can present excellent results when applied in medical image segmentation [25-26]. They are becoming commonly used due to their ability to discard very quickly large amounts of information in an easy and efficient way, while preserving image details [27].

A quadtree is obtained by separating the image into regions based on a given similarity measure, then merge regions based on the same or a different similarity measure. This is achieved through recursive decomposition of the image through the following steps.

1. Determine some similarity measure criteria e.g., mean, standard deviation, variance, texture, etc.
2. Begin with the full image and then split it into 4 partitions (sub-images).
3. Analyse each sub-image.
4. If the sub-image is not uniform, then divide it into 4 new sub-images
5. After each iteration, compare neighbouring regions and the sub-images
6. Then merge if uniform according to the similarity measure.

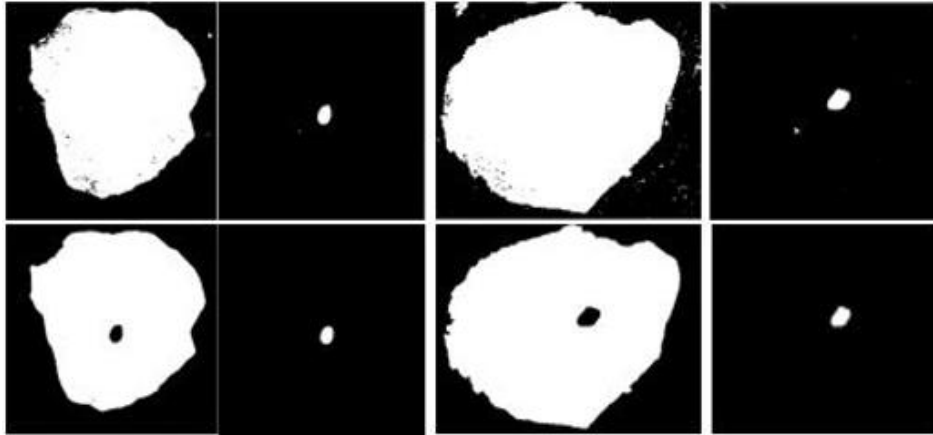
Due to the efficiency of the quadtree approach in image segmentation, researchers are attempting to use it for image segmentation. Gerardo et al. [28] presented a simplified quadtree image segmentation for image annotation. The method was able to efficiently divide the image into homogeneous segments by merging adjacent regions using border and colour information. Spann et al. [29] presented a quad-tree approach to image segmentation by combining a nonparametric classifier, based on a clustering algorithm, with a quad-tree representation of the image. Reza et al. [30] presented a quadtree-based blood vessel detection algorithm using RGB components in fundus images. The technique applied the quadtree on the green component only of the images and the results were promising.

This paper explores the applicability of the quadtree for cervical cell segmentation. It presents a quadtree-based algorithm for nucleus and cytoplasm segmentation of cervical cells from pap-smear images using statistical measures to guide the quadtree decomposition. Segmentation of nucleus and cytoplasm can be used for automated diagnosis and classification of cervical cancer from pap-smear images.

## 2.0 Results

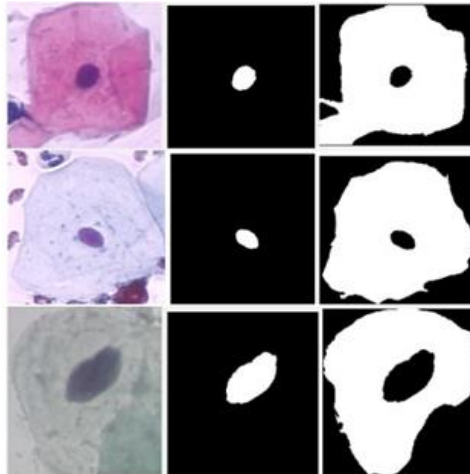
### 2.1 Nucleus and Cytoplasm segmentation

Promising results for nucleus and cytoplasm segmentations were obtained using the proposed quadtree algorithm. Application of morphological operations further improved the segmentation by filling small holes in the nucleus and cytoplasm segmentations and also removing pixels around the boundaries of the nucleus and cytoplasm as shown in Figure 1.



*Figure 1: Nucleus and Cytoplasm segmentations using a Quadtree algorithm and Morphological operations*

To evaluate the accuracy of the nucleus and cytoplasm segmentations (Figure 2), features were extracted from the nucleus and cytoplasm and compared with the ground truth segmentation features that were extracted using CHAMP commercial segmentation software by Jantzen et al. [31].



*Figure 2: Nucleus and Cytoplasm segmentations using a Quadtree algorithm*

The nucleus and cytoplasm features (areas, shortest diameters, longest diameters, roundness and perimeters) were extracted from the segmented cervical cells and compared with the ground truth measurements. The percentage errors in the nucleus area (NA), nucleus shortest diameter (NSD), nucleus longest diameter (NLD), nucleus roundness (NR), nucleus perimeter (NP), nucleus location (NL), cytoplasm area (CA), cytoplasm shortest

diameter (NSD), cytoplasm longest diameter (CLD), cytoplasm roundness (CR) and cytoplasm perimeter (CP) were calculated and shown in Table 1.

Table 1: Percentage errors between the ground-truth and extracted measurements for selected cells.

ID	NA	NSD	NLD	NR	NP	NL	CA	CSD	CLD	CR	CP
1	0.110	0.074	0.075	0.024	0.412	0.070	0.084	0.235	0.002	0.454	0.269
2	0.149	0.152	0.147	0.368	0.050	0.147	0.072	0.486	0.119	0.168	0.394
3	0.046	0.294	0.126	0.321	0.376	0.122	0.005	0.090	0.013	0.298	0.232
4	0.110	0.159	0.377	0.289	0.296	0.432	0.023	0.297	0.206	0.102	0.324
5	0.214	0.348	0.237	0.469	0.131	0.269	0.003	0.362	0.243	0.276	0.378
6	0.035	0.033	0.315	0.079	0.164	0.474	0.042	0.314	0.210	0.034	0.232
7	0.144	0.024	0.195	0.103	0.158	0.247	0.002	0.113	0.368	0.201	0.222
8	0.027	0.298	0.064	0.280	0.018	0.173	0.105	0.344	0.122	0.094	0.175
9	0.212	0.373	0.016	0.295	0.371	0.478	0.112	0.143	0.358	0.220	0.260
10	0.106	0.259	0.105	0.320	0.269	0.154	0.061	0.300	0.250	0.352	0.174
11	0.076	0.266	0.018	0.117	0.270	0.109	0.193	0.138	0.278	0.166	0.264

The average percentage errors of 0.112, 0.207, 0.152, 0.242, 0.229, 0.243, 0.064, 0.256, 0.197, 0.215 and 0.266 were obtained for the NA, NSD, NLD, NR, NP, NL, CA, CSD, CLD, CR and CP respectively. A box plot was obtained to show the shape of the distribution of the percentage error, its central value, and its variability in each extracted feature as shown in Figure 3.

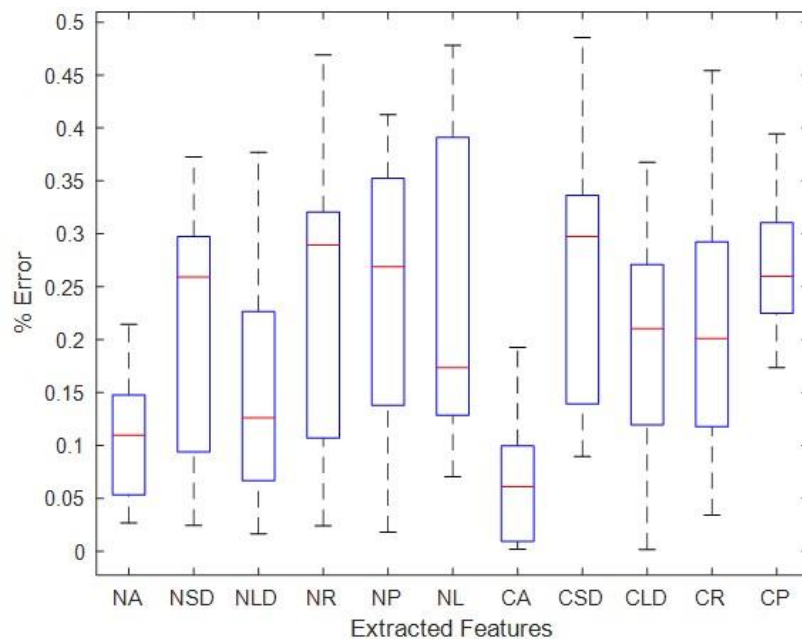


Figure 3: Boxplot for the percentage error in the extracted features.

The variations of the percentage errors in features extracted per image were analysed and are shown in Figure 4.

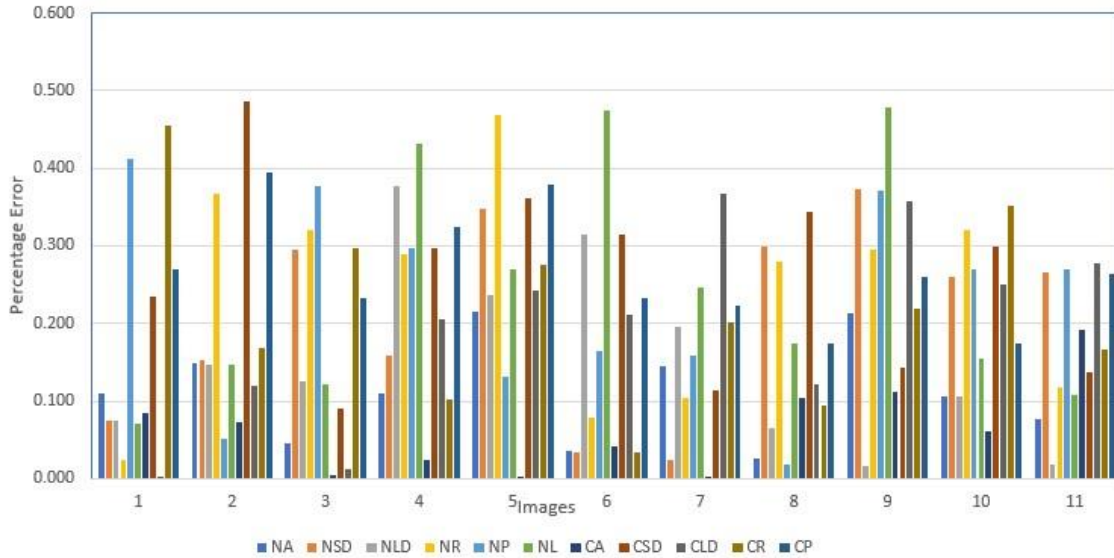


Figure 4: Percentage error in the features extracted per image.

The images in the Herlev dataset belong to 7 classes and were used to test the ability of the Quadtree decomposition algorithm to accurately segment the nucleus regions in each class. As done in [20] and [23], we also use the segment with the highest overlap with the ground truth nucleus region for comparison using the ZSI (Zijdenbos similarity index) which is given by Eq. (1).

$$ZSI = 2 \frac{\#\{X \cap Y\}}{\#\{X\} + \#\{Y\}} \quad (1)$$

where  $X$  and  $Y$  are two sets of segmented pixels. The ZSI computed was compared with that obtained by Asli et al [20] and Kuan et al. [23] who also tested their nuclei segmentation algorithm on the Herlev dataset. The ZSI for the quadtree algorithm has a mean larger than 0.9034 and standard deviation smaller than 0.1735 for all the 7 classes, as shown in Table 2. It can be observed that the quadtree algorithm produces segmentations of acceptable performance compared to the methods in [20] and [23].

Table 2: Comparison of the nucleus segmentation accuracy with methods in [20] and [23]

	cells	Asli et al. [20]	Kuan et al. [23]	<b>Quadtree</b>
Superficial squamous	74	0.93±0.05	0.9524±0.0013	0.9592±0.0184
Intermediate squamous	70	0.95±0.03	0.9578±0.0009	0.9501±0.0726
Columnar	98	0.90±0.07	0.9197±0.0029	<b>0.9034±0.1735</b>
Mild dysplasia	182	0.94±0.08	0.9359±0.0045	0.9239±0.0375
Moderate dysplasia	146	0.93±0.08	0.9334±0.0038	0.9342±0.0826
Severe dysplasia	197	0.92±0.10	0.9333±0.0030	0.9147±0.0443
Carcinoma in situ	150	0.90±0.12	0.9277±0.0039	0.9165±0.1530



The performances of the quadtree algorithm for cytoplasm segmentation was also evaluated using ZSI and compared with results obtained by Shys et.al [19] and Kuan et al [23] who also compared the performance of their cytoplasm segmentation algorithms on Herlev dataset. The statistical results are shown in Table 3.

Table 3: Comparison of the cytoplasm segmentation accuracy with methods in [19] and [23]

Method	$\mu_{ZSI} \pm \sigma_{ZSI}$
Shys et.al [19]	0.8992±0.0348
Kuan et al [23]	0.9545±0.0439
<b>Quadtree</b>	0.9498±0.0921

### 3.0 Discussion

A median filter played a significant role in normalising the cervical cells for the quadtree decomposition algorithm. Unlike in many studies [37-39] where a smaller median filter is used, in this study, a large median filter (27×27) was applied for removing the objects superimposed on the cell background while leaving the very slow variations in the background relatively unchanged. This retrieved the objects of interest while yielding approximately uniform background intensity as shown in Figure.2. Similar to Abramoff et al. [40], application of morphology operations on the initial segmentation helped to overcome inaccuracies in the segmentation by closing small holes in the nucleus and cytoplasm segmentation and closing pixels around the boundaries. This research has shown that the development of an efficient quadtree decomposition criteria has the potential of producing excellent image segmentation as shown in Figure 5 which has also been reported by several studies [29,41-42].

The quadtree decomposition algorithm was able to extract the nucleus and cytoplasm regions accurately and worked well for the different classes of images. The average percentage errors of 0.112, 0.207, 0.152, 0.242, 0.229, 0.243, 0.064, 0.256, 0.197, 0.215 and 0.266 between the segmented images' extracted features (nucleus area, nucleus shortest diameter nucleus longest diameter, nucleus roundness, nucleus perimeter, nucleus location, cytoplasm area, cytoplasm shortest diameter, cytoplasm longest diameter, cytoplasm roundness and cytoplasm perimeter) and the ground truth features show that the measurements are in agreement. From the results shown in Figure 7, it is observed that the algorithm is excellent at nucleus segmentation than the cytoplasm although acceptable results are also obtained for cytoplasm segmentation as compared to other studies. This is also shown by the least variations in the nucleus measurements shown in the boxplot. This could be attributed to the nucleus' brightness compared to other regions. Comparison with other studies showed promising results with the implementation of a quadtree decomposition algorithm for cervical cell segmentation. Given the accuracy of the

classifier in segmenting the nucleus which plays an important role in cervical cancer diagnosis and classification, the classifier can be adapted for automated systems for cervical cancer diagnosis and classification.

#### 4.0 Conclusion

A critical prerequisite in automated analysis of pap-smears is nucleus and cytoplasm segmentation, which is usually considered as the basis of cervical cancer screening from pap-smears. It provides support for various quantitative analyses including calculating cellular morphology, such as size, shape and texture. However, it is difficult to achieve a robust and accurate nucleus/cytoplasm segmentation. This paper articulates a potent approach to the segmentation of cervical cells into nucleus and cytoplasm using a quadtree with statistical measures. Choosing an appropriate quadtree decomposition strategy was a great challenge and a novel task in the proposed approach. As a result, highly efficient and segmentations of acceptable performance were obtained. Comparison of the segmented nucleus and cytoplasm with the ground truth nucleus and cytoplasm segmentations resulted into a Zijdenbos similarity index of greater than 0.9034 and 0.9498 for nucleus and cytoplasm segmentation respectively. The method serves as a basis for first level segmentation of cervical cells for diagnosis and classification of cervical cancer from pap-smear images using nucleus and cytoplasm features.

#### 5.0 Methods

The proposed method was developed through a sequential approach depicted in Figure 5.

##### 5.1.1 Input Images.

The dataset used in the work documented in this paper contains cells obtained from the Herlev University Hospital Cervical Cells Dataset prepared by Jantzen et al. [31]. The dataset contains 947 cervical cells that were obtained by skilled cytotechnicians using a microscope connected to a frame grabber and taken with a resolution of  $0.201\mu\text{m}/\text{pixel}$ . The images were segmented using CHAMP commercial software developed by DIMAC Imaging systems and then classified into 7 classes [31].

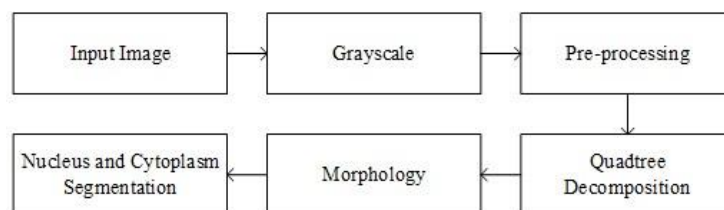


Figure 5: The proposed Approach for segmentation of cervical cell Nuclei from pap-smear images

##### 5.1.2 Grayscale

Since cervical cells are stained in different colours, the input images were first transformed into grayscale. The quadtree algorithm presented in this paper uses pixel level information for decomposition. The grayscale

conversion was carried out to ensure that the value of each pixel was a single sample representing only intensity information each pixel has. This made quadtree decomposition more efficient and reliable. The grayscale conversion was implemented using Eq. (2).

$$\text{New grayscale image} = ((0.3 * R) + (0.59 * G) + (0.11 * B)) \quad (2)$$

Where R=Red, G=Green and B=Blue colour contributions of the new image.

### 5.1.3 Pre-processing.

A pap-smear is stained for easy identification of cell nuclei [32]. The staining usually delineates the nuclei pretty well, however, the staining is not homogenous, as areas of condensation levels can vary across the chromosomes and uneven lighting across the field of view can make the nuclei appear granular [33]. In order to remove the noise, denoising was carried out on the original images using a median filter. In general, the median filtering output is given by Eq. (3) as used in [22].

$$g(x, y) = \text{med}\{f(x - i, y - j), i, j \in w\} \quad (3)$$

where  $f(x, y)$  and  $g(x, y)$  are the original and the output image respectively.  $w$  is a two-dimensional mask of size  $n \times n$ . In order to improve the efficiency of the medium filter, an efficient median filter algorithm proposed by Manfred et al. [34] was implemented. The algorithm computes two neighbouring medians in one step using spatial coherence between the pixels. A large median filter ( $27 \times 27$ ) was applied for removing the objects superimposed on the cell background while leaving the very slow variations in the background relatively unchanged. To yield uniform background within, the difference between the original and the median filtered image was computed. This operation served to retrieve the objects of interest while yielding approximately uniform background intensity. The resultant filtered image is shown in Figure 6.

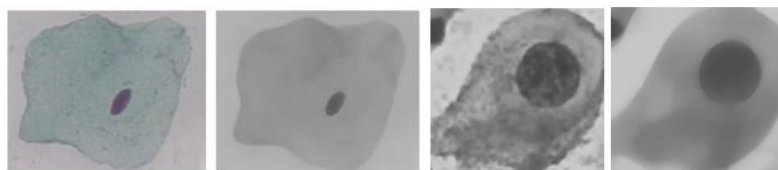


Figure 6: Randomly selected original and median filtered cervical cell.

### 5.1.4 Quadtree Decomposition

A quadtree scanning of the full image is the core step of the proposed segmentation technique. However, the efficient median filter enhanced the decomposition process by reducing noise in the image. We adopt a split and merge quadtree decomposition strategy where the image is divided into four regions, and each of these regions is compared with their adjacent 4-neighbours using a comparison operator. The image is recursively subdivided until smallest unit is reached and in this case it's the pixel information. If two regions are evaluated as similar they

are merged. Regions that are not merged with any other region are divided into four new regions and the same comparison operation with their new neighbours is done. This process is performed recursively until there are no more regions to divide as shown in Figure 7.

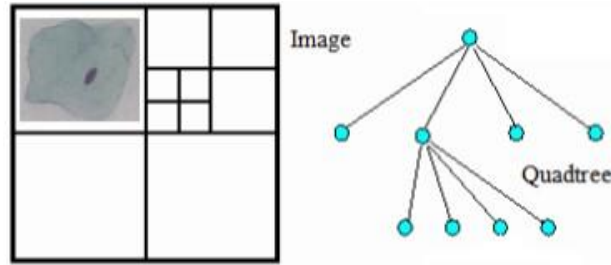


Figure 7: Quadtree decomposition

Recent quadtree decomposition algorithms have been constructed by building a quadtree downwards from each image pixel and storing only the mean value of the subtree pixels in each node. However, mean alone gives no measure of uniformity which is essential in cervical cell segmentation. In the proposed method, the variance and maximum entropy statistical measures of the pixels at each node are stored during the decomposition process to overcome limitations of using only the mean. Hence at each node in the quadtree the algorithm stores the mean, maximum entropy and the variance of the pixels' gray level in the subtree. A homogeneity threshold is automatically determined by the algorithm based on the variance maps of the neighbouring pixels. This is achieved through the following steps:

- a) Construct a quadtree storing at each node the mean, maximum entropy and the variance of the pixel's gray levels of its four children.
- b) Search from the root of the tree to find the node highest in the tree whose mean, maximum entropy and the variance are in the required range (as for the root node), and
- c) Use neighbour finding techniques to merge adjacent nodes whose mean is within the desired range.

The mean ( $\mu$ ) of  $n$ -pixels in a subtree at a node,  $I(x, y)$  is represented by Eq. (4).

$$\mu = \sum i(x, y) / n \quad (4)$$

From equation 3, the variance can be defined by Eq. (5).

$$\alpha = \sum \frac{i(x, y)^2}{n} - \mu^2 \quad (5)$$

Image entropy is a statistical measure of randomness that can be used to characterize the texture of an image. Shannon's entropy [35] defined by Eq. (6) was used in this paper.

$$H(s_m) = -\sum_{n=1}^{256} p_n(s_m) \cdot \log_2(p_n(s_m)), m = 1, \dots, M \quad (6)$$

where  $H(s_m)$  is the entropy of the pixel intensity. The  $(p_n(s_m))$  is the probability of a pixel having entropy  $s_m$ , and  $m$  are all the possible entropy values. The probability density  $p_n$  is calculated using the grey level histogram.

After computation of the above statistical measures, the cervical cell images were segmented using the quadtree decomposition method searching for the node within the required mean, maximum entropy and variance range. To save search time, the highest uniform node in each subtree was recorded as the tree was being built. This eliminated the need for a search after the tree was built. The pixel segmentation was guided by the uniformity of the pixels in each node. The uniformity was recorded by adding a pointer  $\beta$  to each node in the tree which is a membership function for a fuzzy set (statistical measures) for that pixel. The pointer  $\beta$  was computed as follows:

1. If a node (N) and all nodes in its subtree (SN) were out of range of the statistical measures (mean, maximum entropy and variance), then its pointer was set to null as shown in Eq. (7).

$$\text{If } (N \text{ AND } SN) \notin (\text{statistical measures}) \text{ then } \beta = \text{Null} \quad (7)$$

2. If a node was in range, then its pointer was set to point to itself as shown in Eq. (8).

$$\text{If } (N) \in (\text{statistical measures}) \text{ then } \beta = N \quad (8)$$

3. If a node was out of range, but some nodes in its subtree were in range, then its pointer was set to point to the highest node in its subtree which was within range as shown in Eq. (9).

$$\text{If } (N) \notin (\text{statistical measures}) \text{ AND } (SN) \in (\text{statistical measures}) \text{ then } \beta = SN_{max} \quad (9)$$

The pointers helped in the segmentation of the cervical cell into the nucleus, cytoplasm and background based on pixel level information. Taking the node pointed to by the root, we used it as a seed and merged its neighbours and this was used for nucleus segmentation, setting the pointers of each merged nodes to null. We then extracted further regions by moving down the tree and processing similarly any nodes with non-null pointers which were then used for cytoplasm segmentation. To overcome some inaccuracies in the segmentations obtained, morphology operations were applied to these images.

### 5.1.5 Morphology

Morphology is a set-theory approach that considers an image as the elements of a set and process images as geometrical shapes [36]. It is a powerful technique for solving a number of problems in image analysis and computer vision including overcoming inaccuracies in segmentation. The two basic morphological operators are the erosion and the dilation based on Minkowski algebra [39]. A dilation followed by erosion is called a closing operation. Dilation operation of a grayscale image by a two-dimensional point A is defined by Eq. (10) as;

$$(g \oplus A)(r, c) = \max\{g(r - k, c - l) | (k, l) \in A\} \quad (10)$$

Erosion of the grayscale image by a two-dimensional point A is defined by Eq. (11),

$$(g \ominus A)(r, c) = \min\{g(r + k, c + l) | (k, l) \in A\} \quad (11)$$

where  $g$  is the grayscale image,  $A$  is a structuring element [40] and  $(r, c)$  is the pixel of the image  $g$ ,  $(k, l)$  is the size of the element  $A$ . In this paper, the closing operation was utilised to overcome the inaccuracies in the quadtree segmentation by fusing narrow breaks around the nucleus and cytoplasm. It was also used to fill small holes and gaps in the image. Closing is mathematically defined by Eq. (12).

$$g \bullet A = \left[ g \oplus A \right] \ominus A \quad (12)$$

where  $\ominus$  is an erosion,  $\oplus$  is dilation,  $g$  is a binary image and  $A$  is structuring element. After the morphological operation, image shape features, such as blurred edges, holes, corners, wedges and cracks were removed.

### List of Abbreviations

<b>WEKA:</b>	Waikato Environment for Knowledge Analysis
<b>TWS:</b>	Trainable Weka Segmentation
<b>FRF:</b>	Fast Random Forest
<b>FCM:</b>	Fuzzy C-Means

### Declarations

#### Ethics approval and consent to participate

Mbarara University of Science and Technology Research Ethics Committee (MUREC) approved this research (Protocol Number: No.03/06-17)

#### Consent for publication

Not applicable

#### Availability of data and material

The dataset used in the work documented in this paper contains cells obtained from the Herlev University Hospital Cervical Cells Dataset which are freely available for public research use.

#### Competing interests

This paper has the assent of all co-authors and the authors declare that there are no conflicts of interest. The authors declare that there is no conflict of interests regarding the publication of this paper.

#### Funding

African Development Bank- HEST project provided funds for this research. The Commonwealth Scholarships Commission also provided funds for this research while in the UK.

#### Authors' contributions

WW wrote the paper, AW, AHB and JO provided both technical and scientific writing support to this manuscript.

#### Acknowledgements

The authors are gratefulness to the African Development Bank- HEST project for providing funds for this research and the Commonwealth Scholarship Commission for the split-site scholarship for the first author at the University of Strathclyde. The support and exposure from the UK greatly enhanced this research. The authors are also grateful to Mr Abraham Birungi from Pathology department of Mbarara University of Science and Technology, Uganda for providing support with pap-images.

## □ Authors' information

---



[Mr Wasswa William](#) is a senior lecturer at Mbarara University of Science and Technology, Uganda. He has a master's in Biomedical Engineering from the University of Cape Town, South Africa. He has valuable experience in the fields of Medical Devices, Medical Imaging and Machine Learning.



[Professor Andrew Ware](#) is a professor in Computing at the University of South Wales. He has a lead on collaborative projects both within the UK and Europe. He has lectured in many parts of the World including the USA, Canada, Singapore and Hong Kong. His main research interest is the application of AI techniques to help solve real-world problems.



[Assoc. Prof Annabella Habinka Basaza](#) is an Associate professor at the College of Computing and Engineering and St. Augustine International University and in the Department of Computer Science at Mbarara University of Science and Technology. Her main research interests include ICT and Health, ICT for Education, ICT for food security, Data mining and Machine Learning.



[Dr Johnes Obungoloch](#) is a Biomedical Engineer with a PhD (Biomedical Engineering) from Pennsylvania State University, UK. His research interests include medical devices innovations, design and development.

---

## References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA a cancer J Clin* 2015;65(2):87–108.
2. ICO Information Centre on HPV and Cancer (HPV Information Centre) Ecuador 2016. Human Papillomavirus and Related Diseases Report Ecuador. *HPV Inf Cent* 2016;
3. Bengtsson E, Malm P. Screening for cervical cancer using automated analysis of PAP-smears. *Comput Math Methods Med* 2014;842037
4. Chen MK, Hung HF, Duffy S, Yen AMF, Chen HH. Cost-effectiveness analysis for Pap smear screening and human papillomavirus DNA testing and vaccination. *J Eval Clin Pract* 2011; 17(6):1050-1058
5. Lu Z, Carneiro G, Bradley AP. Automated nucleus and cytoplasm segmentation of overlapping cervical cells. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2013;8149 LNCS(PART 1):452–60.
6. Grubb DT. Optical Microscopy. In: *Polymer Science: A Comprehensive Reference*, 10 Volume Set. 2012.
7. Pai PY, Chang CC, Chan YK. Nucleus and cytoplasm contour detector from a cervical smear image. *Expert Syst Appl* 2012;39(1):154–61.
8. Pratt WK. Image Segmentation. *Comput Vis.* 2000; 19:153–164.
9. Zaitoun NM, Aqel MJ. Survey on Image Segmentation Techniques. In: *Procedia Computer Science* 2015; 65:797–806.
10. Sahoo PK, Soltani S, Wong AKC. A survey of thresholding techniques. Vol. 41, *Computer Vision, Graphics and Image Processing* 1988;41:233–60.
11. Meijering E. Cell segmentation: 50 Years down the road. *IEEE Signal Process Mag* 2012; 29: 140–145
12. William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Comput Methods Programs Biomed* 2018;164:15–22.
13. Al-amri SS, Kalyankar N V, Khamitkar SD. Image Segmentation by Using Threshold Techniques. *J Comput* 2010;2(5):83–6.
14. Yogarajah P, Condell J, Curran K, Cheddad A, McKeivitt P. A dynamic threshold approach for skin segmentation in color images. In: *Proceedings - International Conference on Image Processing, ICIP*

- 2010; 17: 2225–8.
15. Zhang L, Chen S, Wang T, Chen Y, Liu S, Li M. A Practical Segmentation Method for Automated Screening of Cervical Cytology. 2011 Int Conf Intell Comput Bio-Medical Instrum. 2011;140–3.
  16. Plissiti ME, Nikou C, Charchanti A. Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images. Pattern Recognit Lett 2011;32(6):838–53.
  17. Lin CH, Chan YK, Chen CC. Detection and segmentation of cervical cell cytoplasm and nucleus. Int J Imaging Syst Technol 2009;19(3):260–70.
  18. Plissiti ME, Tripoliti EE, Charchanti A, Krikoni O, Fotiadis DI. Automated detection of cell nuclei in pap stained cervical smear images using fuzzy clustering. In: IFMBE Proceedings. 2008;637–41.
  19. Yang-Mao SF, Chan YK, Chu YP. Edge enhancement nucleus and cytoplasm contour detector of cervical smear images. IEEE Trans Syst Man, Cybern Part B Cybern. 2008;38(2):353–66.
  20. Kale A, Aksoy S. Segmentation of cervical cell images. In: Proceedings - International Conference on Pattern Recognition 2010: 2399–402.
  21. Bergmeir C, García Silvente M, Benítez JM. Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a web-based software framework. Comput Methods Programs Biomed 2012;107(3):497–512.
  22. Muhimmah I, Kurniawan R, Indrayanti. Automated cervical cell nuclei segmentation using morphological operation and watershed transformation. In: Proceeding - 2012 IEEE International Conference on Computational Intelligence and Cybernetics, CyberneticsCom 2012;12:163–7.
  23. Li K, Lu Z, Liu W, Yin J. Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake. Pattern Recognit 2012;45(4):1255–64.
  24. Samet H. The Quadtree and Related Hierarchical Data Structures. ACM Comput Surv 1984; 16: 187-260
  25. Pham DL, Xu C, Prince JL. Current Methods in Medical Image Segmentation. Annu Rev Biomed Eng 2000;2(1):315–37.
  26. Kwok CK, Khan GN, Gillies DF. Automated Endoscope Navigation and Advisory System from medical imaging. SPIE's Int Conf Physiol Funct fro Multidimens Images 1999; 3660:81-88
  27. Jayaraman PK, Mei J, Cai J, Zheng J. Quadtree convolutional neural networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2018; 6273. 116 - 123
  28. Conde Márquez GR, Escalante HJ, Sucar LE. Simplified quadtree image segmentation for image annotation. In: CEUR Workshop Proceedings. 2011.
  29. Spann M, Wilson R. A quad-tree approach to image segmentation which combines statistical and spatial information. Pattern Recognit 1985; 3: 257-269
  30. Reza AW, Eswaran C, Hati S. Diabetic retinopathy: A quadtree based blood vessel detection algorithm using RGB components in fundus images. J Med Syst 2008; 32:147–155
  31. Jantzen J, Norup J, Dounias G, Bjerregaard B. Pap-smear Benchmark Data For Pattern Classification. Proc NiSIS 2005 Nat inspired Smart Inf Syst. 2005;1–9.
  32. Bengtsson E, Malm P, Bengtsson E, Malm P. Screening for cervical cancer using automated analysis of PAP-smears. Comput Math Methods Med 2014;2014:842037.
  33. Liu XY, Wang WH, Sun Y. Dynamic evaluation of autofocusing for automated microscopic analysis of blood smear and pap smear. J Microsc. 2007;227(1):15–23.



34. Kopp M, Purgathofer W. Efficient 3 times 3 Median Filter Computations. *MVG* 1995; 41(1):74-8
35. Vedral V. The role of relative entropy in quantum information theory. *Reviews of Modern Physics* 2002; 74, 197-234
36. Sonka M, Hlavac V, Boyle R. Mathematical morphology. In: *Image Processing, Analysis and Machine Vision* 1993; 30 (1993):521-539.
37. Perreault S, Hébert P. Median filtering in constant time. *IEEE Trans Image Process* 2007; 16(9):2389-94
38. Hwang H, Haddad RA. Adaptive Median Filters: New Algorithms and Results. *IEEE Trans Image Process* 1995; 4(4):499-502
39. Narendra PM. A Separable Median Filter for Image Noise Smoothing. *IEEE Trans Pattern Anal Mach Intell* 1981; 3(1):20-9
40. Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering* 2010; 3:169-208
41. Smith JR, Chang SF. Quad-tree segmentation for texture-based image query. *Proc Second ACM Int Conf Multimed* 1994; (ACM-GIS 94), (pp. 1-7)
42. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng* 2010; 57, 841–852

### Figure Captions

- Figure. 1: Nucleus and Cytoplasm segmentations using a Quadtree algorithm and Morphological operations
- Figure 2: Nucleus and Cytoplasm segmentations using a Quadtree algorithm
- Figure 3: Quadtree decomposition
- Table 1: Percentage errors between the ground-truth and extracted measurements for selected cells.
- Figure 3: Boxplot for the percentage error in the extracted features
- Figure 4: Percentage error in the features extracted per image
- Table 2: Comparison of the nucleus segmentation accuracy with methods in [20] and [23]
- Table 3: Comparison of the cytoplasm segmentation accuracy with methods in [19] and [23]
- Figure 5: The proposed Approach for segmentation of cervical cell Nuclei from pap-smear images
- Figure 6: Randomly selected original and median filtered cervical cell.
- Figure 7: Quadtree decomposition