Human Papilloma Virus (HPV) and its numerous strains, around 200, have been identified as a causative factor for cervical cancer [12]. Very recently, prophylactic vaccination against HPV has been introduced. The vaccines prevent HPV infections against only a smaller subset of strains [2,13-18]. American Cancer Society and UICC recommend Pap smear screening for vaccinated women also [18,19]. Even though facilities for detecting HPV infection using molecular tests are available, Pap test remains the most widely used screening test.
method. Studies show that even a single screening in a life-time substantially reduces risk of cervical cancer incidence [20]. However, competing health care priorities, insufficient financial resources, weak health systems, and limited numbers of trained providers have made high coverage for cervical cancer screening in most low- and middle-income countries difficult to achieve [2,20-23].

Visual screening of a Pap smear includes careful scrutiny of several thousand Fields of View (FOV) under a microscope, which together contains a few hundred thousand cells, for identifying a few abnormal cells. Screening is a most demanding function of the human eye-brain axis, it is exhaustive and fatigue producing [24]. According to the Clinical Laboratory Improvement Act (CLIA) of 1988 cyto technologist, those who screen the specimen, should not process more than 100 slides per day because of fatigue and habituation factor which deteriorates the quality of screening and can result in high number of false positive and negatives [25]. To give reasonable protection against developing undetected cervical cancers, eligible women need to be screened regularly. Considering 2 billion women population, in relevant age groups, screening programs generate enormous numbers of samples to analyze. Educating and financing sufficient numbers of human screeners create great practical and economic problems which have led to substantial interest in trying to automate the task. Furthermore, the human eye-brain axis is not good at appreciating the early nuclear changes which are the first indications of neoplastic transformations. The quantitative microscopy is much better suited for detection and objective measurement of early changes of malignancy [26].

Ever since the first appearance of computers, significant development efforts have been aiming at supplementing or replacing the human visual inspection of Pap smears by computer analysis [27-29]. But the problem turned out to be lot harder than expected. From the first automated system in 1950’s it took almost another half a century before the first commercially successful system appeared.

The Cytoanalyzer built during the 1950’s was the first attempt towards automation of Pap smear screening. Although the system was able to distinguish the morphological difference between normal and malignant cells it produced too many false alarms. Another early attempt was CYBEST, developed during the 1970’s, which was able to detect malignancy based on morphological features but had problems with the chromatin features primarily caused by poorly focused images. During 1980’s quite a number of systems like BioPEPR, FAZYTAN, LEYTAS, DIASCANNER etc. were developed. Although some of the systems reported accuracy comparable with conventional visual screening none was successful owing to lack of cost effectiveness [30]. Lately, research work on Pap smears images has been done with assisted segmentation where free lying cells with no interference by inflammatory cells were handpicked [31]. Such work may require significantly more effort to develop into a field deployable screening system.

Two United States Food and Drug Administration (FDA) approved automated machines were developed in the 1990’s, the AutoPap 300 QC (NeoPath, Redmond, WA, USA) and the PapNet (Neuromedical Systems Inc., Suffern, NY, USA), both systems were designed to work with conventional cytology slides. AutoCyte also developed a machine known as the AutoCyte-Screen which was able to read AutoCyte-Prep slides (now BD SurePath LBC). The experiences gained from these early commercial efforts led to the merger of the companies into TriPath Imaging Inc. (Burlington, NC, USA) and the first generation products were replaced by the AutoPap Primary Screening System, which is now known as the BD Focal Point GS Imaging System (BD Diagnostics, Franklin Lakes, NJ, USA). Cytyc also developed an interactive system with a computer prescreen that selected the most abnormal looking objects on each specimen for human inspection. In 2003 they received FDA approval for their ThinPrep Imaging System, and in 2007 they became part of the Hologic Company. The system is marketed for increasing detection of abnormalities by improved specimen preparation and screening both visually and by machine [30,32]. Even with numerous attempts, still automated screening is not sufficiently cost-effective to completely replace the visual screening judging from the relatively limited penetration of automated screening systems in the screening operations worldwide [30].

Materials and Methods

The basis for the Pap smear screening is that cancerous or precancerous abnormal cells have larger nuclei and more irregular shape and chromatin structure than normal cells, as from the Figure 1a and 1b. However, the task is not as simple as said owing to the facts that cells in the specimen, even though prepared by mono-layering technique, are often folded, overlapped, covered by blood cells or other artifacts and clustered as in Figure 1c. Moreover, as the task is analyzing a few hundred thousand cells looking for malignant cells even a very low false positive rate will result in all specimens to be classified as malignant. To address the said problems, the automated screening system uses advanced image acquisition, processing and classification technique coupled with novel monolayer slide preparation technique detailed in the subsequent sections to provide a solution, which can be adopted for mass screening of cervical cancer.

Pap smear collection

Pap smears were obtained from women attending the early cancer detection clinic and cancer detection camps of RCC. Cervical scrapes were obtained using cervicobrush and the cells were preserved in the vials provided by Surepath Liquid Based Cytology (LBC) system. A separate scrape of cells were obtained for Mega funnel Technique from a selected group of women whose consent was taken in advance. The samples were processed in the Surepath system according to the manufacturer’s instruction and MFT as described.

Mega-funnel specimen preparation technique

Each of the cell samples in 10 mL of preservative solution containing 50% alcohol, glacial acetic acid and a mucolytic agent was homogenized in a vortex for 30 seconds followed by centrifugation at 2000 RPM for 5 minutes. The cell palette was mixed well with 1 mL of preservative solution and 200-300 µL of the sample was then cyto-centrifuged onto a coated slide using a mega-funnel. The smears were fixed in 95% of alcohol for 15-30 minutes and stained using classical Pap staining method producing a specimen dimension of 22 mm×15 mm. A total of 60 MFT slides were prepared and compared against commercial LBC system to produce image of quality comparable as that of commercial LBC system. Gross appearance of slides and magnified view of smears...
Field of view selection

Specimens prepared on glass slides were magnified through a 40X lens to accurately quantify nuclear chromatin distribution which resulted in an average of 2000 FOVs needed to cover the whole specimen. Data acquisition in this work was manual and as it was impractical to cover the whole specimen with manual repositioning between FOVs, only interesting FOVs were selected from each specimen. The image data was acquired by a person skilled enough to operate a microscope optimally focused manually before acquisition.

Specimen digitization

Each FOV selected and focused manually was digitized using a person skilled enough to operate a microscope. The microscope used was Leica DM2500 with a plan apochromat objective of magnification 40X and numerical aperture of 0.65. The camera used in the digital microscope was Leica DFC495 producing RGB images with a spatial resolution of 3264×2448 pixels and sensor pixel size of 2.7 μm. The whole CMOS sensor of DFC495 has a physical dimension of 8.81 mm×6.61 mm. To capture the maximum possible area in a FOV a demagnifier of 0.63x magnification was also used, resulting in an effective pixel size of 0.1 μm. The workstations which host e-Smear and the slide analysis software were quad core Dell desktops with 4 GB of RAM having a 32 bit operating system.

Pap Image analysis

The images acquired from e-Smear were transferred to an image processing station where each image undergoes a series of processing and analysis steps to finally classify the specimen as either normal or suspicious. A flow chart of the Pap image analysis is shown in Figure 3.

Preprocessing and segmentation

A Laplacian of Gaussian (LoG) filter was used for detecting objects from Pap smear image. The Laplacian operator applied on the image highlight regions of rapid intensity change and was used for edge detection. In order to reduce its sensitivity to noise, the Laplacian operator was applied to an image that has been first smoothened by a Gaussian smoothing filter. Red blood cells, RBC’s are removed using color information from the true color RGB input image [33].

Feature extraction and ranking

The heart of the quantification and automation task is to determine what is to be measured and how it should be measured. Over the past 50 years of quantitative cytometry quite a large set of features have been tried and tested for various applications [34]. Around 40 mathematical features which can accurately determine morphology, texture and densitometry of cervical epithelial cells were identified heuristically. All the identified features were ranked using histogram analysis and Mahalanobis maximization function [35], which is the ratio of difference in mean and sum of standard deviation of normal and abnormal cells.

Classification

A hierarchical multi-stage classification approach was followed for classifying normal smears from suspicious smears. In the first stage, artifacts, microbes and other debris were separated from epithelial cells [36,37]. The epithelial cells were then analyzed using a set of mathematical features to determine suspicious cells from the rest. Apart from the cell level classification, cell clusters were detected for careful scrutiny [38], significant diagnostic information was gathered from count of neutrophils [39] and Koilocytes [40]. Finally the cell distribution of the whole specimen was analyzed for deviation from normal cell distribution. The final classification decision was made by a specimen level classifier taking input from the cell level and slide level classifiers. A flow chart is shown in Figure 4.

The preprocessing, segmentation, feature extraction and classification modules were integrated with a Graphical User Interface (GUI) application called CerviSCAN (Government of India, Copyright Registration No. SW-7352/2013).

Ground truth collection

A Cell Marker utility (Government of India, Copyright Registration No. SW-7458/2013) was developed and used by a team of experienced cytotechnologists to obtain ground truth. The Cell Marker

Preprocessing and segmentation

A Laplacian of Gaussian (LoG) filter was used for detecting objects from Pap smear image. The Laplacian operator applied on the image highlight regions of rapid intensity change and was used for edge detection. In order to reduce its sensitivity to noise, the Laplacian operator was applied to an image that has been first smoothened by a Gaussian smoothing filter. Red blood cells, RBC’s are removed using color information from the true color RGB input image [33].

Feature extraction and ranking

The heart of the quantification and automation task is to determine what is to be measured and how it should be measured. Over the past 50 years of quantitative cytometry quite a large set of features have been tried and tested for various applications [34]. Around 40 mathematical features which can accurately determine morphology, texture and densitometry of cervical epithelial cells were identified heuristically. All the identified features were ranked using histogram analysis and Mahalanobis maximization function [35], which is the ratio of difference in mean and sum of standard deviation of normal and abnormal cells.

Classification

A hierarchical multi-stage classification approach was followed for classifying normal smears from suspicious smears. In the first stage, artifacts, microbes and other debris were separated from epithelial cells [36,37]. The epithelial cells were then analyzed using a set of mathematical features to determine suspicious cells from the rest. Apart from the cell level classification, cell clusters were detected for careful scrutiny [38], significant diagnostic information was gathered from count of neutrophils [39] and Koilocytes [40]. Finally the cell distribution of the whole specimen was analyzed for deviation from normal cell distribution. The final classification decision was made by a specimen level classifier taking input from the cell level and slide level classifiers. A flow chart is shown in Figure 4.

The preprocessing, segmentation, feature extraction and classification modules were integrated with a Graphical User Interface (GUI) application called CerviSCAN (Government of India, Copyright Registration No. SW-7352/2013).

Ground truth collection

A Cell Marker utility (Government of India, Copyright Registration No. SW-7458/2013) was developed and used by a team of experienced cytotechnologists to obtain ground truth. The Cell Marker
study was manually screened by cytologists with over 25 years of experience and the ground truth was recorded. Smears were also analyzed in parallel by the automated system using image processing methods. Manual cytology was considered as the gold standard for benchmarking the efficacy of the automated analysis. All abnormal smears were biopsy proven. Table 1 describes distribution of slides used for validation.

**Results**

The number of smears correctly classified and misclassified is described in Table 2. True positives are those abnormal smears which are classified as suspicious and sent for cytologist's review by automated analysis. True negatives are normal smears which were classified as normal and require no further human intervention. False positives and negatives are misclassified smears. Not processed smears are smears which were rejected from automated analysis either because of poor image quality or insufficient number of image fields.

The system screened out 60% of the normal smears which needs no further human review and classified 80% of the abnormal cases as suspicious which needs further expert human review, as in the Table 3. Detailed analysis of accuracy in normal and different precursors of cervical cancer is elaborated in Table 4.

**Discussion**

**Comparison with commercial Systems**

In a randomized controlled trial by Kitchener et al. [30], automated-

is a GUI application used to generate ground truths, visualization of segmentation results, feature extraction, training set creation and visualization of classification results. A total of 15,708 malignant cells were hand marked by cytotechnologists and close to 300,000 normal cells from normal smears, verified by cytotechnologists, were auto marked using CellMarker. 3092 cells which include 2935 normal cells and 157 abnormal cells of all grades were used to train the classification algorithm. The study protocol was approved by the Human Ethics Committee (HEC) of RCC, Thiruvananthapuram (HEC No. 22/2009). The evaluation protocol is elaborated in Figure 5.

**System validation**

All the Pap smears used for the system validation were obtained from women attending the Early Cancer Detection Centres (ECDC) of RCC from different places in Kerala like Karunagapally, Ernakulam and Palakkad apart from routine examination in RCC, Thiruvananthapuram. The smears were collected after obtaining informed consent as per the recommendation of the HEC of RCC. The semi-automated system for screening of cervical cancer was used in RCC, Division of Cancer Research, since March 2011.

1107 Pap smears were used for the validation. Each slide in the

![Figure 4: Flow diagram of classification approach followed.](image)

![Figure 5: Specimen evaluation protocol.](image)

![Table 1: Category wise slide count.](image)

![Table 2: Classification Statistics.](image)

![Table 3: Accuracy of automated analysis.](image)

![Table 4: Accuracy on different Specimen category.](image)
where a robotic arm transfers each slide from slide tray to a scanning component requiring occasional maintenance due to wear & tare availability. Analysis station contains desktop grade computers which are capable of handling the load of the system. The system is designed to be scalable and can be expanded based on the number of cases to be processed.

Future work

This work has demonstrated a system capable of detecting early pre-malignant changes of cervical smears with acceptable classification performance. However, the operation of the system needs to be made more time-efficient before large scale deployment. We here outline some of the considerations that will be taken into account for that work.

Motorized microscope: The existing system was designed in a semi-automated fashion where a semi-skilled person can operate the microscope, position the stage, focus and acquire the images while the analysis part is taken care of by the image analysis platform. A more sophisticated approach is full automation of slide loading and scanning, where a robotic arm transfers each slide from slide tray to a scanning space which will be controlled by stepper or piezo controlled motors for movement of slide in XY direction for FOV hopping. Image focus on each FOV is controlled by moving either stage or objective in Z direction. Throughput of the system depends to a larger extend on the speed of the image acquisition which requires motorized mechanical movement. However, as the automated platform can work 24/7, excluding time for routine maintenance, human efficiency can very well be breached. From field trial by Kitchener et al., adoption of an automated-assisted system resulted in increase in productivity by 60%-80% [30].

Malignancy associated changes: An alternative approach to exhaustive scan of complete specimen is analysis of the field-effect or malignancy associated changes (MAC) [41,42], which points to the subtle changes in normal cells present in malignant smears. These discoveries were confirmed in the early research on automated cervical screening [43,44]. If the MAC approach is adopted only a small subset of cells from each smear needs to be analyzed instead of the complete smear scan. For MAC analysis it is essential to analyze the chromatin pattern in great detail. A highly accurate artefact removal with perfect focus is a prerequisite to convincingly demonstrate that MAC alone can detect early premalignant changes with sufficient sensitivity [30].

Field trials: The system need to undergo an extensive independent evaluation on around 10,000 smears.

Image analysis throughput: Image analysis throughput can be improved by porting CPU intensive operations to graphics processing units (GPU) where hundreds of dedicated highly parallel cores make the system more efficient with only a marginal increase in cost.

Building a cost effective system: The goal of our project is to demonstrate that a cost effective system for Pap-smear screening can be built. Based on the experiences gained from our study we have made a rough estimate of the component costs for a final automated system. It can be based on a standard microscope with minimal modification for integrating digital camera, quality optics, motorized stage and illumination. Furthermore, commercial XY motorized stages are now available with a travel range sufficient to load multiple slides in which case human intervention is required may be in every 2 hours or so just to re-load slide tray. Such cost optimized systems which avoid the need for expensive embellishments can be very well be implemented under $34,000, Table 6 describes approximate component wise costs. On an economic angle, if such a system will be able to screen a moderate target of 20,000 women per year, the screening cost for each smear can be reduced to under $2 from $7, which is the current cost of slide screening in India. This savings computes to $100,000 per year or rather possibilities of offering screening to many more.

To better address disperse population in low resource setting a Centralized Smear Analysis Station (CSAS) and multiple Satellite Smear Collection Centre (SSCC) model is suggested. CSAS should have all equipments mentioned in Table 6, while SSCC need to have only person(s) collecting cervical smears on a buffer which will be transferred to CSAC. Number of CSAS and SSCC can be decided based on population required to be screened and also based on resource availability. Analysis station contains desktop grade computers which may have evolved into a stable product requiring very less maintenance, if at all required, support will be readily available. Same applies for microscope and its accessories. Microscope XY stage will be only component requiring occasional maintenance due to wear & tear caused by heavy duty slide scanning which is easily addressed being at a centralized location.
References

1. http://globocan.iarc.fr

2. World Health Organization (2013) WHO guidance note: Comprehensive cervical cancer prevention and control: A healthier future for girls and women.

3. http://countrymeters.info/en/

4. http://www.uspreventiveservicestaskforce.org/uspstf/uspscerv.htm

5. Papanicolaou GN, Traut HF, Stanton M, Friedberg A (1943) Diagnosis of Uterine Cancer by the Vaginal Smear. Oxford University Press, New York, USA.

6. Stenkvist B, Bergström R, Eklund G, Fox CH (1984) Papanicolaou smear screening and cervical cancer. What can you expect? JAMA 252: 1423-1426.

7. Christopherson WM, Parker JE, Mendez WM, Lundin FE Jr (1970) Cervix cancer death rates and mass cytologic screening. Cancer 26: 808-811.

8. Nakamura M (1982) Trends in incidence of cervical cancer in Nordic countries. In Magrass K, ed. Trends in cancer Incidence. Hemisphere Publishing Corporation, Washington DC

9. Lynge E, Madsen M, Engholm G (1989) Effect of organized screening on incidence and mortality of cervical cancer in Denmark. Cancer Res 49: 2157-2160.

10. Macgregor JE (1976) Evaluation of mass screening programmes for cervical cancer in N.E. Scotland. Tumour 62: 287-295.

11. Johannesson G, Geirsson G, Day N (1978) The effect of mass screening in Iceland, 1965-74, on the incidence and mortality of cervical carcinoma. Int J Cancer 21: 418-425.

12. Nubia Muñoz, Xavier Bosch F, Silvia de Sanjosé, Rolando Herrero, Xavier Castellsagué, et al. (2003) Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. N Engl J Med 348: 518-527.

13. CUTTS JT, Francischi S, Goldie S, Castellsague X, de Sanjosé S, et al. (2007) Human papillomavirus and HPV vaccines: a review. Bull World Health Organ 85: 719-726.

14. Muñoz N, Bosch FX, Castellsague X, Diaz M, de Sanjosé S, et al. (2004) Against which human papillomavirus types shall we vaccinate and screen? The international perspective. Int J Cancer 111: 278-285.

15. Luisa L Villa, Ronaldo LR Costa, Carlos A Petta, Rosires P Andrade, Kevin A Matti, et al. (2007) Quadrivalent vaccine against human papillomavirus and HPV infections: a randomised controlled phase 2b efficacy trial. Lancet Oncol 6: 271-278.

16. Joura EA, Leodolter S, Hernandez-Avila M, Perez G, et al. (2007) Efficacy of a quadrivalent prophylactic human papillomavirus types (6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. Lancet Oncol 6: 271-278.

17. O’Horo EA, Overdolor S, Hernandez-Avila M, Wheeler CM, Perez G, et al. (2007) A comparison of automation-assisted and manual cervical screening: a randomized controlled trial. JAMA 298: 186-194.

18. Beral V, Prentice A, Cahn J, in press. Screening for cervical cancer in women with low education. Lancet 369: 1702-1704.

19. Agarwal S, Thakur S, Prabhakar AK, Sharma S, Das DK (1993) Estimation of reduction of life-time risk of cervical cancer through one life-time screening. Neoaplasma 40: 255-258.

20. Sujathan K, Kannan S, Pillai KR, Mathew A, Joseph M, et al. (1995) Implications of gynaeological abnormalities in pre-selection criteria for cervical screening: preliminary evaluation of 3602 subjects in south India. Cytopathology 6: 75-87.

21. Prabhakar AK (1992) Cervical cancer in India—strategy for control. Indian J Cancer 29: 104-113.

22. Miller AB, Chamberlain J, Day NE, Nakama M, Prorok PC (1990) Report on a Workshop of the UICC Project on Evaluation of Screening for Cancer. Int J Cancer 46: 761-769.

23. Elizabeth K Abraham Cytoescreening and Evaluation. P 19-25. Handbook on quality assurance in PAP smear

24. Arbyn M, Bergeron C, Klinhamer P, Martin-Hirsch P, Siebers AG, et al. (2008) Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. Obstet Gynecol 111: 167-177.

25. Byju NB, Sujathan K, Malm P, Kumar RR (2013) A fast and reliable approach to cell nucleus segmentation in PAP stained cervical smears. CSI Transactions on ICT.

26. Rajesh Kumar R, Aijith Kumar V, Shashank Kumar PN, Sudhambay S, Ravindra Kumar (2011) Detection and Removal of Artifacts in Cervical Cytology Images Using Support Vector Machine. IEEE ITME 717 - 721.

27. Malm P, Balakrishnan BN, Sujathan VK, Kumar R, Bengtsson E (2013) Debris removal in Pap-smear images. Comput Methods Programs Biomed 111: 120-138.

28. Pournami S Chandran, Byju NB, Deepak R, Rajesh Kumar R, Sudhamany S, et al. (2012) Cluster Detection in Cytology Images Using the Cegograph Method, International Symposium on Information Technology in Medicine and Education, 978-1-4673-2108-2/12 IEEE.

29. Dabzuk RK, SharadKumar PN, RajeshKumar R, Sujathan K (2014) Automated Identification Of Neutrophils In PAP Smear Images. International Conferences for Convergence Of Technology 2014.

30. Sharath Kumar PN, Sujathan K, Rajesh Kumar R (2014) A Novel Approach for Koliocyte Identification in Cervical Smears Using Fourier Descriptors. International Conference on Biomedical Engineering & Assistive Technologies.

31. Hallinan J (1999) Detection of malignancy associated changes in cervical cells using statistical and evolutionary computation techniques, PhD thesis, The University of Queensland.

32. Nieburgs HE, Goldberg AF, Bertini B, Silagi J, Pacheco B, et al. (1967) Malignancy associated changes (MAC) in blood and bone marrow cells of patients with malignant tumors. Acta Cytologica 11: 415-423.

33. Burger G, Jutting U, Rodenacker K (1981) Changes in benign cell populations in cases of cervical cancer and its precursors. Analytical and Quantitative Cytology 3: 368-375.

34. Wied GL, Bartels PH, Bibbo M, Sychra JJ (1980) Cytomorphometric markers for uterine cancer in intermediate cells. Anal Quant Cytol 2: 257-263.