



Cervical Smear Analyzer (CSA) Expert System for identification of cervical cells in Papanicolaou smear test

Abid Sarwar

Department of Computer Science & IT (Bhaderwah Campus)
University of Jammu, Jammu, India

ABSTRACT

The primary objective of this research work is to develop an expert system for identification & classification of the cervical cells in the images of the slides of Papanicolaou smear test, which is done for screening of cervical cancer. The expert system can serve as a potential tool for mass level screening of cervical cancer by characterization and classification of Papanicolaou smear images. The Expert system presented in this work is powered by a novel hierarchical probabilistic artificial neural network that works along with the knowledgebase of novel benchmark database of digitized cervical cells. The primary purpose of employing expert systems in medicine is creation of such artificially intelligent systems which can provide assistance to a medical doctor in delivering expert diagnosis. These artificial intelligent systems support the clinical decision making by anticipating the diagnostic results once they are trained using previously acquired training data. The use of Artificial intelligence in medicine has shown substantial progress in achieving timely, reliable diagnosis and more precise treatment of many diseases.

The expert system developed in this work exhibited a competence of about 92% which has been evaluated by comparing its results with the identification & classification of cervical cells by human experts.

I. INTRODUCTION

An expert system is a computer application that applies artificial intelligence (AI) to

simulate the judgment and behaviour of a human that has expert knowledge and experience in a particular field. It is a branch of applied artificial intelligence, developed by AI community in 1960's [1]. It is a branch of applied artificial intelligence (AI), developed by the artificial intelligence (AI) community in the mid-1960s with an aim to transfer the expertise of a human into a computer. Characteristically, an expert system integrates an inference engine i.e. a set of rules in the form of a program for applying the knowledge obtained from a knowledge base that contains the accumulated experience. The expert systems these days are embedded with machine learning algorithms that allow them to learn from past experience just as is done by humans and thus improve their working efficiency with time [2]. The primary goal of application of artificial intelligence in the field of medicine is developing intelligent tools that can assist a medical doctor in delivering expert and timely diagnosis and thus can prove an aid to both patient and medical expert. These intelligent expert systems are powered by various computational techniques that are trained from the previous instances of clinical cases and are used to perform the prognosis on the unseen cases. The backbones of these systems are the various data sets prepared from various clinical cases which act as practical examples in training the system [3]. The ever increasing expansion in the field of medicine has made it hard for a physician to remain updated with all the knowledge outside the domain. In such a case Consultation with a specialist is a solution when the clinical problem lies beyond the physician's

competence, but frequently expert opinion is either unavailable or not timely available. Attempts have been made to develop computer programs that can serve as consultants.

Cervical cancer is a malignant tumor that occurs when cervical tissue cells begin to grow and replicate abnormally without controlled cell division and cell death. In such a state, the body is unable to use and manage such cells for carrying out their usual function resulting these cells transforming into a tumor. If the tumor is malignant, its cell flow through the blood stream and spread to other parts of body, as a result those parts also become infected. Usually the cervical cancer takes number of years to develop. These infected cells are then distinguished as cervical intra-epithelial neoplasia (CIN) or cervical dysplasia. The cells over the surface of cervix that show unusual changes & potentially precancerous developments are called CIN [3]. The Papanicolaou test (Pap smear) has been the widely used method in cervical cancer screening for many decades and has showed a dramatic lowering of incidents of cervical cancer and hence in related mortality rates in many countries [4]. In taking a Pap smear, cells are scraped from the outer opening of the cervix for microscopic examination

and to lookup for irregularities. The aim of the test is to detect any pre-cancerous or potentially precancerous alterations called cervical intraepithelial neoplasia (CIN) or cervical dysplasia [5].

II. METHODOLOGY

The cervical cancer screening system proposed in this work has is based upon a novel architecture hierarchical probabilistic artificial neural network (HPANN) which has been tailored for the problem under consideration. HPANN works by dividing the complex problem space in such a way that the overall classification problem is partitioned into two phases hierarchically. In the level one of hierarchy, the overall problem of classification of the cells is viewed by a probabilistic neural network as a two class classification problem and the cells are classified as benign or malignant. On the next level of hierarchy two more probabilistic neural networks are employed that further classify the benign and malignant cells into their respective classes according to the latest Bethesda system of classification of cervical cells. Figure-1 shows the block diagram of the overall working of HPANN.

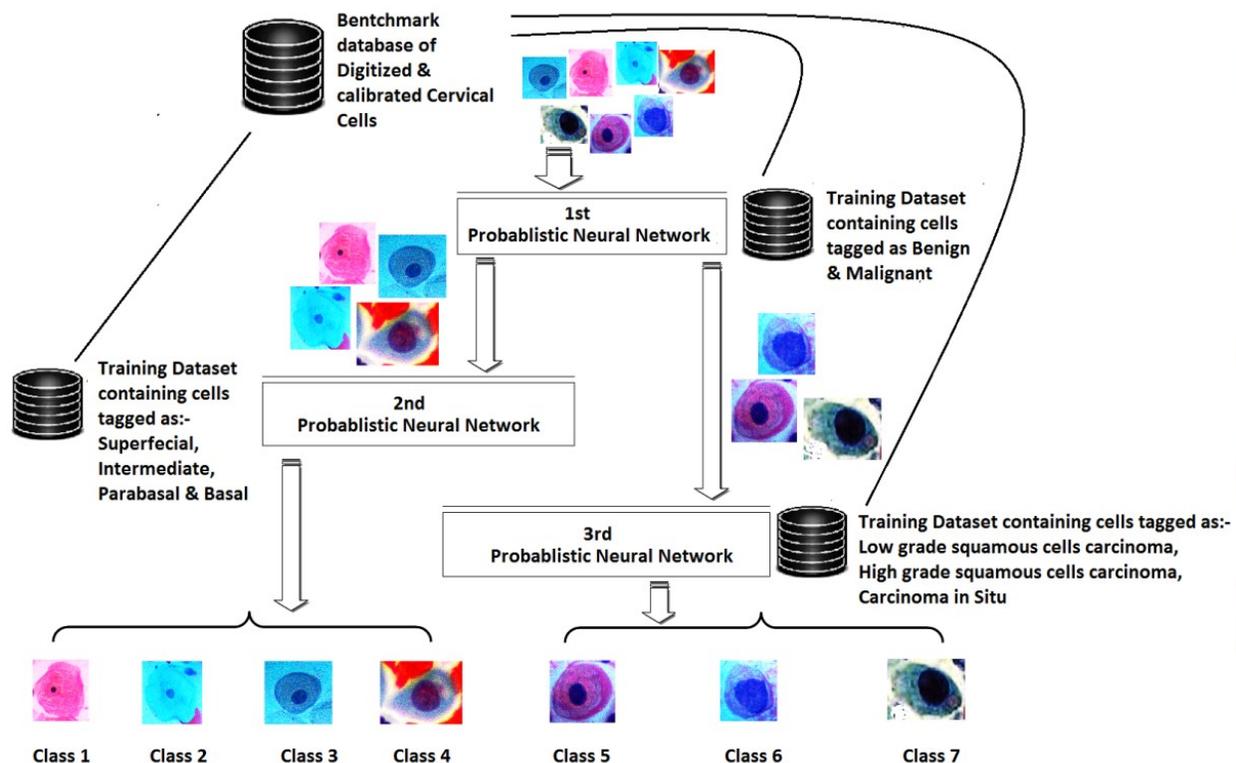


Fig. 1 Block diagram of working of Hierarchical Probabilistic Artificial Neural network

DATABASE FOR STUDY:

For training the HPANN a novel database of digitized and calibrated cervical cells was prepared from the slides of Pap-Smear test [6]. The database consists of 8091 tuples which represent data of about 200 clinical cases; each tuple containing 25 attributes, and is identified by a unique primary key. Among the 25 attributes 12 correspond to the features of cytoplasm, 12 represent the features of Nucleus, one attribute about the ratio of nuclear-cytoplasmic area and the one last attribute identifies the class to which the particular cell belongs. Each tuple corresponding to one cervical cell, extracted from the slides of pap-smear test which were obtained from three medical health care institutions in northern India viz Government Medical College, Jammu, Acharaya Shree Chandra College of Medical Science and Hospital, Jammu and Nijjer Pathology Laboratory, Amritsar. From the hospital records; cases of cervical cancer were identified and their corresponding pathological records were obtained. All the medical ethical issues were taken into consideration so that the identity of the patient is not compromised and revealed in any case. The slides were observed and analyzed under a multi-headed microscope (NIKON Nikon Eclipse E400 DS-F12) having a digital camera mounted over it and connected and configured with an attached computer. After examining the cells under the microscope under different level of magnifications (i.e. 10x, 40x, 100x), images of all the slides were captured at 40x magnification, so as to ensure uniformity and consistency among all the cells. These images were allowed to pass from various pre-processing subroutines so as to obtain distinct individual cells segregated from the cell clusters. While pre-processing the images care was taken that the size and resolution of the images is preserved. The individual cells were cropped-off from the cell cluster obtained from the microscope followed by enhancing their brightness, color and contrast using image processing techniques, where ever required, so as to make the Cytoplasmic and nuclear features easily recognizable. Unique names were assigned to all the cells with an aim to identify each of them distinctly in the database. The database as such, contains a total of 8091 cervical cells which have been carefully differentiated manually into different classes respectively using the 2001-Bethesda system of classification. To ensure the accuracy each cell included in the database was inspected by two trained

Cyto-pathologists, and complex samples were also subjected to multiple subroutines of examination. The diagnosis done by the cyto-pathologists were cross checked with the corresponding diagnosis in the medical records for the clinical case. In case of any difference of opinion in Pap smear reporting the sample was excluded from database.

CERVICAL SMEAR ANALYZER

The Cervical smear analyzer proposed in this work has three components viz de-noising-segmentation-calibration sub-module, training-testing sub-module & cell-identification sub-module. The images of the cervical cells as presented to the system usually have many bodily secretions like RBCs, present on it which needs to be removed from the area of interest. Moreover the texture of the cells image is also not uniform throughout the cell; this makes the identification and calibration of the cell quite complicated. The de-noising and segmentation sub-module de-noises the cell image by unifying the pixel values of all the pixels in the regions of cytoplasm and nucleus to one value. This is followed by segmenting the cytoplasm and nucleus out from the cell image to form two image objects. Once the image segmentation is done to obtain nucleus and cytoplasm as separate objects, they are fed to another sub-module which profiles these image objects on the basis of 24 morphological features 12 each from cytoplasm and nucleus. These features are then concatenated to form a vector that completely represents the morphological characteristics of the cell under consideration. Figure-2 shows Cervical Smear Analyzer in execution.

The training-testing module is used to train the system to get acquainted with the new knowledge gained. Once the system is trained, its performance can be analyzed using evaluated using various performance metrics like confusion metrics, ROC analysis, tuple-wise error etc. If the desired level of accuracy is achieved, the networks can be saved in the system.

The Cell-identification sub-module classifies and tags the fresh cells into their respective classes by de-noising, segmenting & profiling the cells and then feeding them into the trained HPANN. The module picks up multiple cells form a folder and classifies them into their respective classes and presents a summary of all the cells analysed in the form of a report count of number of cells of each type.

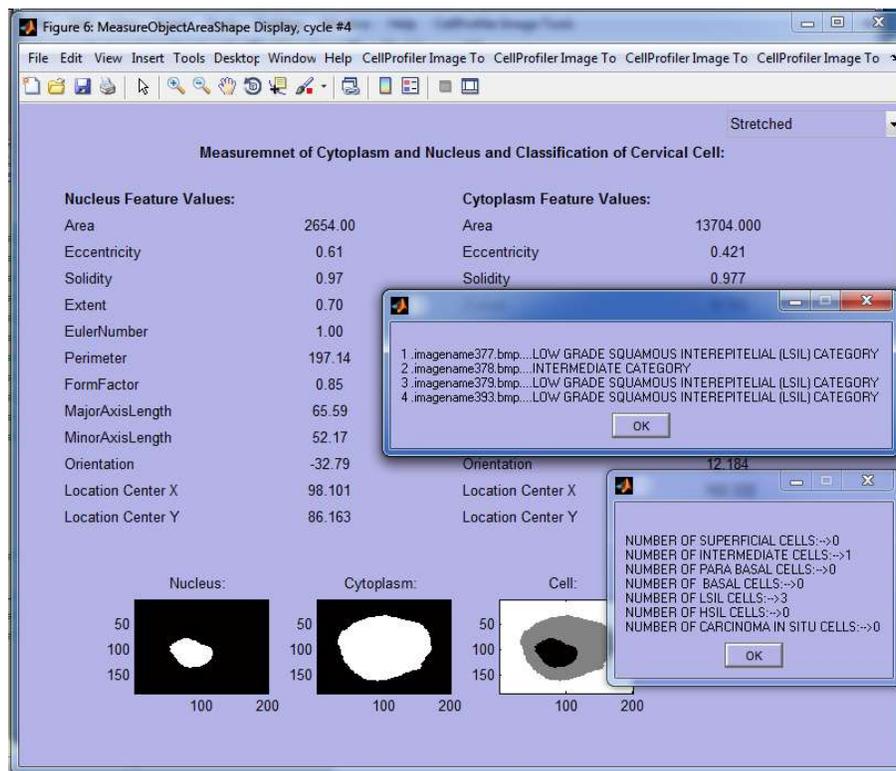


Fig. 2 Cell-identification sub-module of Cervical Smear Analyzer in execution

RESULTS AND DISCUSSION

The working efficiency of the Cervical Smear Analyzer was evaluated by comparing the results obtained from it with the cell identification as done by a human medical expert. The individual networks of the HPANN were tested with 2,617 (20% of the overall number of cells in the database) instances of the cervical cells. The 1st Probabilistic neural network presented an efficiency of about XYZ% while as the 2nd and 3rd Probabilistic neural networks presented efficiencies of about XYZ% & XYZ% respectively. The overall efficiency of the Cervical Smear Analyzer was about 92%.

REFERENCES

- 1) Shu-Hsien Liao, "Expert system methodologies and applications—a decade review from 1995 to 2004", in *Expert Systems with Applications* 28 (2005) 93–103
- 2) Per Winkel, "An application of Expert system in the Clinical Laboratory", in *Clinical Chemistry* 35/8, 1595-1600 (1989)
- 3) Abid Sarwar et. al. "Hybrid ensemble learning technique for screening of cervical cancer using

Papanicolaou smear image analysis", in *Personalized Medicine Universe* 4 (2015) 54-62

- 4) Mat-Isa Nor Ashidi, Mashor Mohd Yusoff, Othman Nor Hayati. An automated cervical pre-cancerous diagnostic system. *Artificial Intelligence in Medicine* 2008;42(1):1e11.
- 5) Chang Chun-Lang, Hsu Ming-Yuan. The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer. *Expert Syst Appl* 2009;36(No. 7):10663e72.
- 6) Abid Sawar et. al, "Novel benchmark database of digitized and calibrated cervical cells for artificial intelligence based screening of cervical cancer", in *Journal of Ambient Intelligence and Humanized Computing* DOI 10.1007/s12652-016-0353-8 (2015)