International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Enhancing the Classification of Pap Smear Images using ENN – TLBO classification Method

Geetha , S. Suganya

Abstract: ***The most common cancer among the women younger than 35 in developing countries is cervical cancer. It is a human papilloma virus disease. It should be identified earlier by Pap smear test and treated earlier to avoid the consequences. Pap test a colonoscopy is widely used to check the vagina and the cervix. The Pap smear test is the most effective medical test, but it causes difficulty under the microscope at the point of analysis. Automatic cancer detection is designed to unravel the downside. This identification process involves some image processing methods, such as segmentation, and an improved SVM classification method. In this paper, an efficient Elman Neural Network (ENN) collaborating with Teaching Learning Based Optimization (TLBO) algorithm is proposed to classify cancer using Pap Smear Test images. At first, an input image of Pap smear is converted into grey level from the RGB. The grey level image is preprocessed to eliminate unwanted noise produced and smoothened with Kuan Filter (KF). Active Contour Method (ACM) has been used to segment the identified cells from the Pap smear image. Features such as GLCM, haralick, solidity, shape, and other mathematical features are extracted for improving the accuracy. Classification has done using ENN-TLBO. TLBO is utilized for getting optimal weights during the training phase. Performance evaluation has done through the experimental outcomes where, ENN-TLBO yields good accuracy of 86.6%, than the prevailing algorithms such as Support Vector Machine, Radial Basis Function Neural Network classifiers.***

Keywords: ***Cervical cancer, Pap Smear test, ENN-TLBO, RBNN, SVM.***

I. INTRODUCTION

Cervical cancer is one of the deadly disease affected to woman throughout the world which is caused by Human Papilloma virus. To identify the cancer at its early stage is Pap test which is used test for testing the cervix and the vagina [1]. The test determines the unusual cells in the cervix and classifies them as normal and abnormal. The aim of the Pap Test Diagnosis Classification is to determine Pap test Cells whether they are infected or not. The word "Pap-Smear" refers to samples of human cells that are stained by the Papanicolaou method. A variety of research papers on segmentation, classification, have been published in the last 35 years [2]. A segmentation process usually involves the localization of the nucleus.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Geetha *, Assistance Professor, Charan's Degree college. Bengaluru, Karnataka.,India.
S. Suganya, Associate Professor, RVS College of Arts and Science, Sulur, Tamil Nadu.,India.

Classification methods only work with already segregated cervical cells, ignoring the rest of the picture. Multiple sexual partners, early sexual contact and pregnancy prevention drugs are the various causes of cervical cancer.Pap test has some of the disadvantages like, low sensitivity, expertise dependent and the requisite for retesting etc., To identify the actual stage of cervical cancer, an efficient segmentation and classification techniques is used.

The extracted cervical cells are studied under a microscope for the manual classification of abnormal and normal cervical cells. Nonetheless, due to human error in cell sorting, this approach is subject to high false positive levels. This approach is very cost-effective and only 4 to 5 slides per day can be counted by pathologists [3]. Due to the cytoplasm and nucleus present in the cell structure, it is difficult to perform the process at a faster rate. Any computer aided cervical cancer screening system includes two fundamental processes: segmentation and classification.

II. RELATED WORK

A lot of research has been done on the classification of cervical cells. Classifying cells into regular or abnormal cells is known as a two-tier problem, and classifying cells into one of the seven groups is known as a seven-tier problem. The nucleus region is represented by nine features, while the cytoplasm region is represented by eleven features.

M.K.Soumya et al. proposed a classification strategy using magnetic resonance imaging for cervical cancer staging. The result was compared to non-linear SVM classification models based on the second-order texture and transformation characteristics of the tumor[4]. The texture and statistical characteristics play a vital role in classifying cancer as normal and abnormal. Obrayan H.Gomez et al. used integrated algorithms to construct a classifier. The results of the experiments show that the use of combination algorithms to bagging cum multilayer perceptron results in a high percentage of correctly classified instances[5]. Martin has introduced both supervised and unsupervised learning processes to categorize cervical cells at different stages. He used three K-fold cross validation classifiers, namely Hard C-means, Gustafson-Kessel (GK) and Fuzzy C-means. The efficiency of the proposed system is better than previous classification techniques[6].

Y.M.S.AlWesabi et al., analyzed and presented various classification techniques and demonstrated the benefits of feature selection approaches for the best prediction of cervical cancer disease. This paper was very much helpful in understanding advantages and disadvantages of classifiers [7].

Jantzenetal., proposed linear network model for cervical cell classification. It is the simplest classifier that uses a linear activation function to minimize squared error. This approach has identified cervical cells in different stages. The performance is calculated by the use of K-fold cross-validation with no reduction of features[8].

Athinarayan et al., proposed model for image processing techniques such as segmentation,enhancement, extraction, and character classification. This method quite clearly identifies specifics of the region of interest seen from the initial cervical cells and arc with greater precision than the original image. Cervical cancer cells that are classified as CIN 0(Normal Dysplasia), CIN 1(Mild Dysplasia). This method is very useful in the selection of features that are most suitable for all classes such as LSIL (low-grade squamous intraepithelial lesion) and HSIL (high-grade squamous intraepithelial lesion) [9].

Expert manual identification of cervical cells for cancer detection is time-consuming and laborious. Due to the presence of a large number of cells on the glass slide, the problem becomes more complex. Other factors that may lead to error by experts include poor contrast and inconsistent staining[10]. Any signs of cervical cancer can be diagnosed at the initial stage. And the difficulty lies in the process of analyzing microscopic images where there is little certainty about the impact of cancer. Medical experts agree that "detection of cervical cancer is an easy task and a diagnosis can be correct, but prejudice against cervical cancer is highly subjective and needs detailed discussion among experts"[11]. In view of all these variables, it is clear that an image classification algorithm is required to identify cells in a short time and to reduce screening errors in the diagnosis of cervical cancer.

III. OVERVIEW OF CERVICAL CANCER

The cervix, also referred to as the uterine cervix is the lower part of the uterus. The fetus develops in the upper part of the body of the uterus. The cervix binds the body of the fetus to the vagina ie., the birth canal. The part of the cervix that is nearest to the body of the uterus is called the endocervix. The part, exocervice (or ectocervix) is next to the vagina. The glandular cells (on the endocervix) and the squamous cells (on the exocervix) are the two main cells that cover the cervix. Transformation zone is the place where these cells meet and where cervical cancers actually begins.

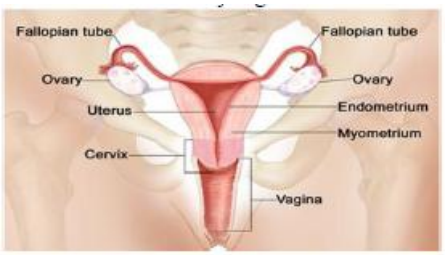

Fig. 1 Female Reproductive System

Such cells do not suddenly turn into cancer. Alternatively, normal cervical cells first slowly undergo pre-cancerous changes that transform into cancer. There are several terminology used to describe these pre-cancer changes, including cervical intraepithelial neoplasia (CIN), squamous intraepithelial lesion (SIL) and dysplasia. These improvements can be detected and treated in the Pap test to prevent cancer growth. Figure 1 displays the female reproductive system

A. Pap Test

Pap (or Pap) is a technique used for the cervical cytology examination of cells from the cervix. The healthcare professional puts the speculum first within the vagina. A speculum is a metal or plastic tube that holds open the vagina, so that the cervix is clearly visible. First, A sample of cells and mucus is gently scraped from the exocervice using a thin spatula (the portion of the cervix closest to the vagina). Then, a tiny brush or cotton-tip swab is inserted into the cervical opening to take a sample of the endocervix (the inner portion of the cervix nearest to the uterus body). Then, cell samples are prepared to be tested in the lab. This is shown in figure 2. The basic classification of Cervical cells on Pap smear test is given in Figure 3.

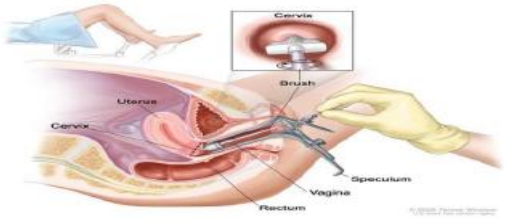

Fig. 2 Pap Test

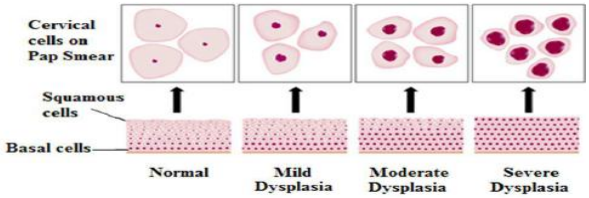

Fig 3 – The Proposed Architecture

B. Proposed Architecture

Figure 3 demonstrates the theoretical method for detecting cervical cancer using images of Pap smear. Conventional Pap smear images are the input images used in this process, which are microscopic optical images in a.bmp format. In this study, an image of a single cell Pap smear is considered to determine the cell as normal / abnormal. The first stage is the pre-processing phase where Kaun filtering is used in the conversion and enhancement of colors to prepare the image for further processing. The next stage is the processing block, which is aimed at separating the nucleus from a single cell. Active Contour Method(ACM) is used to segment the identified cells from the image of the pap smear. Following this stage is the extraction feature where the nucleus region is extracted to identify the cervical cells as normal or abnormal.

The next process is classification. In this, we use Elman Neural Network is used to classify the defected nucleus into seven forms. The main advantage of Elman-NN is that the background nodes can be used to memorize the previous hidden node data, which makes Elman-NN relevant in the fields of dynamic system recognition and prediction control.

i. Dataset Description

The dataset of DICOM image of Pap smear cancer is gathered from the kaggle website. There are around 250 images of various ages are utilized for assessing the cervix cancer identification.

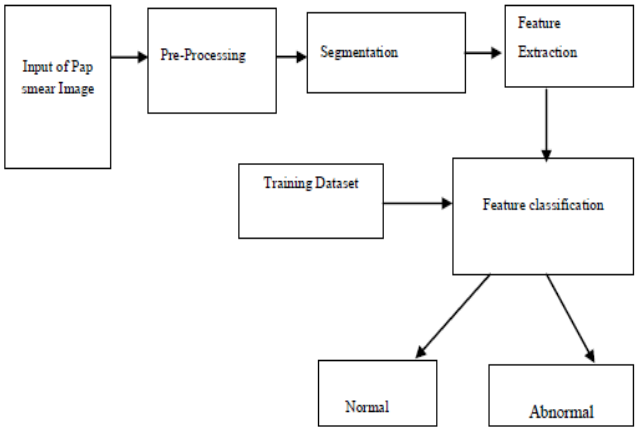

Fig 3 – The Proposed Architecture

ii. Pre-processing stage

Primarily, this stage is performed before certain special processing purpose on image. This stage has enhanced the quality of image and eradicates noise. Before preprocessing, the input Pap smear image is converted to gray scale image. In the anticipated system, Kuan Filter (KF) is executed for image de-noising and smoothing of image quality.

iii. Kuan Filter:

To perform image smoothing, Kuan filter is utilized [13]. This method converts the multiplicative noise model into a signal-dependent additive noise model, and then applies the criterion of minimal mean square error (MMSE). The value of the pre-processed image can be obtained by

$$\tilde{R}(t) = I(t).W(t) + \bar{I}(t).[1 - W(t)] \quad (1)$$

$\tilde{R}(t)$ → De-noised image,
I (t) → image corrupted with noise
$\bar{I}$ (t) → mean image intensity within the filter window.

W (t) → weighted coefficient of Kaun filter determined as:

$$W(t) = \frac{1-\frac{c_u^2}{c_I^2(t)}}{1+c_u^2} \quad (2)$$

Where c_u and $c_I(t)$ are variation coefficients of speckle u (t) and image I (t) respectively. Filtering method aims at smoothed small discontinuities, which is begin with background noise which makes to normalize the image.

iv. Segmentation using ACM

Image segmentation is an essential task for image analysis to partition an image into multiple sub-regions based on a desired feature. Active contours were widely used methods of image segmentation, as it produces sub-regions with continuous boundaries to the methods of kernel-based edge detection.

ACM:

An active contour model is commonly known as snake illustrated by [14] is a spline-minimizing energy guided by external constraints and influenced by image forces that pull it towards features such as lines and edges [15]. The snake can be represented by a set of points $P(s,t) = (x(s,t), y(s,t))$ on Pap smear image parameter relating to s ∈ [0, 1]. The energy function can be written as:

$$E_{snake} = \int_0^1 E_{int}(v(s)) + E_{ext}(v(s))ds \quad (3)$$

When the external energy of the snake come close to object boundary position, to calculate the gradient of the image as

$$E_{external} = E_{Image} + E_{con} \quad (4)$$

The Internal Energy is calculated using equation (5).

$$E_{internal} = E_{cont} + E_{curv} \quad (5)$$

Where E_{cont} represents the snake's stability and E_{curv} represents the snake's elastic degree and the equation are given below:

$$E_{cont} = \alpha(s)|v_s(s)|^2 \quad (6)$$

$$E_{curv} = \beta(s)|v_{ss}(s)|^2 \quad (7)$$

$$E_{internal} = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2)/2 \quad (8)$$

v. Feature Extraction:

The Features of Pap smear images are extracted in favour of detecting the cancer part of the cervix region.

These characteristics are supposed as an input to the classification.

Some of the characteristics are as texture features such as GLCM, haralick Features, size, shape of Pap smear image, and geometric features like mean, median, entropy, irregularity, concavity, convexity, area and perimeter has been extracted.

The above extracted features are used for effective feature classification method using ENN-TLBO**.**

vi. Feature Classification

The Pap smear cancer is ordered with ENN classifier. To advance the preparation method of ENN, TLBO algorithm has been introduced. The classified is displayed the prediction accuracy and compared the performance results with other algorithms.

Elman's neural network is feed forward network with an input layer, a secret layer (undertake layer), an output layer and a special layer called background layer. In the background layer the output of each hidden neuron is copied to a specific neuron.

In an Elman network, the weights from the secret layer to the context layer are set to one and are fixed because the context neuron values must be exactly copied.

The Elman network can be educated using methods of gradient back propagation and optimization.

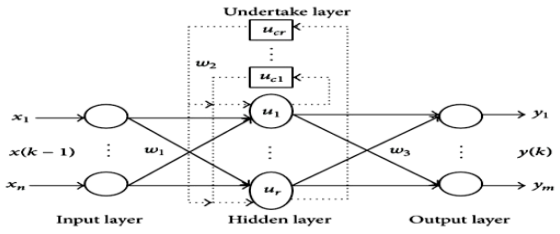

Fig.4: Architecture of ENN

The fig.4 shows the architecture of Elman Neural Network. Suppose with n input, m output, the number of hidden and undertake neurons are r, the weight of hidden layer to output layer is w_3;$u(k-1)$ is the input of neural network, $x_c(k)$ is the output of undertake layer, and $y(k)$is the output of neural network; the weight of input layer to hidden layer is w_1, the weight of undertake layer to hidden layer is w_2, $x(k)$ is the output of hidden layer; then.

$$x(k) = f\left(w_2 x_c(k) + w_1\left(u(k-1)\right)\right) \quad (9)$$

Where $x_c(k) = x(k-1)$

f is the hidden layer transfer function, which is commonly used in S-type function; that is,

$$f(x) = (1+e^{-x})^{-1} \quad (10)$$

g is the transfer function of output layer, which is often a linear function; that is,

$$y(k) = g\left(w_3 x(k)\right) \quad (11)$$

Elman neural network uses BP algorithm to revise weights; the error of network is

$$E = \sum_{k=1}^{m} (t_k - y_k)^2 \quad (12)$$

Where tk is the output vector of object. To enhance the accuracy of ENN prediction, Teaching Learning Based Optimization (TLBO) has been used for optimizing the weight value of ENN.

Optimal Weight selection using TLBO

To optimize the weight of ENN and to enhance the exactness of classification, the TLBO is used. TLBO is also a population-based method and uses a population of solutions to proceed to the global solution. The population is a group of learners or a class of learners. The process of TLBO is divided into two parts: the first part consists of the 'Teacher Phase' and the second part consists of the 'Learner Phase'. 'Teacher Phase' is learning from the teacher and 'Learner Phase' is learning by the interaction between learners.

TLBO algorithm for weight optimization:

In the teacher phase, the best solution in the entire populace is considered as the teacher, where shares their knowledge to learners. (*Xu Chen et al.)*

Assume X_i= $\left(x_i^l, \ldots x_i^d, \ldots x_i^D\right)$ is the position of i[th] learner with the best fitness is identified as the teacher $X_{teacher}$, and the mean position of learner groups can be represented as $x_{mean} = \left(\frac{1}{p}\right)\sum_i^P X_i$. Each learner phase is updated by the following equation:

$$X_{i,new} = X_i + r.\left(X_{teacher} - T_F.X_{i,mean}\right) \quad (13)$$

Where r represents a randomly selected number in the range of 0 and 1 and T_F is a teaching factor

In the student phase, a student randomly interacts with other students for their improvement of performance. The student randomly selects another student and the learning process can be expressed by the below equation:

$$X_{new}^i = X_{old}^i + rand(X^k - X^i), \quad \text{if } (X^i) > f(X^k \quad (14)$$

$$X_{new}^i = X_{old}^i + rand * (X^i - X^k), \quad \text{if } f(X^i) < f(X^k) \quad (15)$$

If the new solution X_{new}^i is better, it is recognized in the populace. The algorithm will continue until the termination condition is met. The proposed ENN-TLBO algorithm is given below

Algorithm: ENN-TLBO based Pap smear test Image classification

Inputs**:** A training set of n classes $A = [A_1, A_2, .., A_n]$, a test sample$T$, hidden feature number$H$, activation function k(x)

Output: Cancer identification

1. The hidden feature parameters (i.e.w_i, r_i) are randomly generated.
2. Set the Objective function f(X),X=$((x_1, x_2, \ldots, x_n)\text{^}V$
3. Initialize the learners(weights) and evaluate all the learners (X)
4. Determine the best learner and acted it as a teacher
5. Compute the mean of all the learners (X) and denoted as mean
6. While (the termination conditions are not met)
7. {teacher phase}
8. Calculate the teaching factor using the equation $X_{new} = X_{old} + r.(X_{teacher} - T_F.X_{mean})$
9. Accept X_{new} if f(X_{new}) is better than f(X_{old})
10. End if (teacher phase)
11. {student phase}

Randomly select the two learners X_i and X_k such that $k \neq i$

If X^i is better than X^j, i.e. $f(X^i) < f(X^k)$

$$X_{new}^i = X_{old}^i + rand * (X^i - X^k)$$

Else

$$X_{new}^i = X_{old}^i + rand(X^k - X^i)$$

End if

IfX_{new}^i is better than X^j, i.e. $f(X_{new}^i) < f(X^i)$

$$X^i = X_{new}^i$$

End if {end of student phase}

19. Accept X_{new}^{i} if f(X_{new}^{i}) is better than f(X_{old}^{i})

20. End if (student phase)

21. Weight optimal results are stored for the ENN

22. Pap smear cancer is classified and accuracy is calculated.

Repeat the different steps until finding the best optimal weight with a minimum error and the cancer is classified in an efficient manner.

IV. OUTCOMES AND DISCOURSE

In this section, the proposed method for detecting cervix cancer automatically based on Pap smear test has been experimented with the traditional classifiers such as RBNN, SVM and compared with our hybrid method as ENN-TLBO. All the Pap smear images are structured and then tested with the MATLAB. In entire, around 250 sample Pap smear images are collected for the evaluation. In this, 150 images are trained and 100 images are tested by ENN-TLBO SVM. In future, to test the Pap smear image with other algorithm that helps in detecting the cancer ease and fast.

classification. Overall performance of cervix cancer detection is performed in Figure 6. It exemplifies that predicted ENN-TLBO has attained more performance with 86.6 %accuracy, 86.83 % of specificity, 64% of precision, 90% of recall and 69.33% of f-measure than RBNN and SVM.

V. CONCLUSION:

In this paper, Cervix Cancer Identification using Pap smear test based on ENN-TLBO classification has been proposed. The input Pap smear image has converted to gray scale image and then preprocessed using Kuan filter which produces the smoothened and normalized image. ACM is utilized to get accurate segmented region from the preprocessed image. Features are extracted and ENN classifier is utilized for classification. In order, to produce accurate results, and issue in allocating weights in ENN, an optimization technique called TLBO has been used. The outcome of the proposed ENN-TLBO has attained high accuracy of 86% than other classifiers such as RBNN and

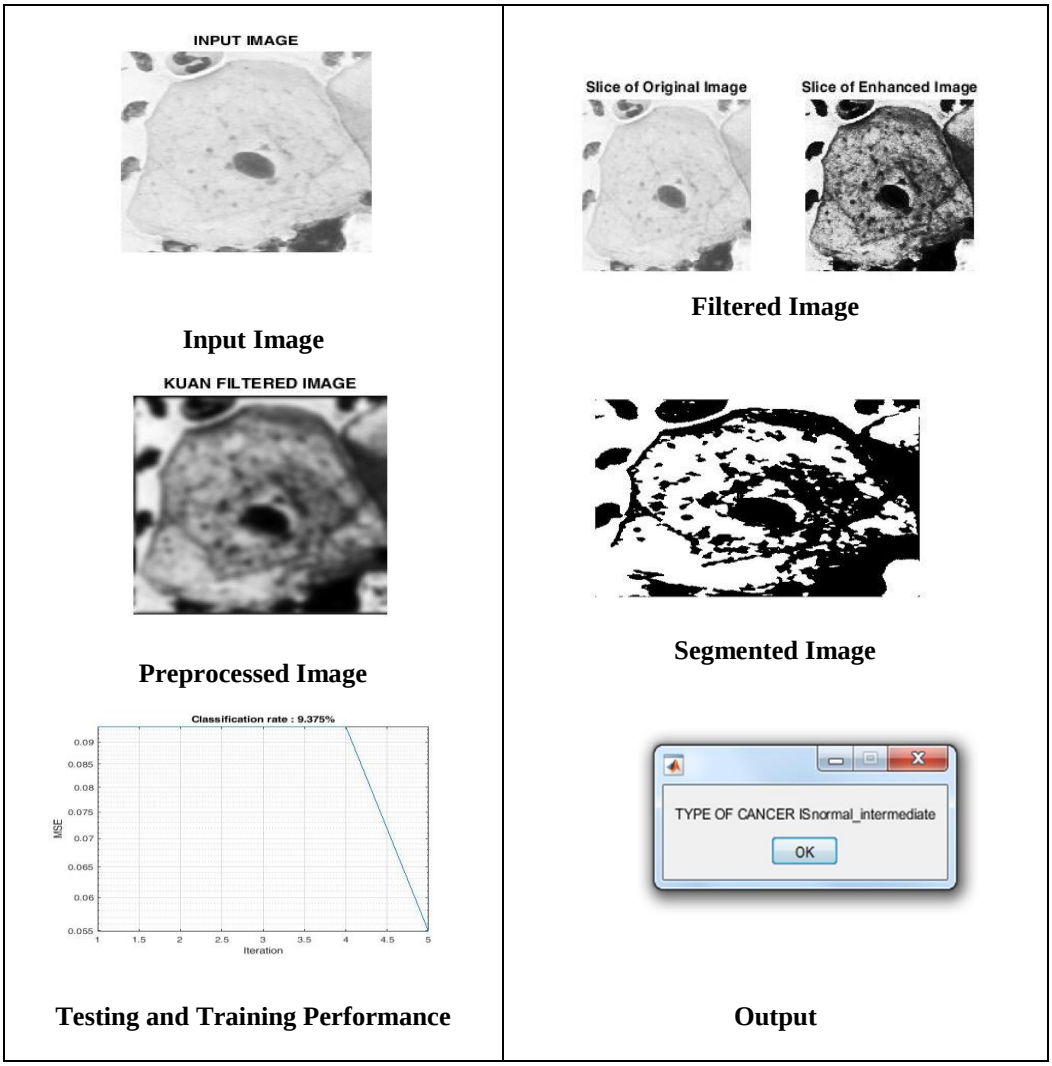

Fig.5: Proposed ENN-TLBO based results for Input Pap smear Image

Table1: Numerical Evaluation of Cervix cancer detection

Performance Metrices	ENN-TLBO	RBNN	SVM
Accuracy (%)	86.67	66.67	80
Specificity (%)	86.83	66.25	72.24
Precision (%)	64	24.96	60
Recall (%)	90	55	77.8
F-measure (%)	69.33	34.12	74.5

REFERENCES:

1. http://www.imaginis.com/cervical-cancer /cervical-cancer-introduction.
2. Sharma.P, Pattanshetty.S.M., 2017, "A Study on Risk Factors of Cervical Cancer Among Patients Attending A Teritary Care Hospital: A Case-Control Study,", Clinical Epidemiology and Global Health (In press).
3. M.Anosouya Devi, S.Ravi, J.Vaishnavi and S.Punitha, "Classification of Cancer using Artificial Neural Network" 1877-0509 © 2016
4. M.K.Soumya, K.Sneha and C.Arunvinodh, "Cervical Cancer Detection and Classification Using Texture Analysis", Biomedical and Pharmacology Journal, Vol. 9(2), 663-671 (2016).
5. Obrayan H. G´omez, Eddy S´anchez-DelaCruz and A. Paulina de la Mata, "Classification of Cervical Cancer Using Assembled Algorithms in Microscopic Images of Paponicolaou", Research in Computer Science, pp. 125–134; rec. 2017-03-14; acc. 2017-05-06.
6. E.Martin, "Pap-smear Classification", Technical University of Denmark:Oersted-DTU, Automation, 2003.
7. Y. M. S. Al-Wesabi, Avishek Choudhur, "Classification of Cervical Cancer Dataset", Proceedings of the 2018 IISE Annual Conference.
8. J.Jantzen, J.Norup, G.Dounias, "Pap-smear benchmark data for pattern classification", in proceedings of Nature inspired Smart Information Systems (NiSIS 2005), 2005, pp. 1-9.
9. S. Athinarayanan and M.V. Srinath, "Classification of Cervical Cancer Cells in Pap Smear Screening Test", ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING, MAY 2016, VOLUME: 06, ISSUE: 04.
10. P.W.Huang andY.H.Lai, "Effective segmentation and classification for HCC biopsy images", Pattern Recognit., vol. 43, no. 4, pp. 1550-1563, 2010.
11. Y.Marinakis, G.Dounias and J.Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification", Computers in Biology and Medicine, Volume 39, Issue 1, January 2009, Pages 69-78.
12. P.Robert, A.Celine Kavida, Published a paper titled "GENETIC ALGORITHM BASED DENOISING OF MEDICAL IMAGE" in International Journal of Control Theory and Applications ISSN: 0974-5572, 13th June, 2016
13. 13. Benes, Radek & Riha, Kamil. (2012). Medical Image Denoising by Improved Kuan Filter. Advances in Electrical and Electronic Engineering. 10. 10.15598/aeee.v10i1.529.
14. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," Int. J. Comput. Vis., vol. 1, pp. 321–331,
15. Chai, Douglas & Ngan, King. (1998). Locating facial region of a head-and-shoulders color image. 124 - 129. 10.1109/AFGR.1998.670936.

AUTHORS PROFILE

Mrs. Geetha is working as an Assistance Professor in Charan's Degree college. She Completed BSc(CSc), MCA, Mphil, Persuing Phd in barathiar university. She has 10 years of IT experience and 1 year of teaching experience in institutions. She have published review paper of cervical cancer using data mining techniques. She is the Research Scholar for Data Mining application.

Dr. Suganya is working as an Associate Professor in RVS College of Arts and Science. She Completed Ph. D from Anna university in the year 2014. Her specialization is moblie computing and data science. She has 20 years of teaching experience in various institutions. So far she published 25 papers in various international journals and 10 papers presented in national and international conferences. She is the supervisor for first author.