

Survey on Detection of Cervical Cancer

Leel Sukel M K¹, Leel Sukesh M K², Naveen R³, Theja N⁴

^{1,2,3}Students, Department of Computer Science and Engineering, Eighth Semester, Vidya Vikas Institute of Engineering & Technology, VTU, Mysuru, Karnataka, India

⁴Assistant Professor, Department of Computer Science and Engineering, Vidya Vikas Institute of Engineering & Technology, VTU, Mysuru, Karnataka, India

Abstract - Cervical cancer is the second deadliest malignant growth prompting woman's morbidity and mortality on the planet. More than 0.2 million individuals lose their life because of cervical malignancy consistently. Pap smear is one of the easiest and most critical examination to altogether decreases the death rate of cervical malignancy. In any case, such conclusion is vigorously reliant on the clinician's understanding, which is very tedious and is exposed to human mistake notwithstanding for experienced specialists. The principle issue with this malignant growth is that it can't be identified as it doesn't toss any side effects until the last stages. This is ascribed to the malignant growth itself and furthermore to the absence of pathologists accessible to screen the disease.

Key Words: Feature Extraction, Support Vector Machines, Thresholding, Segmentation, Nucleus area, Region of Interest

1. INTRODUCTION

Cervical cancer is most common malignancy in female. It arising from the cervix. It is due to the abnormal growth of the cells that have ability to spread to the other parts of the body. Earlier no typically symptoms are seen later on abnormal vaginal bleeding, pelvic pain. Infection with some type of Human Papilloma Virus is the great risk factor for cervical cancer. Cigarette, smoking both active and passive increases the risk of cervical cancer. Symptoms like it may include loss of appetite, weight loss, fatigue, back pain, leg pain, swollen legs, and vaginal bleeding. The growth spread directly to the adjacent structure to the vagina and to the body of uterus. Rarely the virus spread to liver, lungs and bones may occur. There are different treatment like primary surgery, primary radio therapy, chemo therapy and combination therapy. There are many techniques to diagnosis cervical cancer the important ones are Pap smear test, LBC test, HPV test, Biopsy and different screening techniques.

2. LITERATURE SURVEY

In this paper [1], The input image is an ultrasound image. In the pre-processing the input Image has a lot of frail highlights which should be reinforced with the goal that highlights feature more precisely. Likewise, to decrease the

running time it is better that we focus just on the Region of Interest instead of the entire image. There are various strategies like edge detection and fuzzy clustering proposed for identifying these Region of Interest. Anyone of these can be utilized for identifying the Region of Interest. The pre-processing first converts the image to grey scale color model and after that filter the noise with the assistance of trim mean filter. Trim mean filter is the crossover of mean and median filter [1]. In feature extraction is the phase where tumors are recognized. To distinguish malignant growth in each stage you need diverse highlights and consequently they have epitomized a lot of extraordinary highlights that won't just say where the cell has side effects of disease yet in addition it would tell at what organize the disease is in.

In this paper [1], the pre-cancerous stage cells cluster together to end up denser and become tumor cells. Using the mean component extraction one can get recognize when this procedure begins. When you take a look at a cervical cell it would be light shaded. In any case, in the event that you put at least two cells over it, it would look as though it is darker. So as to locate these darker elements utilize a histogram $h(v)$ to graph the circulation of hues in the picture. Then the mean of the color frequency distribution is determined by

$$M1 = \frac{1}{M0} \sum_v v^1 h(v) = \sum_v v H(v)$$

Standard deviation is same as mean but it will in general consider the all number of pixels and look the amount it goes away from the neighboring cells [1]. This measurable information can differentiate between a cell and its neighbors and is given by

$$M2 = \frac{1}{M0} \sum_v v^2 h(v) - (M1)^2$$

Skewness: Every cell procedure a one of a kind shapes yet when cells bunch together, they lose their shape and become sporadically shape [1]. This regularly occurs at the beginning of the pre-carcinogenic stage and is dictated by

$$MB = \frac{1}{(M2)^3} \left(\sum_v v^3 h(v) - 3 \sum_v v h(v) \sum_v v^2 h(v) + 2 \sum_v v h(v)^3 \right)$$

Kurtosis: At the point when the malignant growth enters the precancerous stage, tumors begin to develop [1]. Tumors are typically denser than ordinary cells as they are 3 dimensional. To identify these tumors initially distinguish the thickness which is given by

$$M4 = \frac{1}{(M2)^4} \left(\sum_v v^4 h(v) - 4 \sum_v v h(v) \sum_v v h(v) + 6 \sum_v v h(v)^2 \sum_v v^2 h(v) - 3 \sum_v v h(v)^4 \right)^{-3}$$

Acoustic shadowing: Acoustic Shadowing happens when the sound wave experiences a very echo dense structure; almost the majority of the sound is reflected, bringing about an acoustic shadow [1].

Classification: Support vector machines (SVM) are a set of related directed learning strategies which investigate information and perceive designs, utilized for factual order and relapse examination. SVM is a classifier which develops a hyper plane or set of hyper planes in a high or boundless dimensional space, which can be utilized for characterization [1].

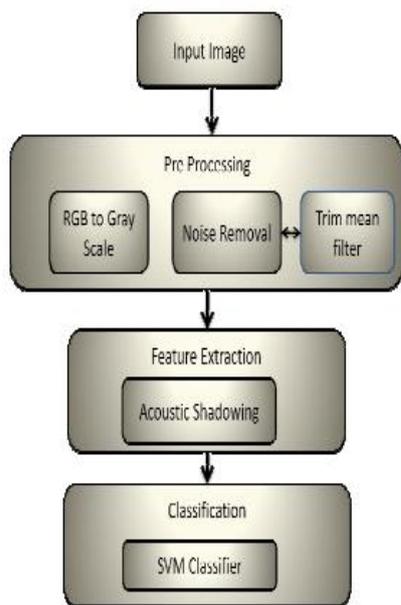


Fig-1: Block Diagram.

In this paper [2], the input images utilized in this technique are typical Pap smear images which are microscopic optical .bmp images. In this work they have considered single cell Pap smear image for assessing the cell as normal/abnormal. The primary stage is the pre-processing which comprises color conversion and improvement to set up the image for further process. The following stage is the processing whose goal is to fragment the core from a unit cell. Following this stage is the feature extraction where the area of the core is separated for grouping of the cervical cells as normal or abnormal.

Pre-processing: The Pap smear images are color images which are of low quality because of stains used to color the cells and uneven lighting over the field of view. The data is inconsequential contrasted with the intensity of the image. So, in the pre-processing stage, the initial step is to change over the RGB picture to Grey scale image. RGB image is changed over to grayscale by forming a weighted sum of the R, G, and B segments utilizing the condition:

$$I_{gray} = 0.2989 * R + 0.5870 * G + 0.1140 * B$$

It is required to have correct limit between the nucleus and the cytoplasm to separate nucleus from the cell. The Pap smear pictures is poor and the absence of homogeneity in image force makes further processing hard. The precision of image segmentation and feature extraction basically relies upon this factor. The following block is the enhancement block and linear scaling is utilized to improve the contrast of the image with the goal that just the core limit is plainly observed from the foundation image making it increasingly appropriate for division.

Processing: Transforming the input data into the set of features is called feature extraction. It includes calculation to distinguish and disconnects different required parameters. The area of the nucleus is the number of pixels inside the nucleus. The zone of the core is the quantity of pixels inside the core. The region of the core of the typical cells and strange cells is determined amid preparing stage. The mean zone of the core of normal cells and abnormal cells are determined using the following equation

$$\text{Mean nucleus area} = \frac{\sum_{i=0}^n \text{Nucleus Area}}{n}$$

Where n = number of images

The principal procedure in this stage is segmentation which is the basic advance in image examination or image understanding. It helps in extracting region of interest for an image. Segmentation should able sperate unwanted things from region of interest. An automatic bi-level thresholding method is used which uses the global mean of the image to calculate the threshold value iteratively. The negative of the sectioned image is increasingly appropriate as it helps in evacuating undesirable cluster using morphological erosion quicker than the original image. The resultant image has nucleus region along with small spurious objects which can be removed by performing erosion operation.

Classification: The arrangement of the image depends on the area of the nucleus of cells. Feature extraction results demonstrated that normal cells have smaller nucleus with area less than the calculated threshold value and abnormal cells have larger nucleus with area greater than the threshold value. In the Testing Phase the area of the nucleus of the input image will be calculated and compared with the mean threshold nucleus area of normal cells (μT_{norm}) calculated. If the nucleus area of the input image is greater

than μT_{norm} then it is classified as abnormal else it is classified as normal.

Performance measures: In medicinal diagnostics, affectability and particularity are the execution estimates which determine the exactness of diagnosis method. It should have a high (100%) affectability and explicitness of any diagnosis method. sensitivity is the potential of a test to effectively recognize those with the malady (true positive rate), though sensitivity is the ability of the test to accurately distinguish those without the disease (true negative rate). The sensitivity test gives the probability of a positive test given that the patient is ill. It is given as:

$$Sensitivity = \frac{Number\ of\ TP}{Number\ of\ TP + Number\ of\ FN}$$

The specificity test gives the probability of the negative test given that the patient is well:

$$Specificity = \frac{Number\ of\ TN}{Number\ of\ TN + Number\ of\ FP}$$

-> True positive (TP): Disease which is correctly identify as disease.

-> False positive (FP): Healthy people incorrectly diagnosed as sick.

-> True negative (TN): Healthy people correctly diagnosed as healthy.

-> False negative (FN): Sick people incorrectly diagnosed as healthy.

Result: The experimentation is done using 80 samples of Papsmear images, out of which 40 (20 normal and 20 abnormal) single cell images are utilized for Training Phase and 40 images for Testing Phase. From the consequences of the Training Phase it is seen that the mean estimation of nucleus area for 20 normal cells is $\mu_{norm} = 1135$ pixels and that for 20 abnormal cells is $\mu_{abnorm} = 21500$ pixels. There is an unmistakable separation between the core zone estimation of ordinary and irregular cells and henceforth grouping of the cells should be possible dependent on territory parameter. Here they considered the mean edge an incentive to be $\mu T_{norm} = \mu_{norm} + T_n$ (where $T_n = 500$ pixels). In the Testing Phase, the proposed method is tested with 40 Papsmear images (20 normal and 20 abnormal images). The area of the nucleus of the input image was calculated and compared with the mean threshold nucleus area ($\mu T_{norm} = 1635$). The cell was classified as normal if nucleus area of the input image was lesser than μT_{norm} and classified as abnormal otherwise. Based on this method of classification a sensitivity of 100% and specificity of 90% was achieved.

In this paper [3], the proposed technique magnetic resonance imaging sweeps of cervical malignant growth were gained and its histogram image is divided in addition with edge location for extraction of tumorous region with accurate edge and shape. The anatomic practical situating just as its impact on other cervix territories are likewise decided. MRI outputs of cervical malignant growth persistent are utilized and are put away in Matrix Laboratory in .jpg format. Images are then changed over into a grey scale image dependent on power intensity of pixels in that image.

Pre-processing: The target image is represented in two-dimensional array $I(a, b)$. Every single component of matrix I speaks to a unit pixel. This image is then changed over to grey scale image of size (361, 512). Pre-processing is done on the images to upgrade the edges and different properties. To get the definite measurable information of every pixel histogram processing is done. Histogram is a plot drawn based on power dissemination of any grey scale image. Mathematically, a histogram is a discrete function $h(k_n)$ of a digital image which is depicted with various intensity level of range.

$$h(k_n) = r_n$$

where, k_n is the intensity level of n th pixel and r_n is the total number of pixels consisting the same intensity levels in MRI image. In this image, intensity range lies in between 0 and 1.

Processing: The image is transfigured into edge image utilizing the impact of the intensity variation. This transformation of image is obtained without going up against any distinctions in other physical properties of image. Filtering is done on the image to remove unwanted noises without losing the genuine edges. Enhancement is on a very basic level vital to decide the intensity changes in the

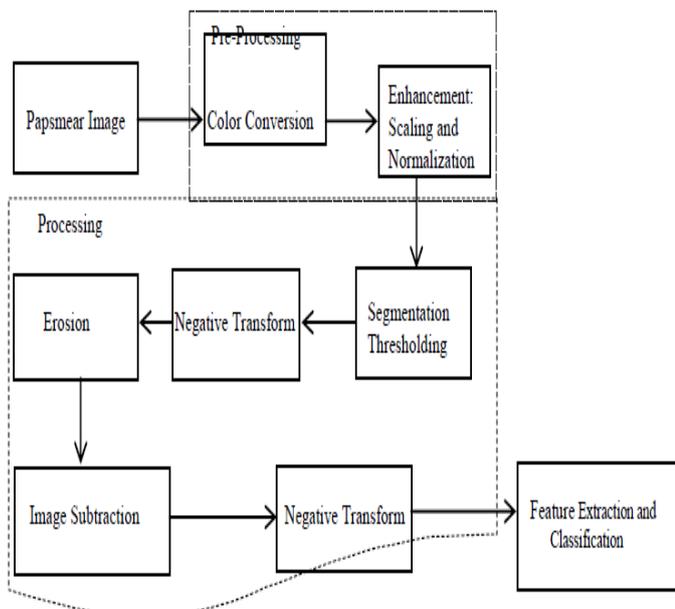


Fig-2: Cervical Cancer Detection System.

area point to encourage proper edge discovery. Detection separates the fake edge focuses from true edges the same number of pixels have nonzero gradient esteems which are not edges. Segmentation techniques are trailed edge identification. It isolates the image into its progressive structures. The engaged highlights are all around removed to find the region of interest utilizing the intensity of division. It is a moving assignment to segment the cervical tumors MRI examines in restorative imaging.

Post processing: Using threshold segmentation the given information MR image is partitioned into different regions based on the intensity values. Let intensity histograms represents an image $A(a, b)$, having low intensity objects on a high intensity background so that these intensities can be grouped into two major modes. This limit isolates items and foundation. This threshold separates objects and background. Multiple thresholding is done in the target image because this image is having multiple objects on the background. Two threshold values have been calculated P_1 and P_2

$$A(a, b) = \begin{cases} l, & \text{if } A(a, b) > P_2 \\ m, & \text{if } P_1 < A(a, b) \leq P_2 \\ n, & \text{if } A(a, b) \leq P_1 \end{cases}$$

where, l , m & n are different intensities. At any point coordinates (a, b) , if $A(a, b) > P_2$ then it belongs to the background, if $P_1 < A(a, b) \leq P_2$ then it is the outer cervix area and if $A(a, b) \leq P_1$ then it is the area of interest.

Watershed segmentation algorithm is connected on the angles or watersheds of the images as the regions in the image are characterized by the intensity varieties.

Result: The proposed system uses an adaptive histogram processing to improve the picture by dividing it into frames or tiles and re-joining the frames after enhancement. The target MR image scan is pre-processed using edge detection tool of image processing. At that point it is portioned utilizing different procedures like thresholding and watershed morphology. Some increasingly morphological operations like erosion and dilation make it conceivable to find appropriate elements of malignant cell zone.

In this paper [4], they have exhibited a method of an automated system which includes a 2-D photonic crystal based bio-sensor in combination with a machine learning algorithm. This mechanized framework can be utilized for malignant growth location without human intercession, decreasing blunders and expanding the precision. There is a need to increase the current plan of photonic crystal based bio-sensors with a tip of a tentacle to make them bio compatible. A photonic band gap (PBG) semiconductor crystal is a structure that could show light emissions by a few orders in magnitude similarly semiconductors control electric flows in channel source setup of MOSFETs or

generally electric potential in emitter base configuration of BJTs. Photonic crystal regularly made out of dielectric materials, that is, materials that act as electrical insulators in which an electromagnetic field can be propagated with low loss. The results of photonic band structure can be altered by filling in certain openings with a compound or natural material from stoichiometric composition. Because of impediment of research examination will limit extent of discourse to straight materials.

The modelling and simulation of sensor has been performed using a tool called MIT Electromagnetic Equation Propagation (MEEP), open source software accessible on a Linux OS. MEEP has unraveled the time shifting Maxwell's condition and actualize Finite Difference Time Domain Method (FDTD), known to be a ground-breaking numerical strategy in computational electromagnetism. In MEEP the space vector is partitioned into discrete matrix and the fields are resolved in time utilizing discrete time steps.

Arrangement in data analysis is the undertaking of allocating a class to instances data described by set of attributes. Classification or supervised-classification incorporates the development of a classifier which is trained on a set of training data that already has the correct class assigned to each data point. Bayesian classification depends on Bayes theorem. Arrangement in data analysis is the undertaking of allocating a class to instances data described by set of attributes. Classification or supervised-classification incorporates the development of a classifier which is trained on a set of training data that already has the correct class assigned to each data point. Bayesian classification based on Bayes theorem. Bayes Theorem is a statement from probability theory that allows for the calculation of certain conditional probabilities. The naive Bayes classifier is addendum with probability models with choice criteria. Among the numerous normal rules, one standard that emerges is that to pick the hypothesis that is most correct; this is known as the maximum a posterior or MAP decision rule.

Sensor design: A 2-D photonic crystal-based sensor of square cross section with line defect. The bars in air are supplanted with cervical tissue sample. The proliferation of light in photonic crystal changes relying on the refractive index of the normal and cancerous tissue. MEEP tool is utilized to design and simulate the sensor. The design specifications are:

Rods in air configuration.

Lattice Constant 'a' = 0.99.

Radius of rods 'r' = 0.2.

Di-electric constant of silicon slab = 12.

Di-electric constant of air is replaced with the sample.

Light source with Gaussian pulse with center frequency 0.4 and width of pulse 0.3.

Design and simulations are done using the following tools.

Tools used for simulation are MIT Electromagnetic Equation Propagation (MEEP) and MATLAB.

Classification: The outcome of the sensor is known to show distinct signature in the frequency and amplitude spectra even for slight change in the refractive index profile of the cervical cell. These characteristics are arbitrarily advanced however consequently extricated with the device under simulation in MEEP and naive Bayes classifier is utilized to group the data. A MATLAB application is developed to characterize the normal and cancerous transmission spectra for the permitted scope of light going inside the pass-band of the artificially fabricated structure of a photonic crystal in MEEP. The qualities of the data so obtained from the sensor are saved in MS Excel as a dot (.xlsx) extension.

Result: The attributes of the output of the sensor is extracted automatically. Based on the outputs the data from the sensor are classified and the class of the data is predicted. Here the classes are 'normal' and 'cancerous'. It is a two-state arrangement. The performance of the algorithm is determined by the confusion matrix. Since it is a two class the accuracy percentage is 100%.

In this paper [5], they utilized automated screening framework for advanced image acquisition, processing and classification techniques to differentiate normal cells from abnormal cells. They designed cervical malignancy detection framework to work with single cervical cell image. The major phases of this paper, Architecture of the Proposed Framework segmentation, feature extraction and classification. In segmentation it is utilized to separate nucleus from cytoplasm, and to detach cytoplasm from background. segmentation is done on a denoised picture. Segmentation is the important step because further classification result depends on segmentation accuracy.

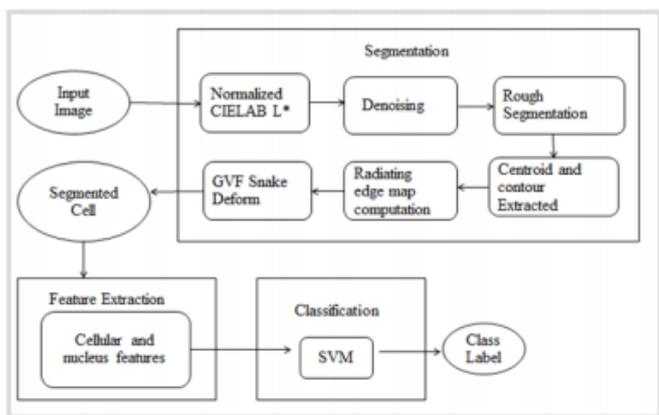


Fig-3: Architecture of the Proposed Framework.

Segmentation is performed by adapting Radiating Gradient Vector Flow (RGVF) snake. First image is inputted and converted from RGB color space to CIELAB space. L* (Luminance) dimension is extracted using converted image, and it is converted to grey scale image. Denoising will improve image quality by removing unwanted noise from the image. A spatial K-means based clustering method is used to separate cell into cytoplasm, nucleus and background. At first it divides cytoplasm and nucleus. Radiating edge map finds the accurate boundaries of cytoplasm and nucleus. GVF snake-based deformation is used later.

Feature Extraction: In this paper [5], Six nuclei based and three cytoplasm-based features are extracted.

Nuclei based features like

- Area of the nucleus A_n = Total No. of nucleus pixels
- Nucleus compactness C_n : Ratio of Perimeter square and area of nucleus $C_n = P_n^2 / A_n$, where P_n is the perimeter of nucleus
- major axis of nucleus L_n = length of major axis of an ellipse cover nucleus
- Minor axis of nucleus D_n = length of minor axis of an ellipse cover nucleus
- Nucleus aspect ratio R_n : Ratio of width and height of nucleus region $R_n = W_n / H_n$, Where W_n is the width and H_n is the height of nucleus.
- Nucleus homogeneity $H_n = \sum_{i=1}^K \sum_{j=1}^K P(i, j) |i - j|$, Where $p(i, j)$ is the probability of occurrence of a pair of pixel values (i, j) in the nucleus region computed from grey-level co-occurrence matrix. K represent number of grey levels of an image.

Cytoplasm based features like

- Nucleus to Cytoplasm (N/C) ratio $NC = A_n / A_{cy}$ where A_n is the nucleus area and A_{cy} is the cytoplasm area.
- Entire cell area $A_{en} =$ Total number of pixels in whole cell

For classification the cell into normal or abnormal, they are 3 different classifiers are used, i.e., Artificial Neural Network (ANN), Support Vector Machine (SVM), and Euclidean Distance (ED) based system. In this paper [5], 3-layer back propagation neural network consisting of 9 neurons in the input layer depending on the number of features used 1 hidden layer with 15 neurons 7 neurons in the output layers depending on the number of classes being considered. Multi class SVM with one against rest approach have been considered. For ED based system, Euclidean distance was found out between the feature vectors of the test image and the corresponding feature vectors of the database image.

Result: In this paper [5], proposed a framework for automatic analysis of single cellular pap smear slides. Segmentation phase they used RGVF snake method to segment the cell into 3 regions like cytoplasm, nucleus and background. For the classification phase they used SVM based model which they yielded about an accuracy of 93.78 % and sensitivity of 98.96 % and specificity of 96.69 %.

In this paper [6], the diagnosis system is divided into five parts. The first step is providing input images to the system. In the first step, all slides are dyed with Papanicolou dye passing on and after that scanned on a Nano Zoomer NDP Image framework with 20x resolution. There are 90 slides. They comprise of 61 negative images and 29 positive ones. The second step is pre-processing in which segmented images are magnified and nuclei and cells are extracted from them. The third step is image feature calculation. Image feature calculation is required to distinguish abnormal cells and to judge image positive or negative. The area of each nucleus is determined, and those area with the pixel range between 150-600 pixels are characterized as nucleus enlargement. In the cores of dangerous cells, the measure of chromatin increments, and it is darker than typical. Moreover, if the B component is 90 or less, the region is viewed as dangerous cells. At the point when no less than one malignant cell is identified from inside the slide, then the slide image is made a decision as positive. If not, it is made a decision as negative.

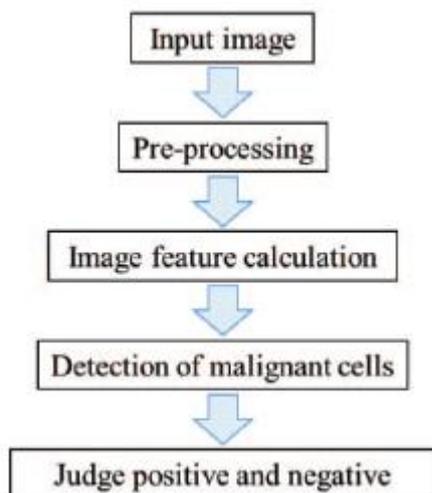


Fig-4: Flowchart of the system.

Result

In this investigation [6], we performed nucleus extraction by utilizing superpixel segmentation, and proposed the technique for identifying threatening cells with nucleus expansion and color density of cell cores. Subsequently, a positive separation rate was 97% and a negative segregation

rate was 55%. FN rate which turns into an issue in screening did not move toward becoming 0%, and it ought to be improved.

In this paper [7], the normal and abnormal cells are classified based on image processing techniques. The process includes adjustment of color histogram with respect to the reference image, detection of nucleus by K-Means Clustering, detection of cytoplasm by creating the edge profile identifying with geometrical pivot, calculating the ratio of nucleus area to cytoplasm area (N/C Ratio), and finally, the abnormal cells are detected by the system.

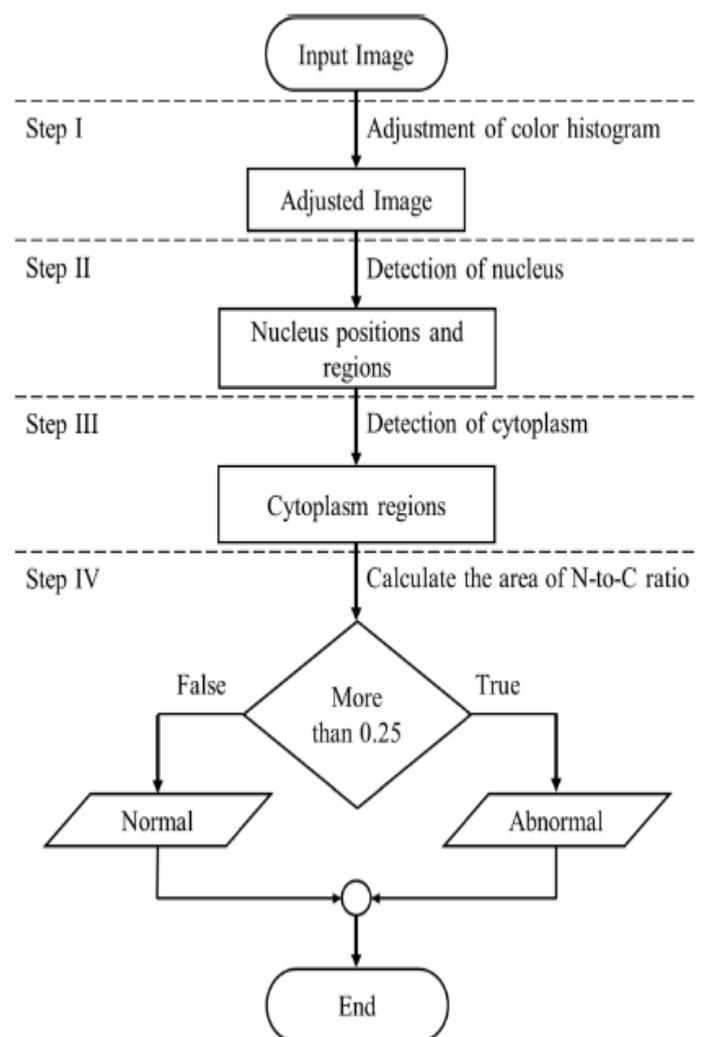


Fig-5: Flowchart of image processing system.

Result

In the issue of factual grouping, a confusion matrix, otherwise called an error matrix is a particular table design that permits visualization of the execution of a calculation [7].

| | | | |
|---------------------|----------|----------------|----------------|
| | | True Condition | |
| | | True | False |
| Predicted Condition | Positive | True positive | False positive |
| | Negative | True negative | False negative |

Fig-6: Confusion Matrix

The importance of every condition is as per the following.

True positive (TP) is an abnormal cell and the calculation will show corrected recognition as a cell. False positive (FP) is an ordinary cell however the calculation will indicate inaccurately discovery as an abnormal cell. True negative (TN) is a typical cell and the calculation will demonstrate effectively location as a normal cell. False negative (FN) is an abnormal cell yet the algorithm will show incorrectly identification as an ordinary cell [7].

The automatic screening of the cervical malignant growth cells has been created by means of morphological image processing techniques [7]. The programming procedure used the calculation of pre-modifying image and discovering core-to-cytoplasm proportion. The procedure incorporated the identification of core by k-means clustering technique and the location of edge profile including geometric revolution relationship [7]. The precision of imaging test was up to 79% for just individual cell count [7]. The affectability was 60% and the particularity is 100% if there should arise an occurrence of without collapsed or covered cells [7]. For applications in therapeutic network, this calculation can be coordinated with whole slide imaging (WSI) microscope. As per the screening of abnormal cells utilizing WSI, the benefit of this calculation would give efficient and minimize human mistakes. Besides, this calculation could be grown further to build the discovery accuracy if there should arise an occurrence of covering cells to improve the execution [7].

3. CONCLUSIONS

Cervical most cancers are screened manually with the aid of the usage of the Pap smear test and LCB check which does not deliver correct classification effects in classifying the normal and uncommon cervical cells inside the cervix region of the uterus. The manually screened technique suffers from excessive faux fee because of human errors and also value effective to be executed by means of the usage of the professional cytologist. In this paper, several methods are proposed for the automatic detection of cervical cancer the usage of image processing techniques. The automatic techniques are achieved to supply correct outcomes and to make effective type of ordinary and atypical cells.

REFERENCES

- [1] N.Sakthi Priya, "Cervical Cancer Screening and Classification Using Acoustic Shadowing", Volume 1, Issue 8, IJIRCCE, October 2013.
- [2] Sreedevi M T, Usha B S, Sandya S, "Papsmear Image based Detection of Cervical Cancer", Volume 45- No.20, IJAC, May 2012.
- [3] SetuGarg, ShabanaUrooj, RituVijay, "Detection of Cervical Cancer by Using Thresholding & Watershed Segmentation",978-9-3805-4416-8/15/\$31.00, IEEE, 2015.
- [4] Nithin S, Poonam Sharma, Vivek M, Preeta Sharan, "Automated Cervical Cancer Detection using Photonic Crystal based Bio-Sensor",978-1-4799-8047-5/15/\$31.00, IEEE, 2015.
- [5] Sajeena T A, Jereesh A S, "Automated Cervical Cancer Detection through RGVF segmentation and SVM Classification",978-1-4673-7309-8/15/\$31.00, IEEE, 2015.
- [6] Ayaka Iwai, Toshiyuki Tanaka, "Automatic Diagnosis Supporting System for Cervical Cancer using Image Processing",978-4-907764-56-2 PR0001/17 ¥400, SICE, 2017.
- [7] Manas Sangworasil, Chayanisa Sukkasem, Suvicha Sasivimolkul, Phitsini Suvarnapaet, Suejit Pechprasarn, Rujirada Thongchoom, Metini Janyasupab, "Automated Screening of Cervical Cancer Cell Images", 978-1-5386-5724-9/17/\$31.00, IEEE, 2018.