Cervical Cancer Segmentation and Classification Using Machine Learning

[1]Mr. A Arul Prakash, [2]M. Tejaswini, [3]A. Navya Sivani

[1]Assistant Professor
[1,2,3]Department of Computer Science and Engineering,
Bharath Institute of Higher Education and Research Chennai, India

Abstract: A large group of women population, in the world is affected by the cervical cancer in their different age groups. Hence, most of the researchers, pathologists and also more number of collegiate have provided more number of solutions, to identify this cancer from the test images of pap smear screening test. But their results represent, only to an extent, whether the patient is affected by cancer or not, and do not specially identify the exact nature and severity of the cancer. In this project, it is proposed, to identify the cancer and classify the stage of the cancer from the pap smear test images with high accuracy. The proposed work is based on segmentation, enhancement, and feature extraction and classification process. Test results show the mean and area value-based classification of the normal and abnormal cell and also to identify the stage of the cervical cancer. Hence the results would be more useful for the pathologists to reduce their work load and minimize the human error while improving the accuracy of diagnosis.

Keywords: Machine Learning, Cervical Cancer, Support Vector Machine, Convolution Neural Network

INTRODUCTION

Cancer refers to the uncontrolled multiplication of a group of cells in particular location of the body. A group of rapidly dividing cells may form a lump, micro calcifications or architectural distortions which are usually referred to as tumours. Cervical cancer is any form of malignant tumour which develops from cervical cells. Cervical cancer is one of most hazardous types of cancer among women in the world. The world health organization's International Agency for Research on cancer (IARC) estimates that more than 400,000 women expire each year with cervical cancer. Today, there is an urgent need in cervical cancer control and it is achieved primarily by knowing different risk factors. Secondly, there is need to detect this disease in early stage by knowing different symptoms of this disease, so it can be cured. Cervical cancer is mainly of two types: Invasive and Non-invasive. Invasive type is the one in which cancerous cells break through normal cervical tissue barriers and spread to other parts of the body. While in non-invasive, cancerous cells remain in a particular location of the cervical and do not spread to surrounding tissue, ducts or lobules. Cervical analysis techniques have been improved over the last decade. Number of automated classification systems has been developed over last years. Different techniques have varying results. However, there still are issues to be solved: developing new and better techniques. The comparison between different systems helps us to know better system with high performance; this will assist radiologists to take accurate results regarding the disease. Radiologists still produces some variation in reading images. So, there is a need for automatic interpretation of images or automated classification system, and for this purpose classifier is required. Nowadays many techniques are used for classification but Neural Network (NN) and Support Vector Machine shows better results in many instances. This paper gives comparative analysis of NN and SVM.

LITERATURE REVIEW

Nord Screen - An Interactive Tool for Presenting Cervical Cancer Screening Indicators in the Nordic Countries Veil-Matti Partanen ; Maiju Pankakoski ; Zurab Bzhalava ; Piret Veerus ; Ahti Anttila ; Tytti Sarkeala ; Ameli Tropé ; Stefan Lönnberg ; Sirpa Heinävaara ; Joakim Dillner ; Ágúst Inga Ágústsson

Quality assurance and improvement of cancer screening programmes require up-to-date monitoring systems and evidence-based indicators. National quality reports exist but the definition and calculation of indicators varies which makes comparison between countries difficult. The Nord Screen project aims to facilitate comparison of cancer screening across countries and support quality improvement. It has developed a publicly available Web-based interactive application to access standardized performance and outcome indicators of cervical cancer screening, based on up-to-date Nordic cancer screening register data.

Intrinsic motivated cervical cancer screening intervention framework Md. Rajib Hasan ; Hamid Gholamhossein ; Nurul I Sarkar ; S M Sacituzumab .

Cervical cancer turned into a reason of extreme mortality even though it is preventable. The expansion rate of cervical cancer is at alarming rate internationally, including both developed (e.g., New Zealand) and developing (e.g., Bangladesh) countries. This study considers survey data collected from Chittagong Medical College Hospital in Bangladesh and other secondary data from open sources. Studies have uncovered that younger women aged from 25 to 34 tend not to react to cervical screening program. Without regular cervical cancer screening, the early changes that could progress to cervical cancer suffering would not be picked up. This study aims to address these concerns by proposing a cervical screening intervention program to encourage women of all age group to participate in the screening program. The intervention could contribute to the health sector by decreasing the cervical cancer mortality rate and associated cost. The preliminary investigation shows that about 96% screening take-up following such intervention program suggesting that a proper intrinsic motivated cervical cancer intervention program is required.

Digital health game on cervical health and its effect on American women's cervical cancer knowledge Rohit Nirmal ; Chang Yun ; Martin Le ; Patipol Paripoonnanonda ; Jenny Yi
We develop a digital health game, a potential standalone and/or supplementary tool to conventional static educational tools such as pamphlets and videos, to investigate its feasibility and effectiveness in disseminating truth and dispelling myths about human-papillomavirus (HPV)/cervical cancer prevention and treatment to female population in US, both native and those of
Stratifying Cervical Cancer Risk with Registry Data Nicholas Baltzer ; Mari Nygard ; Karin Sundstrom ; Joakim Dillner ; Jan Nygaard ; Jan Komorowska .
The cervical cancer screening programmes in Sweden and Norway have successfully reduced the frequency of cervical cancer incidence but have not implemented any form of evaluation for screening needs. This means that the screening frequency for individuals can be suboptimal, increasing either the cost of the programme or the risk of missing an early-stage cancer development. We developed a framework for assessing an individual's risk of cervical cancer based on their available screening history and computing a primary risk factor called CRS from a data-driven separation model together with multiple derived attributes. The results show that this approach is highly practical, validates against multiple established trends, and can be effective in personalizing the screening needs for individuals.
Prediction of Cervical Cancer using Voting and DNN Classifiers Komala Rayavarapu ; Kishore K.V. Krishna .
Cervical tumour is one of the far-reaching sicknesses among ladies in India and around the world. Early analysis is good for better treatment, yet due to vulnerability in detecting cancer cells becoming complex. Machine Learning (ML) systems can be used to create tools for doctors in the conclusion of cervical growth, which will enormously upgrade the survival rate of patients. In this work, two popular ML techniques like Voting Classifier and Deep Neural Network (DNN) Classifier are used to predict cervical growth. The cervical cancer (risk factors) datasets from UCI store was utilized as a preparation set to assess and look at the execution of the two ML classifiers in terms of accuracy and other false rate measures. Outcomes acquired in this paper give how effectively ML systems predict cervical malignancy.
Computer Based Automatic Segmentation of Pap smear Cells for Cervical Cancer Detection Anupama Bhan ; Divyam Sharma ; Sourav Mishra .
Cervical Cancer is the fourth leading cause of death due to cancer among women worldwide. Pap Smear Test is the commonly used method for Cervical Cancer screening. But Pap Smear pathology screening is very time-consuming process. Therefore, an automatic detection method of nucleus of cervical cell is proposed in this paper which mainly focuses on time consumption which is an important parameter when it comes the automatic segmentation. The pre-processing is achieved using edge map with double threshold for de-noising of edges, and then segmentation of the nucleus of cervical cancer cell is achieved using Gradient Force Model and Balloon force Model. The two parametric deformable models are used to check the trade-off between the number of iterations and accuracy. Further, geometrical features like perimeter, area, eccentricity, mean intensity etc. are calculated followed by segmentation using both methods to detect whether cell is cancerous or normal. The calculated features are contrasted with each method. The experimental results show time consumption is reduced using gradient force model in terms of number of iterations used for segmentation with the accuracy of 0.92 which is significant for clinical interpretation.

EXISTING SYSTEM

Support Vector Machine (SVM):

To detect an ideal hyperplane for different distinct examples in a high dimensional space is the main process of the SVM. To fulfill this model there is more than one hyperplane. This process depends upon the bolster vector which the information that lies nearest on the closed surface and coordinating with the ideal choice surface. It performs classification by planning the input vectors into a high dimensional space and constructing the hyperplane to separate the data. This strategy is mainly used to solve a quadratic programming problem and non-convex, unconstrained minimization problem. The SVM is the most effective method in the classifier process.

DISADVANTGES OF EXISTING SYSTEM

- Only edges are been detected.
- Time consumption.
- More complexity.
- Less efficiency.
- Noise is high.

PROPOSED SYSTEM

Convolution Neural Network (CNN):

In deep learning, a **convolutional neural network** (**CNN**) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as **shift invariant** or **space invariant artificial neural networks** (**SIANN**), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and recognition, recommender systems, image classification, Image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series.

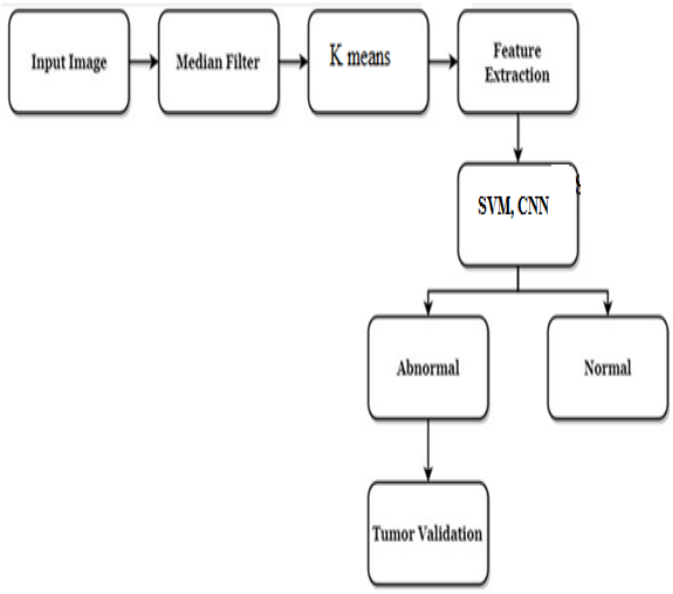
CNNs are regularized versions of multilayer perceptron's. Multilayer perceptron's usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to overfitting data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a

restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

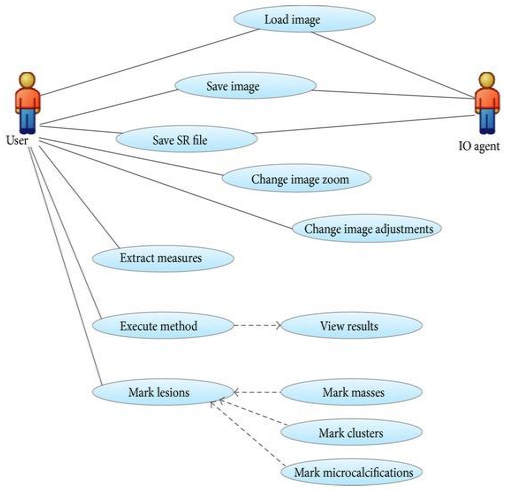
ADVANTAGES OF PROPOSED SYSTEM

- Prediction of tumor is accurate.
- Time consumption is less.
- Noise will be less.
- No Overlapping of images.
- Can be implemented to all images

SYSTEM ARCHITECTURE



USE CASE DIAGRAM



SYSTEM REQUIREMENTS HARDWARE REQUIREMENTS

- Processor : Core i3/i5/i7
- RAM : 2-4GB
- HDD : 500 GB

SOFTWARE REQUIREMENTS

- Platform: Windows Xp/7/8/10
- Coding Language: Python, Machine Learning

MODULE DESCRIPTION

There are 5 components in the system. They are

(i) Image Acquisition
(ii) Image Preprocessing
(iii) Image Segmentation
(iv) Feature Extraction

(v) Classification.

1. Image Acquisition Image

Acquisition is the process of collection of images. These images are downloaded from the online dataset provider called Kaggle.com.

2. Image Pre-processing

Image pre-processing includes converting normal images into resize images. Grayscale images have the combination of black and white. Grayscale images help to reduce noise and also make the background neutral. It also helps to improve brightness of the image. Data augmentation is a way of creating new data which has benefits like the ability to generate more data from limited data and it prevents over fitting.

3. Image Segmentation

Image segmentation breaks the image down into meaningful regions. It divides digital image into multiple segments. The goal is to simplify or change the representation into more meaningful image. It differentiates between the objects we want to inspect further and the other objects or their background. It consists of segmenting the converted grayscale images using K means segmentation.

4. Feature Extraction

Feature extraction is extracting or showing of the segmented portion of the image so that classification becomes easy. Features are extracted in order to differentiate between the images. Features extraction is used in almost all machine vision algorithms. The common goal of feature extraction and representation techniques is to convert the segmented objects into representations that better describe their main features and attributes.

5. Classification

Here we use the concept of different classifiers for classification method. The last module includes the classification in which Tensor Flow and Machine Learning algorithm will be used. Tensor Flow is a MATLAB-friendly open-source library for numerical computation that makes machine learning faster and easier. Tensor Flow allows developers to create dataflow graphs - structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor.

REFERENCES

1. Sarvestan Soltani A, Safavi A, Parandeh M N and Salehi M, “Predicting Cervical Cancer Survivability using Data Mining Techniques”, IEEE 2019.
2. Software Technology and Engineering (ICSTE), 2nd International Conference, Vol.2, pages 227-231,2019.
3. Werner J C and Fogarty T C, “Genetic Programming Applied to Severe Diseases Diagnosis”, In Proceedings Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP), 2019.
4. Iran pour M, Almassi S and Analoui M, “Cervical Cancer Detection from fan using SVM and RBF Classifier”, In 1st Joint Congress on Fuzzy and Intelligent Systems, 2019.
5. Joachims T, Scholkopf B, Burges C and Smola A, “Making large-scale SVM Learning Practical, Advances in Kernel Methods-Support Vector Learning”, Cambridge, MA, USA, 2019.
6. Soman K P, Loganathan R and Ajay V, “Machine Learning with SVM and Other Kernel Methods”, PHI, India, 2019.
7. Crammer Koby and Yoram Singer, “On the Algorithmic Implementation of Multi-class Kernel based Vector Machines”, Journal of Machine Learning Research, MIT Press, Cambridge, MA, USA, Vol.2, pages 265-292, 2019.
8. Riccardo Poli, William B and Langdon Nicholas F McPhee, “A Field Guide to Genetic Programming”, Lulu Enterprises, UK, 2019.
9. Markus Brameier and Wolfgang Banzhaf, “A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining”, IEEE Transactions on Evolutionary Computation, Vol.5, Issue No.1, 2019.
10. Hussein A Abbas. An evolutionary artificial neural networks approach for cervical cancer diagnosis, Artificial Intelligence in Medicine, 25(3):265{281, 2019.
11. Pedro Henriques Abreu, Hugo Amaro, Daniel Castro Silva, Penousal Machado, and Miguel Henriques Abreu, Personalizing cervical cancer patients with heterogeneous data, In The International Conference on Health Informatics, pages 39-42. Springer,2019.