Cervical Cancer Analysis Using Image Processing Techniques

Ankit Sharma\textsuperscript{1}, Rimjhim Agarwal\textsuperscript{2}, Saloni Arora\textsuperscript{3}, Nagesh Kumar\textsuperscript{4}

\textsuperscript{1,2,3,4}Electronics and Instrumentation Engineering, Galgotias College Of Engineering And Technology, INDIA

ABSTRACT

The purpose is to develop some vigorous tool to analyse the nature of cancerous cell by using image processing methods so that they can be used in automate diagnostic techniques, reducing the cost and increasing the accuracy of conventional test procedures. The approach taken combines a mix of image processing techniques and methods. Uterine Cervical Cancer is one of the most commonforms of cancer in women worldwide. Papanicolou test abbreviated as Pap test has been used and has decreased the death rate among women. We focus on the cervical borders of the digital colposcopic images so that we have a measurable boundary on the image. When a cell has some abnormalities, it is the nucleus which shows some changes from its normal one and thus it is important to study the nucleus frontier. Further, the Pap test reduces the number of trainers to analyse the samples.

Keywords------ image processing, Uterine cervical cancer.

I. INTRODUCTION

Cervical cancer is one of the most preventable forms of cancer. This is because it is a slow-onset disease whose precursor signs can be detected by inspecting visually, under magnification, samples of cervical cells. The properties such as nuclei to cytoplasm ratio, nucleidiameter, shape are to be studied. The anatomy of the cell lean to turbulent growth because of which nuclei to cytoplasm ratio is higher than normal. Nuclei is circular in shape in general but because of these abnormalities it may be oval, axle or tadpole structure. The screening program is highly victorious in decreasing the death rate. Now the visual examination of cell is a human work so it is highly prone to error. So there exits an automated system as the computer never gets tired and never gives one sided results. The tool developed classifies the cell in three parts: normal, pre-cancerious and infectious. Pap smear is used for the early detection of cancereous and pre cancereous cell before it enters into the protuding stage.

II. LITERATURE REVIEW

The presence of more than one nucleus and the absence of uniformity in image intensity in the cancer images. The determination of these modifications in the nucleus study and density helps in the separation of normal and affected cells in PAP smear images. Medical images have many limitations such as poor quality, noise and human error examination. Digital image processing can help the doctors to a great range. Some ideas for future improvement may be: to design system where a doctor can provide his own grayscale threshold or to automate the process by computer using histogram or fuzzy logic. The images handled are magnified and the computations are done in pixels. So, if a relation among magnification, pixels and actual size is established, the analysis will be more efficient[1].

Cervical cancer is a leading cause of mortality and morbidity, which comprises approximately 12% of all cancers in women worldwide according to World Health Organization (WHO). cervical cancer is a cancer of the cervix which is commonly caused by a virus named Human Papillomavirus (HPV). The virus can damage cells in the cervix, namely, squamous cells and glandular cells that may develop into squamous cell carcinoma (cancer of the squamous cells) and adenocarcinoma (cancer of the glandular cells), respectively…..classification tools (i.e., ANN, SVM, logistic regression, KNN, LDA, and decision tree) generally can achieve good performances to classify the cervical precancerous data. The screening systems based on neural network technique are frequently applied due to the better results and potential of the technique to build a real time system.

One of the most important features of identifying the cancer affected cells is by analysing the N/C ratio.
Manual processes in this field may not always satisfactory as it may be time consuming to visualize millions of cells and also it is error prone. So an automatic analysis may help pathologist to identify cancer affected cells correctly with accuracy. The conclusion is that the Medical images have various limitations such as low quality, presence of noise and human error in interpretation. So this type of automatic analysis may help the pathologist to a great extent. But this system totally depends upon the quality and contrast of the image. Based on the contrast the output may be different for different users. Some idea for further studies may include considering the whole Pap smear image for analysis at the same time instead of taking one cell at a time from that image[2].

A current image processing and analysis extracting the histological image are needed to provide prognostic features are used in this research such as nuclei cytoplasm ratio (NC ratio), nuclei area, nuclei diameter, shape factor of nuclei and pleomorphism. A few studies discuss analysis histology cervical images, although some have shown the grading of pre cancer. Now, reports on the methodologies and outcome of study aimed at developing a robust tool to evaluate and classify histological cervical images into normal, pre cancer and malignant. The conclusion is changes nuclear cytoplasm ratio and morphology features indicates each types of normal, pre cancer and malignant. It is known that higher N/C ratios are indicative of malignancy. In this study, the cervical histology images are classified into three categories: 1) normal 2) pre cancer 3) malignant. The results achieved using N/C feature is very promising and indicate the high discriminatory power of this feature[3].

PAP smear test analysis by visual examinations of images is time consuming, expensive, and tedious. In order to overcome these problems, medical images made the doctors to analyse the internal portions of the body for painless diagnosis. A common shortcoming in the imaging system is unwanted non linearity in the sensor and the display system that produces noises in the images. The conclusion is that after denoising by these filters, PSNR values and MSE values are compared and it is found that Wiener is more efficient than other filters for removing the Poisson noise in the Pap smear images and also it enhances the visual quality of the Pap smear images. Then, the nucleus is segmented from the denoised image with optimal threshold value obtained from the shrinkage algorithms like VisuShrink thresholding, BayesShrink thresholding and SureShrink thresholding. Based on the subjective results of threshold values and MSE it is confirmed that the performance of adaptive wiener filter in combination with Sure Shrink threshold gives the better result than the other adopted shrinkage methods[4].

There are many different types of cell present on a Pap smear slide, including epithelial cells at different stages of maturation, white and red blood cells, and cells from other parts of the reproductive system, together with pathogens (bacteria and fungi) and cellular debris. The vast majority of cells on a slide will be completely normal, but a smear from a woman with cancer may include cancer cells. The task of Pap smear screening appears on the surface to be extremely suitable for automation, and much research is currently underway in this area. The computer system would have the advantages of endless patience and consistent interpretation of images, a particularly desirable trait given the subjective nature of the screening process, and the demonstrated variability between human cytologists[5].

The prognosis of cervical cancer is influenced by local disease extent as determined on the basis of tumor volume, depth and degree of tumor invasion, parametrial invasion, pelvic side wall extension, lymph node involvement, and presence of distant metastases. The current study is aimed at identifying the most influential risk factors among the other factors in order to reduce the number of deaths and creating awareness among women. Accuracy achieved by J48 algorithm is better than any given in the literature. Sensitivity and specificity analysis on these algorithms provided us with the prioritized importance of the prognostic factors that lead to the staging of the cancer. This analysis was not performed in any given in the literature. Data analysis was done using 10-fold cross validation[6].

The normal cells of the cervix first gradually develop pre-cancerous changes that turn into cancer. Doctors use several terms to describe these pre-cancerous changes, including cervical intraepithelial neoplasia (CIN), squamous intraepithelial lesion (SIL), and dysplasia. These changes can be detected by visual inspection after acetic acid application (VIA) on uterine cervix and treated to prevent the development of cancer. Cervical cancers and cervical pre-cancers are classified by how they look under a microscope. The resultant images are useful to gynecologists and oncologists while diagnosing the cancer patient. The work done here is applicable to all kinds of medical images where reflection due to camera flash is present For instance; specular regions are to be discarded when image segmentation is performed so that they are not mistaken for AW lesions. On the other hand, when extracting texture measures, the specular regions have to be not only identified, but completely eliminated. For effective and correct diagnosis it is necessary to segment and find the different types of tissues and organs in the image. The cervix region, which is the main region of interest within the cervigram, is located in the central part of the image. Segmentation is one of the first steps in image analysis. It refers to the process of partitioning a digital image into multiple regions (sets of pixels). Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. We can segment color images by reducing the number of colors in the images, and then...
recognizing contiguous pixels of the same color as a region[7].

An automated method of classifying normal from abnormal tissue is required so the images do not have to immediately interpreted by a trained pathologist. The goal of this research is to exploit differences in cell organization between normal and cancerous tissues using image processing techniques, thus allowing for automated diagnosis, interpreted by a trained pathologist. The goal of this research is to exploit differences in cell organization between normal and cancerous tissues using image processing techniques, thus allowing for automated diagnosis in account to discover the abnormality issues in target images. Additional enhancements were needed to improve contrast between the power spectrum of normal and cancerous tissues (averaging). After these enhancements, clear differences could be seen between the normal and cancerous power spectrum[8].

The hypothesis that morphologic stability of both normal and tumor tissue in vivo requires a uniform level of substrate concentrations through a robust vascular network. Morphologic instability may occur during solid tumor growth and during response to treatment as the result of oxygen, glucose, acid, and drug concentration gradients driving spatially heterogeneous cell proliferation, migration, and death, and reducing cell adhesion and other mechanical forces in hypoxic and acidic regions due to disruption of cell-cell and cell-matrix interactions. This results in invasive fingering and branching and even fragmentation and migration of cell clusters into the surrounding tissue because of differential proliferation along the gradient[9].

III. METHODOLOGY

Cervical cancer and screening:
Cancer Research UK states that “Cervical cancer takes time to develop. There is usually a period when some of the cells lining the cervix develop abnormal changes but are not yet cancerous – these can give rise to cervical cancer later on. Doctors can pick up these changes through screening, and a simple treatment can prevent cancer developing.” A sample of cervical cells is collected and then either smeared and fixated onto a glass slide (smear method) or preserved in a fluid (Liquid Based Cytology or LBC method) and sent to a laboratory. At the laboratory, the samples are stained, usually with the Papanicolaou (‘Pap’) stain. As a result of the staining process, the cells and their major components (cytoplasm, nucleus) are made visible. The slides go through a strict screening process, whose aims are (1) to detect any abnormal cell changes, (2) to assess the type and severity of abnormal cell change, when it is observed and (3) to report the presence of a number of infectious agents, when detected. Cervical cancer is associated with Human Papilloma Virus (HPV) infection. Only a minority of HPV infections will subsequently result in cancer; however nearly all women who develop cervical cancer have had past infection with a high-risk strain of HPV. A vaccine to prevent HPV infection (and hence cervical cancer) has now been licensed for use within the European Union. The UK Government is in the process of launching a program of mass vaccination.

Image processing:
We analyze digital images of cervical sample slides to extract the image properties that will drive the system. Features which may be indicative of abnormality include:
• Size: In particular an increased relative size of the nucleus compared to the cytoplasm.
• Shape: Normal nuclei tend to have a smooth, circular or oval outline.
• Texture: Abnormal nuclei may have rough textures (due to chromatin clumping).
• Chromaticity: Abnormal nuclei are often darker (hyperchromatic) and sometimes take a reddish color instead of their usual blue/green.

Image Segmentation:
A number of choices were made regarding the desired outcome of the segmentation process.
• It will split the image into one background region and other regions which delineate the cells’ nuclei.
• It is allowed to miss very bland nuclei (very likely normal) and out-of-focus nuclei (scanning artifact).
• The nuclear regions can be over-segmented (cut up into several pieces) because our system will analyze the texture features of groups of connected regions twice: once considering the group as a single region and once considering each region separately.

The overall scheme for the segmentation process was as follows:
• Perform an initial segmentation of the image where each region contains either nuclear material or background material (including cytoplasm). The regions can be much smaller than a typical nucleus.
• Discard, as far as possible, the background regions.
• Deal with the remaining regions by inspecting groups of connected regions with the aim of discarding remaining background regions and fusing nuclear regions into regions representing, as far as possible, a single, whole, nucleus.
IV. STEPS FOLLOWED

In this procedure, some digital image processing techniques such as image conversion, morphological operations, criteria selection, image analysis and image classification have been applied. First, a cervical cells image was loaded. This was then converted to grayscale and to binary image. Boundary of all ‘elements’ in the image has been detected and marked using green color line. ‘Elements’ here are referred to all features appeared in the Pap smear slide including cervical cells, blood cells, leukocytes and other structures. The next step was the measurement of all of the elements’ properties such as color intensity level and area. This was followed by a criteria selection. Criteria selection refers to the criterion that had been chosen to distinguish between cervical cell and other cells. The criterion used was cells’ area and cells’ color intensity distribution. This allows only cervical cells to appear in image and other structures will be eliminated. Only cervical cells are left in the image after this step is completed.

The algorithm has two stages, first is pre-processing of given MRI image and after that segmentation and then perform morphological operations. Steps of algorithm are as following:

1. Give MRI image as input.
2. Convert it to grey scale image.
3. Apply high pass filter for noise removal.
4. Apply median filter to enhance the quality of image.
5. Computer segmentation and morphological operation.
6. Finally output will be a tumoured region.

V. RESULT AND CONCLUSION

The pre-processing step excludes all the background and leaves for further processing the parts of the image which contain isolated cells or cell clusters and as such is a fast process which results in the reduction of the region of interest in the image. The method has been applied in several PAP smear images defined by an expert observer. The step for the detection of the cell nucleus centroid has exposed that the resulted points of the image indicate the area of the nuclei, as it is confirmed by the expert observer.

The comparative study of multiple classifiers identifying the stage of cervical cancer using a dataset of size 221 records provided us with an insight into the predictive ability of different data mining methods. Accuracy achieved by J48 algorithm is better than any given in the literature. Sensitivity and specificity analysis on these algorithms provided us with the prioritized importance of the prognostic factors that lead to the staging of the cancer. This analysis was not performed in any given in the literature. Data analysis was done using 10-fold cross validation. We can conclude saying that by applying data mining algorithms the invaluable efforts of the medical professionals can be enhanced to save more human lives by giving proper treatment at the right time.

REFERENCES