Novel Benchmark Database of Digitized and Calibrated Cervical Cells for Artificial Intelligence Based Screening of Cervical Cancer

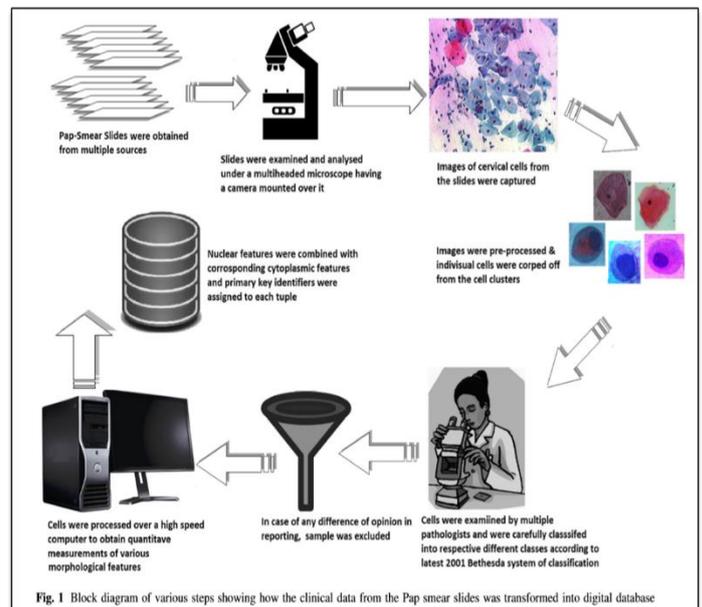*Abid Sarwar*
University of Jammu, India

Abstract

The primary objective of this research work is to develop a novel benchmark database of digitized and calibrated, cervical cells obtained from slides of Papanicolaou smear test, which is done for screening of cervical cancer. This database can serve as a potential tool for designing, developing, training, testing and validating various artificial intelligence-based systems for prognosis of cervical cancer by characterization and classification of Papanicolaou smear images. The database can also be used by other researchers for comparative analysis of working efficiencies of various machine learning and image processing algorithms. The database can be obtained by sending a request to the corresponding author. Besides developing a rich machine learning database, we have also presented a novel artificial intelligence-based hybrid ensemble technique for efficient screening of cervical cancer by automated analysis of Papanicolaou smear images. The correct and timely diagnosis of cervical cancer is one of the major problems in the medical world. From the literature it has been found that different pattern recognition techniques can help them to improve in this domain. Papanicolaou smear (also referred to as Pap smear) is a microscopic examination of samples of human cells scraped from the lower, narrow part of the uterus, called cervix. A sample of cells after being stained by using Papanicolaou method is analyzed under microscope for the presence of any unusual developments indicating any precancerous and potentially precancerous developments. Abnormal findings, if observed are subjected to further precise diagnostic subroutines. Examining the cell images for abnormalities in the cervix provides grounds for provision of prompt action and thus reducing incidence and deaths from cervical cancer. It is the most popular technique used for screening of cervical cancer. Pap smear test, if done with a regular screening programs and proper follow-up, can reduce cervical cancer mortality by up to 80 % (Arbyn et al. Ann Oncol 21:48–458, 2010). The contribution of this paper is that we have created a rich machine learning database of quantitatively profiled and calibrated cervical cells obtained from Pap-smear test slides. The database so created consists of data of about 200 clinical cases (8091 cervical cells), which have been obtained from multiple health care centers, so as to ensure diversity in data. The Pap-smear slides were processed using a multi-headed digital microscope and images of cervical cells were obtained, which were then passed through various data preprocessing subroutines. After preprocessing the cells were morphologically profiled and scaled to obtain separate quantitative measurements of various features of cytoplasm and nucleus respectively. The cells in the database were carefully classified in different corresponding classes according to the latest 2001-Bethesda system of classification, by multiple cytotechnicians and histopathologists. In addition to this, we have also pioneered to apply a novel hybrid ensemble system to this database in order to evaluate the effectiveness of both novel database and novel hybrid ensemble technique to screen cervical cancer by categorization of Pap smear data. The paper also presents a comparative analysis of multiple artificial intelligence-based classification algorithms for prognosis of cervical cancer. For evaluating the effectiveness and correctness of the digital database prepared in this work, authors implemented this database for training, testing and validating fifteen different artificial intelligence-based machine learning algorithms. All the algorithms trained with this database presented commendable efficiency in screening of cervical cancer.

Fig. 1 Block diagram of various steps showing how the clinical data from the Pap smear slides was transformed into digital database

For two-class problem all the algorithms trained with the digital database showed the efficiencies in the range of about 93 to 95 % while as in case of multi class problem algorithms expressed the efficiencies in the range of about 69 to 78 %. The results indicate that the novel digital database prepared in this work can be efficiently used for developing new machine learning based techniques for automated screening of cervical cancer. The results also indicate that hybrid ensemble technique is an efficient method for classification of pap-smear images and hence can be effectively used for diagnosis of cervical cancer. Among all the algorithms implemented, the hybrid ensemble

approach outperformed and expressed an efficiency of about 98 % for 2-class problem and about 86 % for 7-class problem. The results when compared with the all the standalone classifiers were significantly better for both two-class and multi-class problems. Figure 1 show the block diagram of how the clinical data was transformed into a digital database.

Biography:

Abid Sarwar is working at the Department of Computer Sc & IT Bhaderwah Campus University of Jammu, India.

World Congress on Gynecology and Obstetrics
October 19-20, 2020.

Abstract Citation:

Abid Sarwar, Novel Benchmark Database of Digitized and Calibrated Cervical Cells for Artificial Intelligence Based Screening of Cervical Cancer, Gynecology Obstetrics 2020,
World Congress on Gynecology and Obstetrics; October 19-20, 2020