Cervical Cancer Segmentation using Fuzzy Support Vector Machine Algorithm

Gowri Sankar P.A.[1], Aparna R.[2], Sathishsharama K.[3], Aravinth K.[4], Thirumalai V.[5], Sathishkumar M.[6]

[1]Associate Professor, Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, Tamil Nadu

[2]PG Student, Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, Tamil Nadu

[3,4,5,6] UG Student, Department of Electrical and Electronics Engineering, Knowledge Institute of Technology, Salem, Tamil Nadu

E-mail: [1]pagseee@kiot.ac.in [2]aparnaravi129@gmail.com

Abstract

Cervical cancer is a dangerous disease, particularly prevalent in developing countries where public awareness is low. The Papanicolaou test, commonly known as the Pap test, is the most widely used method to detect cervical cancer, which develops in the cervix and affects many women. Image processing algorithms play an important role in the segmentation of the cancerous region in cervical images. The fuzzy support vector machine (FSVM) algorithm is used to segment the cancerous regions in cervical cancer images. This method effectively separates the cervical cancer regions from the background in these images. The K-means classification algorithm is another existing method applied to cervical cancer images. The results of the existing and proposed segmentation algorithms are compared using quality measurement techniques such as accuracy and precision. The proposed FSVM algorithm demonstrated the highest accuracy (98%) compared to the previous algorithms.

Keywords: Cervical Cancer, Image Segmentation, FSVM, Image Quality, k means.

DOI: https://doi.org/10.36548/jscp.2024.2.007
Received: 03.05.2024, received in revised form: 29.05.2024, accepted: 14.06.2024, published: 25.06.2024

1. Introduction

Cervical cancer and precancerous cells that may eventually become malignant can both be found using screening tests. Cells that begin to proliferate in the cervix might develop into cervical cancer. The lowest portion of the uterus that joins the vagina is called the cervix. The virus known as human papillomavirus (HPV) is responsible for most cervical malignancies and comes in different strains. If cervical cancer is discovered earlier and addressed quickly, it can be cured. In the decades to come, nations worldwide are striving to expedite the eradication of cervical cancer, including a set of three goals that must be accomplished by 2030. If detected and treated in its early stages, cervical cancer is curable.

A crucial first step is identifying symptoms and getting medical help to address any concerns. If a woman notices, the symptoms she should consult a healthcare expert. The symptoms of cervical cancer is given below,

- Abnormal vaginal discharge.
- Chronic back, leg, or persistent of pelvic pain.
- A typical bleeding between periods after menopause.
- They get weight loss, exhaustion, and appetite loss.
- Discomfort during vaginal swells legs.

In this research, we discuss the image acquisition, image pre-processing, and segmentation of cancer regions using fuzzy-based support vector machines and quality measurement for the segmentation algorithm. The flow of proposed work is given in the below Figure 1.

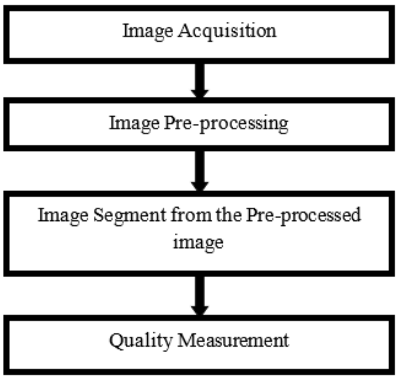


Figure 1. The Flow of Proposed Work Methodology

The Computer-aided diagnosis (CAD) tools are developed to detect cancer in the cervix region, where portable technology is needed to identify aberrant cervix alterations that may lead to invasive carcinoma at a low cost. The developed tools are assisting with secondary examinations for expert assessments for early detection. The proposed work consists of picture collection to obtain cervical cancer data for examination of the cancer region. Then Pre-processing is one of the most efficient ways to improve the cancer digital image. The goal of pre-processing is to enhance the picture data by suppressing unwanted distortions or enhancing certain visual elements that are relevant for later processing. The proposed methods of fuzzy-based support vector machine (FSVN) and k-means classification were used to segment the cancer region from the image background. This algorithm quality was tested in terms of precision and accuracy.

2. Literature Review

One of the primary causes of cancer-related deaths in women globally is cervical cancer, which behaves epidemiologically like a low-infectious venereal illness [1, 2]. Human papillomavirus (HPV) 16/18 and smoking are currently significant factors in the multifaceted, progressive carcinogenic theory in the cervix uteri. As a result, it is advised to implement screening programmes, HPV vaccination campaigns, and society-based preventive and control measures [1,2].

The small portion of the uterus that links to the vagina is called the cervix, and here is where cervical cancer starts. The image segmentation is used extract the region of cancer so medical image segmentation is used for together therapeutic and diagnosis reason. The region of interest, classification, threshold methods are applied in a medical imaging were identified with the use of image segmentation. This phase is essential for further analysis because it enables the emphasis on a particular region that is pertinent to the screening of cervical cancer. Cervical abnormalities or lesions are common presentations of cervical cancer. These lesions can be located and identified via image segmentation[3].

Healthcare practitioners can identify regions that could need more investigation with the use of automated segmentation techniques. In the event that an anomaly or tumour is present, segmentation aids in defining its limits. Precise segmentation facilitates comprehension of the tumour form and size, which is important for treatment planning. Segmentation aids in defining the limits of any tumours or abnormalities that may be present.

For the purpose of treatment planning, accurate segmentation helps determine the size and form of the tumour.

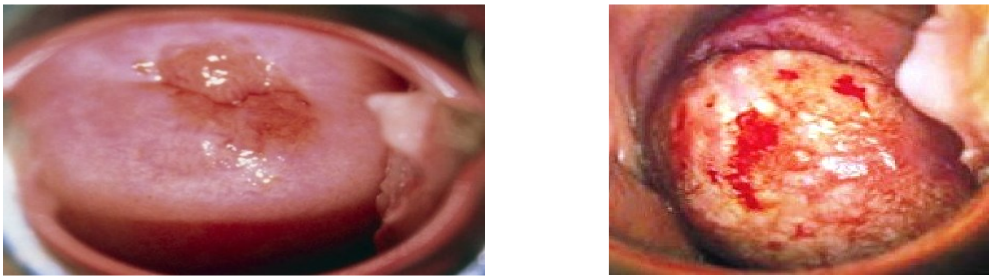
Multi class support vector machine (multiclass-SVM) was applied for cervical images where detect and segment the abnormal region as object from the background region [13]. Sometimes, the image features are extracted from the cancer images where SVM classification is applied for segment the cancer regions. Before used of SVM, select the kernel type leads to better classification. There are important standard kernel types such as polynomial kernel, radial basis kernel, gaussian and linear kernel [10]. The research is focused on cancer detection utilising image processing algorithms to find the cancer nucleus region. Most machine learning approaches are used to identify cancer such as SVM, CNN architect (Convolutional Neural Network) and it has been one of the autonomous processes that predict cancer at an early stage, improving patient safety [7]. The KNN (K-Nearest Neighbor) algorithm is provided classification precision of eighty-four percentage for segment the cancer cervical images [8]. The standard k means algorithm is applied for colposcopy images with accuracy 87.25% [9]. This literature study reveals that many common algorithms are utilised to segment cancer regions in cervical images, and this research raises awareness about the need for experimental work in medical-based cancer research.

3. Methodology

Follow-up scans and other serial imaging techniques can be used to track the advancement or regression of malignancy. Observing changes in tumour features and size over time is crucial for assessing how well a treatment is working. Image segmentation helps to segment the cancer region from the cervical cancer images.

3.1 Image Acquisition

The National Cancer Institute (NCI) generated the Guanacaste dataset (https://knoema.com/atlas/Costa-Rica/Guanacaste/datasets) in 1997, and this collection of cervical photos includes photographs of patients with cancer and those without any abnormal abnormalities. It contains ground-truth images that have been examined by the experts [4]. The Figure 2 illustrates the cervical cancer image1 and image2.



a) Cervical Cancer Image b) Cervical Cancer Image(at stage 2nd)

Figure 2. Cervical Cancer Images

3.2 Image Pre-Processing

Computers process digital images and characterise them numerically. Digital image processing is used to process digital images on computers. The digital image processing algorithms help enhance the image, or if the image contains errors, they reduce noise and enhance the brightness or luminance. In this paper, the RGB (red, green, and blue) image is converted to hue, saturation, and value.

3.3 Image Segmentation

The SVM is used for segmentation tasks, including cancer image analysis and then medical imaging. SVM is one of the supervised machine learning algorithms that is applied for both segmentation and classification tasks. The specific region object is extracted from the background image and assigned to a specific category or based on classes. The Pap smear test examines the cells in the cervix area and diagnoses the nuclei sections. The image processing technique uses cervical images to identify cancer in the cervix. In the current environment, cervical cancer can be detected at the terminal phase, resulting in untimely death. In numerous cases, people do not experience any discomfort or soreness in the cervix region, and no symptoms develop. If the cancer is diagnosed early, the patient's life can be preserved [4, 5]. Researchers and scientists spent more time detecting cancer in its early stages and improved the automatic tools for segmentation and prediction.

4. Modules

4.1. Standard SVM

Cervical cancer is becoming more common every day. Segmenting the nucleus and cytoplasm and detecting the cancer cells is a difficult task in cervical cancer. A novel radiating gradient vector flow (RGVF) technique that seeks to recover the cytoplasm and nucleus from a single cervical smear cell was used.

After extracting the cervical picture, the k means algorithm is used for pre-processing to roughly locate the nucleus and cytoplasm. Utilizing a novel edge map computational stack-based refinement approach, the RGVF algorithm effectively locates obscure boundaries in the cytoplasm and nucleus while disregarding false negative rates. The sample used to extract the uneven borders between the cytoplasm and nucleus of a single cervical cell is the 917 images in the Herlev dataset [1].

4.2. Fuzzy Support Vector Machine Algorithm

The diagnosing of cancer cells in the cervical cancer is a tedious job as it has to be diagnosed in the precancerous stage. The challenging task is to segment a single cell image into nucleus, cytoplasm and background. Three different types of datasets are collected and extracted to differentiate the nucleus, cytoplasm and background regions. While every data point in SVM is treated equally, for many real-world classification issues, some data points may be more crucial than others. To address this issue, Lin and Wang [12] created the fuzzy support vector machine (FSVM), a fuzzy variant of SVM. Based on each data point's significance, the fuzzy membership function in FSVM assigns a fuzzy membership value. To create the fuzzy membership, we'll employ the class center approach.

4.2.1. Design Fuzzy Membership Function

The support vectors determine the best classification hyperplane, according to the fundamental idea of the support vector machine. These support vectors are located far from the center of the two classes on the respective boundaries of the two convex sets, if each class in the two classification problems is thought of as a convex set. The membership functions of both positive and negative sample points are constructed according to equation (1) and (2).

$$s_i^+ = \begin{cases} 0.6 * \frac{D(\varphi(x_i^+))}{d(\varphi(x_i^+))} * \frac{d(\varphi(x_i^+))}{R^+} + 0.4, & d(\varphi(x_i^+)) \le R^+,\ i = 1,2,\cdots,l_{new}^+ \\ 0.4 * \left[\frac{1}{\frac{d(\varphi(x_i^+))}{R^+} + \frac{D(\varphi(x_i^+))}{d(\varphi(x_i^+))}} \right]^P, & d(\varphi(x_i^+)) > R^-,\ i = 1,2,\cdots,l_{new}^+ \end{cases} \tag{1}$$

$$s_i^- = \begin{cases} 0.6 * \frac{D(\varphi(x_i^-))}{d(\varphi(x_i^-))} * \frac{d(\varphi(x_i^-))}{R^-} + 0.4, & d(\varphi(x_i^-)) \le R^-,\ i = 1,2,\cdots,l_{new}^- \\ 0.4 * \left[\frac{1}{\frac{d(\varphi(x_i^-))}{R^-} + \frac{D(\varphi(x_i^-))}{d(\varphi(x_i^-))}} \right]^P, & d(\varphi(x_i^-)) > R^-,\ i = 1,2,\cdots,l_{new}^- \end{cases} \tag{2}$$

4.2.2. Method for Feature Extraction

Feature extraction is a crucial step in machine learning, particularly for algorithms like the Fuzzy Support Vector Machine (FSVM). FSVM is an extension of the standard SVM that incorporates fuzzy logic to handle uncertainty and noise in data. When using Fuzzy Support Vector Machines (FSVM) for image processing tasks, feature extraction plays a pivotal role in transforming raw image data into a format that the FSVM can effectively utilize. Here are some common feature extraction methods used in image processing, specifically tailored for training FSVMs is Principal componenet analysis methods.

4.2.2.1.Principal Component Analysis (PCA)

PCA is a statistical technique used to reduce the dimensionality of the data while retaining most of the variance. This is achieved by transforming the original features into a new set of uncorrelated features called principal components.

Step 1: Flatten the image into a 1D array.

Step 2: Compute the covariance matrix of the dataset.

Step 3: Calculate the eigenvectors and eigenvalues.

Step 4: Project the data onto the principal components to form the feature vector.

4.2.3. Method for Feature Selection and Training

After extracting features, it's often beneficial to perform feature selection to retain the most informative features and reduce dimensionality. In this work, Recursive Feature Elimination (RFE) Techniques is used for feature selection and training.

4.2.4. Implementation of FSVM in Software Tools

Implementing a Fuzzy Support Vector Machine (FSVM) algorithm requires a combination of data pre-processing, feature extraction, fuzzy membership assignment, and the application of the SVM algorithm with a fuzziness factor. FSVM is implemented using Python with popular libraries such as NumPy, scikit-learn, and potentially some fuzzy logic libraries.

The software/tool implementation workflow is described in the below steps for image segmentation:

Step 1: Pre-process the Image: Normalize and resize the image as needed.

Step 2: Extract Features: Use one or more feature extraction methods described above.

Step 3: Normalize Features: Normalize the feature vectors to have zero mean and unit variance.

Step 4: Assign Fuzzy Membership Values: Determine fuzzy membership values based on the extracted features.

Step 5: Train FSVM: Use the extracted features and fuzzy membership values to train the FSVM.

Selecting and combining feature extraction methods can enhance the performance of FSVMs in various image-processing tasks such as classification, detection, and segmentation.

5. Results and Discussion

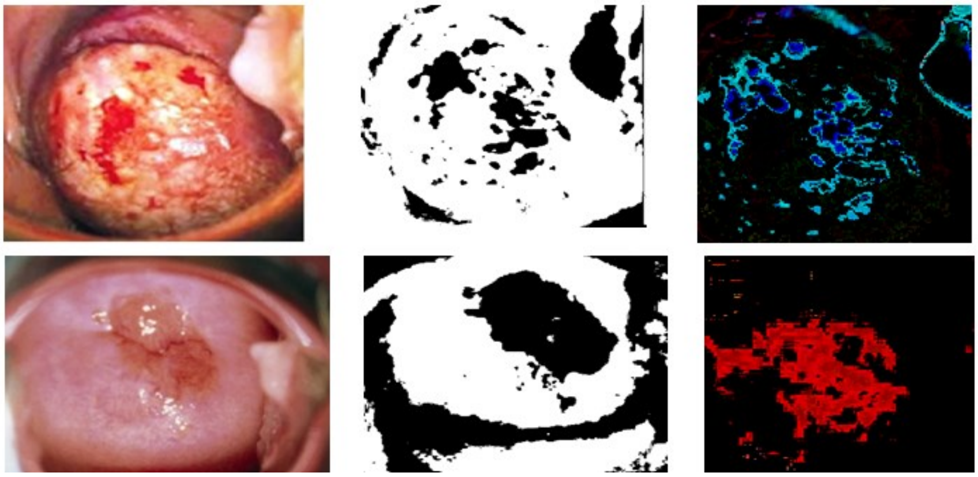
A collection of specified pixels is all that makes up an image. The similar pixels are clustered together with meaningful information from the original image. The pixel-by-pixel mask was created for each item in the image, and the object was divided from the image at a deeper and more detailed level. Image segmentation is one of the important tasks to detect the cancer-affected region from the cancer cervical images. We discussed the applied segmentation algorithm and the comparison of quality based on measurement. The k means segmentation is one of the standard methods: randomly assign the data points and then calculate the cluster centre point. Then calculate the Euclidean distance between the centre points. This process continuously produces data points that cannot be changed. It is a very complex and time-consuming process to handle cancer image data.

To begin with, the original image is converted to an HSV image, and the resultant image is adjured for a segmented process. The SVM has the capability to segment the images, and the intelligence approaches are an SVM and neural network. SVM is more sensitive and one of the standard methods, besides. The SVM method has one complication, which is sensitivity to pixel initialization, which can be overcome by applying fuzzy-based SVM [12]. From the proposed algorithm, we measured the hue, saturation, and values, extracted them, and applied fuzzy and SVM for cancer in the cervical region. The fuzzy-based SVM yields better segmentation results compared to the existing K-means clustering algorithm. The algorithm steps are given below:

- To set the initial parameter and then the number of clusters.
- The number of colour pixels is clustered through the membership function.
- Classifying the pixels in the image by membership value.
- To create the whole training set, combine the training samples from each cluster. As the test set, retain the remaining image pixels.
- Then train the SVM classifier (with a train set with FCM class labels by SVM)

5.1 Experimental Results for Cancer Region Detection using K-Means Classification and FSVM.

The proposed and existing method applied for cervical cancer images to extract the cancer region shown in Figure 3. Compared the existing and proposed method, the cancer region is better segmented from fuzzy based SVM.



Original Image k-means Segmentation FSVM Segmentation Image

Figure 3. Segmentation Results for K Means and then FSVM Algorithm

5.2 Quality Measurements between Existing and Proposed Algorithm

The result of the suggested cervical cancer segmentation system is assessed in terms of precision and accuracy, as shown below Table 1 and Table 2.

Table 1. Quality Measurement Value for Cervical Cancer Image 1

Algorithms	Precision	Accuracy	
		This Paper	Ref [9]
K-means	87%	96.2%	87.25%
FSVM	89%	98.6%	88%

Table 2. Quality Measurement Value for Cervical Cancer Image 2

Algorithms	Precision	Accuracy	
		This Paper	Ref [9]
K-means	89%	96.2%	87.25%
FSVM	97%	98.6%	88%

When compared to the current k-means segmentation approach, the FSVM-based cancer area segmentation method performs well based on quality measurements. When applied k-means for cervical cancer image 1, the precision value is 87%, followed by accuracy of 96.2%. As a result, the suggested FSVM segmented measurement value of accuracy is 98.6%, while the precision value is 89%. When k-means is used for cervical cancer image 2, accuracy is 94.4% and precision is 89%. As a result, the suggested FSVM segmented measurement value of accuracy is 98%, and precision is 97%.

6. Conclusion

Cervical cancer is the second most common malignancy among women worldwide. This cancer originates in the cervix, which is located at the lower part of the uterus. Since cervical cancer patients typically exhibit few symptoms, early diagnosis is crucial to detecting the disease. In these cases, automated detection employing image processing algorithms is used to identify the malignancy. In this research, we select the cervical cancer images and detect the cancer region by using image processing algorithms such as existing method of k-means and then proposed FSVM. The FSVM method is performed well compared to the existing one. Based on quality assessments, the FSVM-based cancer area segmentation method outperforms the existing k-means segmentation strategy. Using k-means to cervical cancer image 1, the accuracy is 96.2% and the precision is 87%. Consequently, the accuracy value of the proposed FSVM segmented measurement is 98.6%, while the precision value is 89%. For Image 2 of cervical cancer, k-means yields accuracy of 94.4% and precision of 89%. This leads to a proposed FSVM segmented measurement value of 98% for accuracy and 97% for precision. Early detection of cancer improves the patients' outcomes.

Reference

[1] Zhang Shaokai, Xu Huifang, Zhang Lugao, Qiao Youllin, "Cervical cancer: Epidemiology, risk factors and screening", Chinese Journal of Cancer Research Dec Vol.31, Issue 32,pp.720-728, 2020

[2] Zur Hausen H, "Papillomaviruses in anogenital cancer as a model to understand the role of viruses in human cancers", American Journal of Cancer Research (AACR); Volume 49, Issue 46, pp. 77–81, 1989

[3] Mattiuzzi Camilla, Lippi Giueppe, “Cancer statistics: a comparison between World Health Organization (WHO) and Global Burden of Disease (GBD)”, European journal of public health. 2020;30:1026–7.

[4] Jaya, B Karthiga, and S Senthil Kumar. “Image Registration based Cervical Cancer Detection and Segmentation Using ANFIS Classifier.” Asian Pacific Journal of Cancer Prevention: , volume19, issue 11, pp. 3203-3209. 2018.

[5] Kashyap, Debashree, Abhishek Somani, Jatin Shekhar, Anupama Bhan, Malay Kishore Dutta, Radim Burget, and Kamil Riha. "Cervical cancer detection and classification using Independent Level sets and multi SVMs." In 2016 39th international conference on telecommunications and signal processing (TSP), pp. 523-528. IEEE, 2016.

[6] Soumya MK, Sneha K, Arunvinodh C. Cervical cancer detection and classification using texture analysis. Biomedical and Pharmacology Journal. volume 9, pp.123–34, 2016.

[7] Wei, Low Zhe, Wan Azani Mustafa, Mohd Aminudin Jamlos, Syed Zulkarnain Syed Idrus, and Mohd Hamzari Sahabudin. "Cervical cancer classification using image processing approach: A review." In IOP Conference Series: Materials Science and Engineering, vol. 917, no. 1, p. 012068. IOP Publishing, 2020.

[8] Meenakshi Sharma, Sanjay Kumar Singh, Prateek Agrawal and Vishu Madaan, “Classification of Clinical Dataset of Cervical Cancer using KNN”, Indian Journal of Science and Technology, Volume 9, issue 28, pp.45-56, 2020.

[9] Bai, Bing, Pei-Zhong Liu, Yong-Zhao Du, and Yan-Ming Luo. "Automatic segmentation of cervical region in colposcopic images using K-means." Australasian physical & engineering sciences in medicine 41 (2018): 1077-1085.

[10] Schölkopf, Bernhard, and Alexander J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.

[11] Cai, Weiling, Songcan Chen, and Daoqiang Zhang. "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation." Pattern recognition 40, no. 3 (2007): 825-838.

[12] Lin, Chun-Fu, and Sheng-De Wang. "Fuzzy support vector machines." IEEE transactions on neural networks 13, no. 2 (2002): 464-471.
[13] NCCC. Cervical tumor. 2010. Available at: http://www.nccc-online.org/index.php/cervicaltumor.

Author's biography

Dr. P.A. Gowri Sankar is currently working as an Associate Professor in the Department of Electrical and Electronics Engineering at Knowledge Institute of Technology, Salem. He is a highly accomplished professional with a strong background in Electrical and Electronics Engineering. He is Holding a Ph.D. in Electrical Engineering, with an impressive research portfolio. His research interests span various domains, including Intelligent controllers for Power Electronics and Power System Applications, Machine learning, Artificial Intelligence, Deep learning, Image Processing, Nanoelectronics, Nanosensor & materials. He has published 15+ research articles at national and international level and 06+ International Conferences. His two design Indian patents are granted. Currently, He is supervising 06 PhD Scholar at Anna University Chennai. In addition to their academic achievements, He has also been involved in consultancy projects with the industry's top companies and research projects. With over 11 years of work experience, his strengths include a positive attitude, teamwork, commitment, and confidence.

R.Aparna - is currently pursuing her Master degree in Embedded Systems Technologies in the Department of Electrical and Electronics Engineering at Knowledge Institute of Technology, Salem. He is a highly accomplished professional with a strong background in Electrical and Electronics Engineering, Image processing and Embedded C programming.

K.Sathishsharama, K.Aravinth, V.Thirumalai, M.Sathishkumar is currently pursuing his Undergraduate degree in the Department of Electrical and Electronics Engineering at Knowledge Institute of Technology, Salem. His strong background in Electrical and Electronics Engineering, Image processing and Aurdino programming.