Chapter I

Introduction
1.1 Overview

Cervical cancer is the third most diagnosed cancer and the fourth leading cause of cancer death among women worldwide accounting for 9% of all malignancies among females in 2008 [1]. The World Health Organization (WHO) promotes many programs for the early detection of cervical cancer worldwide. In developing countries due to the weak health systems and untrained pathologists, it is difficult to make this program in success. The mortality rate of cervical cancer in India is high next to the breast cancer (Figure 1.1). Also in South Asia, the numbers of deaths of cervical cancer by age group in India is the highest (Figure 1.2). There has been a regular campaign against cervical cancer for the last 30 years in India, but this had little impact on the morbidity and mortality from the disease, with India ranking fourth worldwide [2]. Papanicolaou test (Pap-smear) was the first mass screening test introduced during 1950’s which aims at the early detection of cervical cancer before it reaches the invasive stage. But due to inadequate sample collection, errors in screening and interpretation of smears, a high false negative rates ranging from 13 to 70% have been reported from this screening test [3]. Also overlapping of cells and backgrounds, limited computing power and lack of conceptual understanding of morphologic abnormalities creates these systems in failure. Countries with well-organized programs to detect and treat precancerous abnormalities and early stage of the cervical cancer can prevent more percentage of the cancers. The automated system is in need which focuses on screening of large number of samples which reduce biased decisions and to improve the accuracy of results.

Due to the collaborative works of experts from the fields of biology and medicine, medical image analysis has produced various important research results. Even though the primary role of image analysis is to start with the quality images, it extends its work towards acquiring any kind of images and make them suitable for further processing. One of the basic
problems of research over last few decades in cytology image analysis was identifying set of descriptors in classifying cytological samples. However huge advances in medical imaging and computing power have given tremendous growth in accuracy in automated cytology systems. The quantitative analysis of cell nucleus and cytoplasm’s texture has shown the most promising areas in the past and continues to be the challenging work.

In medical image analysis, most of the color and shape features are not suitable for processing cervical cancer cytology images. In detection and classification of edges, corners and junctions color features will takes the important part. Since the cervical cancer cytology image samples do not have all those edges and corners, color feature is not suitable. Further depending on the color system, one or more histograms are employed to quantify the color distribution, defined by the number of bins used. Differences in color distribution are, sometimes, essential to determine differences between images. However such distribution can lead to errors when different images present similar histograms. The shape of an object has its apparent boundary, outline or external surface. In the image domain extracting shape consists in the identification of lines and curves. Due to the factors like the presence of blood, inflammatory cells or thick cell clumps in the cervical cancer image make shape features may still fail to reflect the differences between normal and cancerous cells.

The aim of this thesis is to present a detailed analysis of various unique texture features which could help in identifying abnormalities and classification of various stages of Pap smear images. The essence of this analysis paved the way to categorize the various texture features or to rank the features which are most suited for different stages of cervical cancer classification.

1.2 Cervical Cancer Screening Methods

The aim of the screening cervical cancer is to reduce the number of incidence and mortality rates by detecting and treating precancerous lesions, and the ultimate proof of
success is to do this in a cost-effective manner. This is done by encouraging women to have regular cervical cancer screening so that conditions that might otherwise develop into invasive cervical cancer can be identified and treated. The screening results should have adequate sensitivity and specificity for detection of precancerous lesions, yield reproducible results, is cheap, simple and easy to apply, be without side effects or complications, be as painless as possible and be socio-culturally acceptable. It is normally recommended that cervical cancer screening should begin three years after vaginal intercourse is initiated and no later than the age of 21 [4].

1.2.1 Conventional Cytology: Cervical (Pap) Smear

The Pap smear test or Papanicolaou test is the most common form of diagnosis for detecting cervical cancer in its early stages is a procedure called a Papanicolaou test or Pap smear. In this procedure, cervix cells are collected and view under the microscope to find whether the cells are affected with cancer or not. Normally in conventional cytology, the collected cells are spread directly onto a glass microscope slide and spraying it with a preservative. Instead of putting them into the slide, the collected samples are put into the preservative liquid are called liquid based cytology. Women who are 18 or older or who are sexually active are recommended to undergo annual Pap smear tests [5].

By this method, it has been proved by the regular screening procedures the mortality rate came down to 60% approximately in women aged 30 and older. Even though this test is more successful, it has certain limitations. Cells on the slides can be spread too thickly making them difficult to see clearly. Since the results need to be examined by the humans, an accurate analysis of hundreds of cells in each sample is not always possible.

1.2.2 Pelvic Examination

A pelvic examination is an important method of detecting cervical cancer, in which the doctor will manually examine the abdomen and pelvic area for any nodules or bumps,
which are explored in greater detail with imaging technology. During this examination, doctor manually examines the vagina, cervix, uterus, fallopian tubes, ovaries and rectum. Two types of benefits may be predictable: the detection of cervical cancers which produce false-negative smears, and the detection of other gynecologic pathology. A pelvic exam alone will not help to find out the abnormal cells of the cervix or cervical cancer at an early stage; it can be done along with Pap smear test.

1.2.3 Liquid based Cytology

Liquid based cytology (LBC) technology has been in use since 1996 in the USA and parts of Europe. It is the way of preparing cervical samples for examination in the laboratory. The samples are collected using a special device (spatula) which brushes cells from the neck of the womb. From the head of the spatula, the cells are lodged, is broken off into a small glass vial containing preservative fluid, or rinsed directly into the preservative fluid. In the laboratory it is spin, remove the unclear material and the remaining cells are taken onto the slide. The slide is examined in the usual way under a microscope by a cytologist. LBC is more generally preferred by smear takers and laboratory staff who read the smears. There is no need for smear takers to spray/’fix’ cells on a glass slide. This method has been developed to overcome the problem of ‘unsatisfactory’ Pap tests, where microscopic examination of the cervical cell smear is difficult and sometimes impossible due to the presence of blood or mucus.

SurePath™ and ThinPrep® are the two major systems which use LBC technologies. In SurePath system, a sampling device is used to take a sample of cervical cells. The head of the device is put into a vial of fixative liquid and it is then processed in a SurePath machine to free cells from the sampling device. A sample of the fluid is centrifuged (spun) and the cells then allowed settling onto a glass slide. In ThinPrep system, A sampling device (cytobrush or cervibroom or combo) is used to take a sample of cervical cells. The device is rinsed into a
vial of fixative fluid that is then processed in a ThinPrep machine where it is mixed and filtered to deposit a sample of cells onto a glass slide.

The ThinPrep Imager identifies 22 areas on a slide that are most likely to contain abnormal cells.

1.2.4 Colposcopy

Colposcopy is a gynecological procedure that illuminates and magnifies the uterine cervix in order to detect and examine abnormalities of these structures. A colposcope is a low-power, stereoscopic, binocular field microscope with a powerful light source used for magnified visual examination of the uterine cervix to help in the diagnosis of cervical neoplasia. The clinical use of colposcopy for the evaluation of cervical cytologic abnormalities allows the identification and successful management of most premalignant cervical lesions. It is a very safe procedure where gynecologist get close look at the cervix for analysis. The expert inset a instrument called speculum into the cervix during this procedure. It is not a painful one but it creates a discomfort to the patient. Most of the women continue the daily routine work after the test has been done. The powerful light source with a microscope is used to detect any abnormalities and any detected tissue spots are removed and send for further investigation.

Hinselmann (1925) first described the basic colposcopic equipment and its use, establishing the foundation for the practice of colposcopy. The key ingredients of colposcopic practice are the examination of the features of the cervical epithelium after application of saline, 3-5% dilute acetic acid and Lugol’s iodine solution in successive steps. It helps in coagulating and clearing the mucus. Acetic acid is thought to cause swelling of the epithelial tissue, columnar and any abnormal squamous epithelial areas in particular. Special tests are done during colposcopy, including acetic acid wash, use of color filters, and sampling (biopsy) of tissues. Cervical abnormalities include pre-cancer (dysplasia), which can be rated
as mild, moderate, or severe, and cancer. The type of treatment procedure chosen by the physician depends on the severity of the cervical abnormality, which is determined by analysis. The colposcopic diagnosis of cervical neoplasia depends on the recognition of four main features: intensity (colour tone) of acetowhiteness, margins and surface contour of acetowhite areas, vascular features and colour changes after iodine application. The major drawback of colposcopy is usually the lack of experience with a large enough patient population to become familiar with changes in premalignant and malignant conditions [6].

1.2.5 HPV DNA Testing

The most cause of cervical cancer is due to the infection with Human Papilloma Virus (HPV). HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 59 and 68 are known to be frequently associated with HSIL and invasive cancers of the cervix [7]. Thus it is necessary to test the presence of HPV in cervical cells which could be incorporated in cervical cancer screening programs. Using a sample of cells, this test specifically identifies HPV 16 and HPV 18 and concurrently detecting 12 other types high-risk HPVs.

Number of techniques available for HPV testing and Southern Blot hybridization is regarded as a gold standard. This test can be done in two situations: First, It can be combined with Pap test to screen for cervical cancer. Second, The HPV DNA test can also be used in women who have slightly abnormal Pap test results. The specimen for HPV-DNA testing can be obtained in two ways, either by using a cell suspension from liquid based cytology or by using the endo-cervical cytobrush [8].

1.2.6 Polar Probe

The probe technology is based upon the fact that the tissue impedance to electrical stimulation differs between normal and abnormal tissues. Investigators have tried to utilize spectral and electrical stimulation of the cervical tissues as an adjunct to conventional Pap smear testing. The innovative technology utilized in the design of the Polar probe includes
optical elastic backscattering techniques; electrical measurements; and expert system classification of tissue. This instrument is operated with three light emitting diodes (LEDs) each operate at a different optical frequency in the red, green and infrared sections of the spectrum respectively.

A low level electrical pulses and optical signals are passed through the cervical tissues during the scanning of polar probe. The responses are compared with the already stored real time databank of cervical tissue types. Based on the matched result, they are classified as normal, low grade abnormality and high grade abnormality cells. In USA and Europe, Polar probe is used as the secondary screening tool in a population of Pap smear positive women and it is considered as the adjunctive test to enhance the accuracy of screening information.

1.2.7 Laser Induced Fluorescence

An abnormal pap smear followed by colposcopy produce only limited predicted value even the diagnosis has done with the experienced people. For accurate diagnosis of CIN and to improve the predictive values, biopsy and histologic analysis are needed. The optical technique by which the predictive value of colposcopy can be improved is called Laser Induced Fluorescence method. Optical methods of detection of cervical cancer are considered as the standalone technique since this method can be automated and faster. Several groups [9], [10], have studied fluorescence spectra of normal and malignant tissue to develop an optical pathology method for early detection of cancer. Fluorescence emission spectra of cervical tissue at multiple excitation wavelengths contain information that can enhance the diagnostic content of tissue spectra at a single excitation wavelength [11]. Based on chemical and morphological composition of tissues, there is varying fluorescence spectral characteristics from the tissues. Based on the spectroscopic difference, the tissues can be divided as normal or abnormal [12].
1.2.8 Visual Inspection with Acetic Acid (VIA)

Visual inspection of cervix, after application of 5% acetic acid (VIA) and/or of Lugol’s iodine, provides simple tests for the early detection of cervical precancerous lesions and early invasive cancer. Such procedures eliminate the need for laboratories and transport of specimens, require very little equipment and provide women with immediate test results. VIA has emerged as a promising, cost effective, non-cytology based “see and treat” alternative for economically unprivileged geographic reasons [13]. This has been demonstrated in various studies where trained physicians and mid-level providers correctly identified between 45% and 79% of women at high risk of developing cervical cancer [14]. VIA can offer significant advantages over Pap in low resource settings, particularly in terms of increased screening coverage, improved follow-up care and overall program quality. VIA has shown to have in several studies a low specificity compared to cytology and a high rate of false positives. Entities such as inflammation, cervical condyloma and leukoplakia can give false positive results of VIA test [15].

VIA has demonstrated high sensitivity for detecting CIN and cervical cancer, but it is limited by low specificity. VIA has the advantage of requiring only low-technology equipment and the result is available within a couple of minutes.

1.2.9 Visual Inspection with Lugol’s Iodine (VILI)

VILI test is otherwise called Schiller’s test, which involves in temporarily staining of cervix with Lugol’s iodine. VILI is a simple screening test which is based on ability of trained health care personnel to detect yellow, non iodine uptake areas in the cervical transformation zone [16]. The aim of this test is to determine whether the test result is positive or negative for possible precancerous lesions or cancer. Since the normal cells contain glycogen and iodine is glycophilic, the normal cells changes its color to mahogany-
brown. On the other hand due to the lack of glycogen, abnormal cells (cancerous cells) do not absorb iodine and appear as yellow.

The advantages of VILI test are, it involves high sensitivity results in a low proportion of false negatives, and most of the test results are available immediately. Moderate specificity may result in over-referral and over-treatment in a single-visit approach is the main drawback of this system. Further, there is a need of properly designed studies on VILI are essential to evaluate the effectiveness in reducing cervical cancer incidence and mortality.

1.2.10 Cervicography

Cervicography is a technique of taking photographs of the cervix using a 35mm special camera. Before doing this procedure, the cervix is swabbed with the application of 5% acetic acid to identify and visualize the tissue changes in the cervix. The photographs are developed on 2x2 meter screen and analyzed by the experts of colposcopy. Cervicography is considered as a static version of a colposcopy, and the image is referred as cervigram photographic cervical slide. It is considered as a laboratory procedure and is performed in clinical settings only. The results of cervicography are reported as positive, negative, atypical, or defective. This is recommended where there is a need of improved sensitivity and to clear the uncertainty in low grade lesions on Pap smear analysis.

1.2.11 Speculoscopy

Speculoscopy, or magnified chemiluminescent examination (MCE), is a new visual method for the detection of cervical neoplasia. It is a procedure by which a special blue-white light (Speculite) is used to examine the cervix for cancerous or pre-cancerous lesions. This test can be performed complementary to pap smear test and a negative speculoscopy along with negative pap smear provides greater assurance of absence of diseases. It was developed in the year 1988, and recently approved by the U.S. Food and Drug Administration (FDA). The combined results of pap smear with speculoscopy in the screening examination
designated PapSure, can allow for the identification of more women who are appropriate referrals for colposcopy or close follow up than pap smear alone.

1.2.12 Cervical biopsy

A cervical biopsy is a surgical procedure in which a small amount of tissue is removed from the cervix to test the abnormality. This can be suggested after the abnormality can be detected during the pap smear or pelvic examination. To identify the abnormalities, Schiller test has been done which is nothing but washing of cervix with a solution of vinegar and water, and may be swabbed with iodine. It is not a painful process, but the cleaning process may burn a bit. There are three different methods in this category: punch biopsy, cone biopsy, and endo-cervical curettage. Pre-cancerous changes in a biopsy are called cervical intraepithelial neoplasia (CIN). It is graded on the scale of 1 to 3 depends on the depth of abnormality viewed under the microscope. Based on this biopsy report, the doctor advised for any two types of general treatments. The first type is the destruction of abnormal area and the second type is the removal of affected area.

1.3 Automated Methods and Techniques of Cervical Screening

Automation assisted screening methods are developed to: increase sensitivity and specificity of pap-smear screening, decrease the workload of cyto-technicians and cyto-pathologists, decrease the cost of the screening programmes, and to reduce the incidence and mortality rates of cervical cancer. This is essential for reducing the false positives and false negatives that invariably result from inter- and intra- observer variation [17]. Pap smear screening test is the most effective cervical cancer prevention measure developed so far. The visual examination of the smears is the time consuming and expensive process and the necessity for automation is highly required. After the acceptance of Pap smear test as the golden standard, proposals for automating the screening through image processing techniques were generally accepted [18]. The only aim of the system is focusing the automated screening has to be done at lower cost and high accuracy.
1.3.1 First Generation Systems

Automated screening devices have been developed from the beginning of 1990’s and they have been commercially available from mid 90’s. The Cytoanalyzer project in the US was the first automated screening device developed for PAP-smears images [19]. In this system, based on nuclear size and optical density, the cells were classified as normal or abnormal. The System has the automated slide feed and the circuits for automatic focusing. The video processing circuits produces two-dimensional histograms of nuclear size versus nuclear optical density. By this method, with the spatial resolution of 5 micrometer it was able to detect the difference in size between normal and abnormal cell and this system was the first uses fully automated microscope. Unfortunately, tests with the Cytoanalyzer revealed that the special purpose fixed logic pattern recognition produced too many false alarms on the cell level [20]. This project failed in early sixties since of many reasons which includes: (1) the system treats both malignant and normal cells as same, (2) contents of clumps of blood cells, strands of tissue and mucus and overlapping epithelial cells made the poor analysis and (3) samples, including the normal ones, was thus found to be suspicious for abnormality. After the failure of Cytoanalyzer, the automation attempts were made by Europe and Japan. In late sixties, one parameter automatic screening device was developed and gets failed [21]. The system CYBEST, was developed in Japan by Watanabe and Toshiba and the first version used special-purpose electronic circuits and later versions used digital computers [22]. The device used four different features from the cell images: nuclear area, nuclear density, cytoplasmic area, and nuclear/cytoplasmic ratio. They also realized that nuclear shape and chromatin pattern were useful parameters but were not able to reliably measure these features automatically mainly because the automatic focusing was unable to reliably produce images with all the cell nuclei in sufficiently good focus. The chromatin pattern measure that was proposed by this group was the number of blobs within the nuclear region. Four generations
of prototype systems were developed over a 15-year period. The last one used strobe illumination and nonstop scanning motion to reach high scanning speeds. The prototypes were used in large field trials in the Japanese screening program and showed promising results but none of them became a product [23].

1.3.2 New Generations Systems

During early 1980’s new generation of system such as BioPEPR [24], FAZYTAN [25], Cerviscan [26], LEYTAS [27], and Diascanner [28] were developed with the improved image segmentation, feature extraction, and image classification algorithms. The better focus and high resolution were the two important requirements of the above systems. In Diascanner, this can be achieved by dual resolution approach in which an initial low resolution search scan followed by high resolution scans. All the above systems reached an operational prototype stage in the mid eighties and none reached the market due to low processing power, lack of automated microscope, and high cost. The improvements in computer display technology create in making the interactive systems and led to possibly of developing new automated systems. During this time the concept of “pre-screening” system, in which large fraction of specimen would be able to classify without the human intervention. Due to sufficient computer memory capacity and advancements in computer displays during late eighties, it is easy to save any kind of cell images and good enough for the human to judge whether the image could be normal or malignant one. The PAPNET system from Neuro-medical Systems was the first to introduce interaction into automated screening [29].

Initially it used the low resolution search, and then high resolution fields were processed first by algorithmic classifier and then by a neural network classifier. As an output 64 most abnormal cells were stored on the magnetic tapes which later become the review station. There the decision whether the specimen should be classified as normal or suspicious
was taken. For the abnormal cases, cyto-pathologist would do the final analysis and make the decision whether the woman should be called for follow-up or not.

For economical and legal reasons, the automation of cytology images creates the great impact during the late eighties in United States. During this time, the new technique for preparing the samples called ThinPrep was introduced which was based on liquid cytology [30]. At the same period of time, similar sample preparation method called PREP developed by AutoCyte was introduced. During this period of time, a conventional and neural network classifier called AutoPap 300 from NeoPath was used in automated screening process [31]. This system used the strobe illumination mechanism, which uses two resolution levels, low resolution which is done initially and high resolution done on most interesting parts of the specimen. The mathematical morphology operations were used on most of the processing, and as a result 68 features are extracted from the image and send to the classifiers. However the final decision requires the visual inspection of the specimen level.

1.3.3 Commercially Available Screening Devices

During the 1990’s, Food and Drug Administration (FDA) has put tough approval processes to sell the screening devices in the american market. Huge money has been spent on developments and field trials and there was a shakeout; the companies merged and were acquired by larger companies. In 1988, Tripath was the first company received the FDA approval for its screening product BD FocalPoint Slide Profiler [32]. Another specimen preparation device called SurePath was added into this category as it received the approval from FDA. The system can also be used for quality control and claims increased sensitivity in detecting abnormalities [33]. Cytyc was quite successful with their improved liquid based preparation technique and could demonstrate better performance for that technique as compared to conventional smears. In 2003, it has received the a FDA approval for ThinPrep imaging System and in 2007, it become the part of Hologic Company [34].
1.4 Screening- Technical Challenges

Even though many cervical cytology screening programmes have been introduced in low- and middle-income developing countries, very limited success rate have been recorded. Only 5% of women in developing countries undergo cervical screening compared with 40-50% in the developed world. In developing countries, because of the lack of trained cytotechnologists and cytology laboratories, there is often a long interval (1-3 months) between the Pap screenings and when the test result is available. A huge amount of money is invested in providing the infrastructure, inspection for cervical screening programmes and screening devices. The effective screening program will have to meet the technical challenges to improve cervical screening in developing countries.

Nowadays, identification of both normal and abnormal cervical cancer cells can be missed due to technical errors. Many studies have shown that some Pap smear cytology images are blurred and highly affected by unwanted noises, such as blood, air artifacts’, vagina discharge etc. In Pap smear imaging, starting from image preparation to identification of normal and abnormal cells, there are number of technical challenges are in front of us to solve; else the Pap smear cytology image will be referred as inadequate samples for cervical cancer screening process.

1.4.1 Specimen Preparation

During the preparation of Pap smear, the cell collected from the cervix regions are manually spread over the glass slide. It is important to note that during the specimen acquisition process, cells are obtained from the transformation zone. The samples either stained manually or by using machines with varying degrees of complexity. There is a big variation in specimen quality due to manual smearing and staining. Sometimes the cellular material may be unevenly distributed leading to dense clumps which light cannot penetrate while other parts of the slides may be empty [35]. Even the smear preparation is done with
care; there are some regions of denser and more overlapping. Only experts can diagnose these variations, but it is very tough to analyse these uncertainty using automatic methods.

The solution on this, many liquid based cytology (LBC) preparation methodologies was developed. The collected cells from the cervix regions are submerged into some liquids before it is set down into the glass slide, fixed and stained. By this procedure, the samples are easier interpreted to any forms and it can adapt to any machines. Over the years, two leading techniques SurePath and ThinPrep were developed based on liquid based cytology.

The more operational costs and materials cost are the disadvantage of the liquid based cytology methods. This causes significant economic problems in regions with limited resources. There are many alternatives to liquid based preparations were proposed which are competing with lower costs [36].

1.4.2 Scanning

In order to extract the image features, the image should be scanned with high resolution. A normal smear with size of 25x50 mm with the pixel size of 0.2 microns, will occupy 31 billion pixels. Even with the high transferring techniques, it will take more time in transferring the whole data from camera to computer. A high resolution microscope lens gives a field of view (FOV) with a diameter of around 0,5mm and with a matching 6 megapixel sensor we will get 5000 image fields. It is very hard to reposition and capture the image which can be reduced by nonstop motion and flash illumination to freeze images. This idea was used in two most popular scanning system; CYBEST4 [37] and AutoPap [38]. In the similar way Cerviscan[39], Diascanner [40] and Aperio [41] used the methodology of using a 1D sensor with the length of 2000 pixels and the movement of microscope in the orthogonal direction.

Extraction of features from the cell image is the very important step in Pap smear screening. This process requires very good focussing effect, requires high quality autofocus
instruments and process in less time. One way to increase the focussing effect is to use only the smaller part of the slide for the specimen. Another approach called dual resolution approach in which minimizing the resolution between 10x and 40x lenses. The only drawback of this approach is that if the low resolution scan type missed any abnormalities, it will not be even seen in high resolution analysis.

1.4.3 Segmenting Cells and Nuclei

Image segmentation is the crucial step in medical image processing system. This process is nothing but the delineation of each cell or cell nucleus from the specimen image. Due to the smear thickness, staining intensity and obscuring elements, it is very difficult to do the segmentation process effectively. The earliest system used threshold based gray scale for the segmentation procedure. Prewitt and Mendelsohn [42], proposed a concept of fixed threshold system was used on later systems.

Recently more advance concepts were used. Bergmeir et al. [43] use mean shift and morphological filtering and later try Canny edge detection followed by the randomized Hough transform [44]. Bamford and Lovell [45] use a dual active contour algorithm. In recent reviews five different classes of approaches to cell segmentation are identified and it is demonstrated how they have appeared and gained popularity over the years.

None of the methods are reliable as expected and the need of developing new methods is still necessary. The identification of nucleus and cytoplasm borders are still a toughest job. The main requirements of segmentation algorithms are accurately detect and segment the cell nucleus under different staining conditions. The second main requirement of segmentation is the time complexity; the whole segmentation process must be completed in milliseconds of time. The possibilities of hardware parallelism help to improve the performance much better.
1.4.4 Artifact Rejection

The main aim of the segmentation is to accurately split both cell nuclei and cytoplasm. Many of the segmentation methods get failed in accuracy because most of the cell images are corrupted by overlapping cells and other artifacts. This may cause due to poor preservation of cells or due to the wrong outline of the object. The other objective is to detect artifacts and discard the corresponding image from the database; else it will lead to unreliable classification results. This process of analyzing the segmentation results in order to remove erroneous results is called artifact rejection.

Artifact rejection is not a simple task because there are many ways by which the artifacts influences the cells; inflammatory cells, folded and distorted cells, overlapping objects, mucus and staining mistakes. But without this step, the outputs would produce undesirable results. In the medical image diagnosis, even if the results produce 1% false positive result, the system will be treated as useless one. One solution to this problem is to make the classifier highly asymmetrical between false positive and false negative. This is acceptable when the false positive rate is virtually zero, less than 0.001%. Technology must be improved to create such a classifier for accurate segmentation and to extract the features.

There are many research work done on artifact rejection for cervical screening. Most of the work ignores the automatic artifacts rejection and working on visually selected or verified images of nuclei. Some other research papers include the problem of rejection during the segmentation or classification steps which could not worked out for a real automated screening system.

Malm et al. [46], recently presented such a study where they demonstrated a specificity of 99.38% on smears and 99.83% on LBC specimens, while maintaining a sensitivity of around 98% based on a material of around 12,000 automatically detected and
segmented images of objects visually classified into cell nuclei and artifacts. With that kind of performance we would still have a few hundred artifacts corrupting the data if we analyze 100,000 objects, so it may be hard to achieve the sensitivity of detecting a few abnormal cells without getting too many false positive samples. Still it points in the direction of what is necessary to achieve for a useful system.

1.4.5 Feature Extraction

The extraction of unique features from the images such as color, size, shape and texture helps in accurate segmentation. The normal cells have the regular features while the malignant cells have the irregular structures. The healthy images can focus all the features even it can be acquired in low resolution. In the first generation systems, due to the lack of perfect artifact rejection principles, the feature extraction stage got many failures. In recent years [47], different kinds of features extraction methods are introduced and they are systematically classified.

By knowing the texture of nucleus and chromatin pattern, it is easy to identify whether the nucleus is normal or malignant. Also by the distribution of DNA inside the nucleus, it is able to identify the malignant cells. On the other hand, measuring the chromatin distribution is difficult in Pap smear images. The originally proposed methods were based on a statistical description of neighboring grey levels typically measured through so-called transition probability matrices [48]. In graph analysis method [49], the individual chromatin graduals are segmented as objects, the spatial relations between these objects are described. A very important aspect of the chromatin analysis is that the need of perfect focus and very high quality images to reliably represent this pattern which is at or beyond the optical resolution limit.
1.5 Burden of Cervical Cancer

According to the world health organization cervical cancer is said to be the world’s second deadly cancer with an estimate of about 493,243 women diagnosed with it and 273,505 dying from it per year. India has a population of 432.20 million women aged 15 years and older who are at risk of developing cervical cancer. Current estimates indicate that every year 122844 women are diagnosed with cervical cancer and 67477 die from the disease. Cervical cancer in India ranks as the 2\textsuperscript{nd} most frequent cancer among women and the 2\textsuperscript{nd} most frequent cancer among women between 15 and 44 years of age. Based on India studies performing HPV detection tests in cervical samples, about 7.9\% of women in the general population are estimated to harbor cervical HPV infection at a given time, and 84.1\% of invasive cervical cancers are attributed to HPVs 16 or 18.

1.5.1 Incidence of Cervical Cancer in India

Globally, in many economically developing regions of the world, including Africa, Asia, South and Central America, and parts of the Pacific region, cervical cancer is the single most common cancer among women with an age standardized rate (ASR) ranging from 30 per 100,000 women in parts of Eastern and Western Africa to 24.6 per 100,000 women in South-Central Asia [50].

Current data suggest that out of the total global burden, 21 percent of all cervical cancer registered are from India (Table 1.1) and Age standardized Incidence rate (ASR) of India on cervical cancer is 22.0 (Table 1.2) which is even higher than the global record. In many economically developing countries, cervical cancer represents 13 percent of all females cancers compared to less than six percent in other regions of the world. Among the cervical cancer registered in India, top metros cities rank in the top list. During 1993-1996, New Delhi ranks top with the record of 4280 new cases, Mumbai recorded 3338 new cases; Chennai
recorded 2145 cases and Pune with the record of 1035. Among the mortality rate, India ranks first in South Asia (Table 1.3).

Table 1.1 Incidence of cervical cancer in India (estimations for 2012)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>India</th>
<th>South Asia</th>
<th>World</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual number of new cancer cases</td>
<td>122,844</td>
<td>145,946</td>
<td>527,624</td>
</tr>
<tr>
<td>Crude incidence rate(^a)</td>
<td>20.2</td>
<td>17.1</td>
<td>15.1</td>
</tr>
<tr>
<td>Age-standardized incidence rate(^b)</td>
<td>22.0</td>
<td>19.3</td>
<td>14.0</td>
</tr>
<tr>
<td>Cumulative risk (%) at 75 years old(^b)</td>
<td>2.4</td>
<td>2.1</td>
<td>1.4</td>
</tr>
</tbody>
</table>

\(^a\)Rates per 100,000 women per year.
\(^b\)Cumulative risk (incidence) is the probability or risk of individuals getting from the disease during ages 0-74 years


Table 1.2 Incidence of cervical cancer by cancer registry in India (observed cases during the specified period)

<table>
<thead>
<tr>
<th>Cancer Registry</th>
<th>Period</th>
<th>N-cases(^a)</th>
<th>Crude Rate(^b)</th>
<th>ASR(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmedabad(^1)</td>
<td>1993-1997</td>
<td>799</td>
<td>9.1</td>
<td>13.4</td>
</tr>
<tr>
<td>Bangalore(^2)</td>
<td>2005-2007</td>
<td>1,541</td>
<td>15.9</td>
<td>20.6</td>
</tr>
<tr>
<td>Barshi(^3)</td>
<td>1988-1992</td>
<td>252</td>
<td>23.4</td>
<td>27.4</td>
</tr>
<tr>
<td>Paranda and Bhum(^2)</td>
<td>2003-2007</td>
<td>214</td>
<td>18.7</td>
<td></td>
</tr>
<tr>
<td>Bhopal(^2)</td>
<td>2004-2007</td>
<td>426</td>
<td>13.5</td>
<td>19.5</td>
</tr>
<tr>
<td>Chennai(^2)</td>
<td>2003-2007</td>
<td>2,145</td>
<td>19.4</td>
<td>21.4</td>
</tr>
<tr>
<td>Delhi(^1)</td>
<td>1993-1996</td>
<td>2,983</td>
<td>16.7</td>
<td>25.8</td>
</tr>
<tr>
<td>Dindigul, Ambillikai(^2)</td>
<td>2003-2007</td>
<td>1,215</td>
<td>24.1</td>
<td>24.5</td>
</tr>
<tr>
<td>Karunagappally(^2)</td>
<td>2003-2007</td>
<td>108</td>
<td>9.9</td>
<td>8.9</td>
</tr>
<tr>
<td>Mizoram(^2)</td>
<td>2003-2007</td>
<td>364</td>
<td>15.4</td>
<td>19.7</td>
</tr>
<tr>
<td>Mumbai(^2)</td>
<td>2003-2007</td>
<td>3,388</td>
<td>11.7</td>
<td>13.5</td>
</tr>
<tr>
<td>Nagpur(^4)</td>
<td>1998-2002</td>
<td>741</td>
<td>15.2</td>
<td>18.4</td>
</tr>
<tr>
<td>New Delhi</td>
<td>2003-2007</td>
<td>4,280</td>
<td>12.3</td>
<td>17.7</td>
</tr>
<tr>
<td>Poona(^2)</td>
<td>2003-2007</td>
<td>1,035</td>
<td>10.6</td>
<td>13.4</td>
</tr>
<tr>
<td>Sikkim State(^2)</td>
<td>2003-2007</td>
<td>87</td>
<td>6.5</td>
<td>10.2</td>
</tr>
<tr>
<td>Trivandrum(^2)</td>
<td>2005-2007</td>
<td>188</td>
<td>10.8</td>
<td>10.0</td>
</tr>
</tbody>
</table>

ASR: Age-standardized rate. Standardized rates have been estimated using the direct
method and the World population as the reference.

\(^a\)Accumulated number of cases during the period in the population covered by the corresponding registry.

\(^b\)Rates per 100,000 women per year.

Data sources:


### 1.5.2 Mortality Rate

Table 1.3 Cervical cancer mortality in India (estimations for 2012)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>India</th>
<th>South Asia</th>
<th>World</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual number of deaths</td>
<td>67,477</td>
<td>79,958</td>
<td>265,653</td>
</tr>
<tr>
<td>Crude mortality rate(^a)</td>
<td>11.1</td>
<td>9.4</td>
<td>7.6</td>
</tr>
<tr>
<td>Age-standardized mortality rate(^a)</td>
<td>12.4</td>
<td>11.0</td>
<td>6.8</td>
</tr>
<tr>
<td>Cumulative risk (%) at 75 years old(^b)</td>
<td>1.4</td>
<td>1.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

\(^a\)Rates per 100,000 women per year.

\(^b\)Cumulative risk (mortality) is the probability or risk of individuals getting from the disease during ages 0-74 years

Figure 1.1 Cervical cancer mortality compared to other cancers in women of all ages in India (estimations for 2012)
Figure 1.2 Annual numbers of deaths of cervical cancer by age group in India and Southern Asia (Estimations for 2012)

1.6 Risk Factors in Cervical Cancer Screening in India

Lack of efficient high quality precancerous screening techniques, poor infrastructure and cost affordability makes the count to the heights. Cervical cancer is well preventable by screening in the precancerous cervical cancer lesions early detection leads to faster and more successful treatment. According to multiple studies that have been carried out women that have been screened for at least once in their lifetime between ages 30 and 40 reduce cancer risk by 25-36%. The primary reason is lack of awareness of the disease and access to screening and health services. Even though the screening programs are available in India, we are facing the following problems.

1.6.1 Poor Design in Screening Programs

The screening programs are offered through government agencies, still quality and the services are poor. At least once in three years if a woman wish to do screening for cervical
cancer definitely the incidence rate be reduced much. The practices and methodologies of the screening program designs will make the public to have more awareness on this issue. The design of the screening program will make it compulsory for all women. According to the Program for Appropriate Technology in Health (PATH), an international non-governmental organization, “An important reason for the high incidence in developing countries is the lack of effective screening programs to detect precancerous conditions and treat them before they progress to cancer”. According to the report of ICMR in India the incidence rate of cervical cancer varies from 20 to 35 of 1000 women between the age group of 35 to 64 years while in developed countries it is just 1 to 8 of 1000. This clearly shows the need of good design in cervical cancer screening could only help Indian sub-continent women population. In India, 132,000 new cases are reported annually with 74,000 deaths occurring each year hence, every 7th minute a woman dies due to cervical cancer. It is predicted that figures are expected to double by 2020 if no action is taken [51].

1.6.2 Lack of Knowledge and Awareness

In India more than 60 percentages of people live only in villages. But almost all cervical cancer screening programs here reach only a small fraction of the population, usually urban women community only. This is due to the lack of knowledge and awareness towards cervical cancer screening, as well as of the lack of accessible and acceptable screening services. Even if the woman is much aware of cervical cancer they are likely to face the socio-cultural barriers to actually participating in cervical cancer screening. Only minimum percentage of women undergoes.

1.6.3 Shortage of efficient equipment in Health Centers

The screening program required to have the equipment starting from the collection of pap smear, applying the chemicals to the slides to avoid the overlapping of cells, special microscopes to get the clear view and many more advanced medical instruments for further
processing. In primary health centers of India have only very limited quantity of these kinds and even if it is available, it has some problems. The cervical cancer screening population ratio and the available medical equipment are not satisfactory. Initially these instruments are provided to the health centers and furthermore no maintenance and future enhancements. Women population feels tired to move to these places for screening since they find these places are not germ-free and not the place for screening. The women who have heard of cervical cancer form them are not aware of the availability of screening techniques as a method to prevent cervical cancer.

1.6.4 Shortage of Cytologists and Training

There is a big shortage of cytologists in the developing countries like India. In the curriculum of nursing the cervical cancer screening process is not included and no doctors and nurses are trained on the process of taking Pap smear images and further analysis. Also the doctors and nurses are lack in biological background of cervical cancer and the management of precancerous diagnosis. According to the report in India (1995), there were about 200 cyto-technicians. Even if all women would be screened only once in lifetime, the country would need four times more of them [52]. The cytology laboratories are available only at multi-specialty hospitals in very small count and also lack in mechanisms for quality control. Also the people who are affected with these diseases have to travel long distance and also to wait a long time in the hospitals to get them treated. Thus people used to move to the private hospitals to avoid these unpleasant situations or even avoid cervical cancer screening.

1.6.5 Lack of Follow-ups and Financial Backgrounds

Many patients never come back to pick up the smear result, and, especially in urban areas, it might be difficult to trace patients with an abnormal smear. As a result, many patients never receive adequate follow-up. Inadequate follow-up is a factor contributing to reduced programme impact in a variety of countries such as Brazil, India, Indonesia, Lesotho,
and Nigeria [53]. Many people in the country are also facing multiple economic barriers to get these screening services since of work and family. In many governments funded health care centers do not account for low level rates for these treatments and which are not affordable to the normal poverty line population in India.

1.7 Cervical Cancer Screening Programs in India

It is very much necessary to implement a population based cervical cancer control program to reduce the mortality rate in countries like India. The new program should ideally be integrated into the existing health services and should have components as well defined target population, linkage between detection and treatment and appropriate quality control [54]. In last five years government of India has taken various steps and programs in cervical cancer prevention and control. Among those programs, validation of VIA as the alternative, low cost screening tests, Cryotherapy effective technology to treat cervical pre-cancers and the vaccine against HPV are most important.

1.7.1 NCCP

The National Cancer Control Program (NCCP) formulated and funded by the Ministry of Health, Government of India has stressed upon the implementation of community based cervical screening program at least in select districts of each state [55]. A national guidelines for cervical screening was prepared by the expert committee of the Ministry that included representatives from the Regional Cancer Centres, Federation of Obstetrics and Gynecologists of India, Indian Academy of Cytologists, World Health Organization and International Agency for Research in Cancer, France [56]. The primary goals and objectives of NCCP in cervical cancer are given more importance in strengthening of existing cancer treatment facilities which are inadequate and the early detection and diagnosis of cervical cancer.
Existing Schemes under National Cancer Control Program are:

1. **Recognition of New Regional Cancer Centers (RCCs):**

   To enhance the cancer treatment facilities across the country and reduce the geographical gap in the country in the availability of cancer care facilities, New Regional Cancer centres are being recognized. A one-time grant of Rs. 5.00 crores are being provided for New RCC’s.

2. **Strengthening of existing Regional Cancer Centers**

   A one-time grant of Rs.3.00 crores is provided to the existing Regional Cancer Centers to further strengthen the cancer care services.

3. **Development of Oncology Wing**

   Government Hospitals and Government Medical Colleges are provided with a maximum grant for the development of Oncology Wing.

4. **District Cancer Control Program**

   The DCCP will be implemented by a nodal agency, which may be a Regional Cancer Centre or Government Medical College or Government Hospital with radiotherapy facility. A cluster of 2-3 districts are taken up for prevention, early detection, minimal treatment and provision of supportive cancer care at district levels. A grant-in-aid of Rs. 90.00 lakhs spread over a period of 5 years is provided per DCCP proposal.

5. **Decentralized NGO Scheme**

   A grant of Rs. 8000/- per camp will be provided to the NGOs for IEC activities. The funds are released through a nodal agency which could be a Regional Cancer Centre, Government Medical Colleges or Government hospitals with radiotherapy facilities.

1.7.2 **NPCDCS**

   The Ministry of Health and Foreign Welfare of India launched a National Program for Prevention and Control of Cancer, Diabetes, cardiovascular diseases and stroke has among its
major objectives cervical cancer control through opportunistic screening of women above 30 years. The Objectives of NPCDCS [57] are,

1. Prevent and control common NCDs through behavior and life style changes,
2. Provide early diagnosis and management of common NCDs,
3. Build capacity at various levels of health care for prevention, diagnosis and treatment of common NCDs,
4. Train human resource within the public health setup through doctors, paramedics and nursing staff to cope with the increasing burden of NCDs.
5. Establish and develop capacity for palliative and rehabilitative care.

1.7.3 NCRP

The National Cancer Registry Program was promoted by Indian Council of Medical research with a network of cancer registries across the country. The main objectives of this program aim at:

1. To generate reliable data on the magnitude and patterns of cancer.
2. Undertake epidemiological studies based on results of registry data.
3. Help in designing, planning, monitoring and evaluation of cancer control activities under the National Cancer Control Program (NCCP).
4. Develop training program in cancer registration and epidemiology.

1.7.4 CCFI

Cancer Care Foundation of India is a charitable trust which spreads to 50 cities in India. It aims at providing free treatment to needy and helpless cancer. They provide Ayurvedic cow urine therapy for the treatment of all types of cancer. Cow Urine has been researched and patented for its ability to improve efficacy and absorption of anti-cancer drugs of modern medicine (allopath) and herbal. The research wing AYUSH (Government of India)
has researched on herbs for its anti-cancer properties and pharmacological activities as per modern science.

1.8 Automated Cervical Cancer Screening

An automated screening system should ideally be relatively fast, tireless, and give reproducible results. Research into Pap smear screening automation has been underway since the late 1950’s with most systems attempting to reproduce the way in which humans perform the screening task.

The automated Pap smear screener takes a slide and uses a microscope incorporating a digital camera to capture a digital image of a single microscopic field of view. Image processing algorithms are then used to “clean up” the image, correcting for irregularities such as uneven lighting across the field of view, and poor contrast. The boundary of the objects is detected and pixels are classified as either belonging to the object or to the background of the image. Further image analysis can then be restricted to the object of interest. Features are measured using features extraction principles from the prior knowledge of the problem characteristics. The feature dataset is often then post processed in some way, such as by scaling or normalizing the feature values. A classifier is then constructed by selecting thresholds with which the feature values can be compared in order to assign individual cells to one of a number of classes. A second threshold, for the number of abnormal cells per slide, is then established, and any slide with more than the threshold number of “abnormal” cells is classed as an “abnormal” slide. This process is known as training. Once trained, the classifier thus constructed is saved, and can be used to classify previously unseen slides on the basis of their measured feature values.

1.9 Organization of Thesis

The thesis is divided into seven major chapters.

Chapter I discuss the need of automation in cervical cancer cytology images and the shortcomings on this analysis in developing countries like India. Various statistics on
incidence and mortality rates were shown and the various risk factors of this deadly disease in India. It also elaborates various screening techniques available as of now. It explains various programs designed and the research programs on this study.

Chapter II reviews the biological aspects of cervical cancer cytology. It has given more focuses on Papillomaviruses the root cause of this cancer and its various stages where the cells get lot of modifications. It also discusses the various international staging systems available.

Texture features analysis is the crucial step in medical image processing. Chapter III converse various techniques by which the texture features are analyzed in Pap smear images. It also focuses number of performance measures available in medical image analysis.

In recent years, great attention has been paid to the texture features and their extraction in the field of medical image processing, and the essence of feature vectors extraction from images using texture can be the base of many other processes such as classification, segmentation and identification of objects. The Chapter IV discusses the importance of texture features in Pap image analysis and elaborates the texture features extracted from cervical images.

Feature selection has been an important research area in pattern recognition, machine learning, and medical image processing communities. The sensible feature selection techniques will enhance the learning efficiency, increasing the accuracy, reducing the learning efficiency, and enhanced generalization by reducing over fitting and is explained in chapter V.

The experimental results and analysis are discussed in Chapter VI. Examples of the images used in this analysis are shown. The step-by-step method used in this work is elaborated. The experimental setting and results are also stated. An analysis of the result is presented with discussion of the ranking used in this study. Chapter VII concludes the thesis. The limitations of the system and further enhancement are stated.