International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 4, July- August 2017 ISSN 2278-6856

Machine learning Technique for detection of Cervical Cancer using k-NN and Artificial Neural Network

Priyanka K Malli , Dr. Suvarna Nandyal

[1]Department of Computer Science & engineering , PDA Engineering college Kalaburag i, Karnataka ,India

[2]HOD Department of Computer Science & engineering , PDA Engineering college Kalaburag i, Karnataka ,India

Abstract

Cervical cancer along with micro classification are one of the 2 major forms of cancer being observed amongst women across the globe. A cervical cancer results in dead nucleus or change in the morphology of the cells in the cervix. Such cells may have multiple nucleuses ,faulty cytoplasm ,lack of cytoplasm ,dissolved lack of cytoplasm and so on. Detection of cervical cancer in a microscopic smear Test (fluid taken from the cervix)smear is analyzed to microscope is extremely challenging because such cells does not offer significant color or texture variations from the normal cells .Therefore high level Digital Image Processing technique are required identify abnormalities in human cell related cancer detection system. Therefore an automated, comprehensive machine learning technique has been proposed in this work. The proposed technique gives that color and shape features of nucleus and cytoplasm of the cervix cell. The nucleus and the cytoplasm are separated from the cell use the advanced fuzzy based technique. KNN and Neural network are trained with the shape features and color features of the segmented units of the cell and then an unknown cervix cell samples are classified by this technique. The classification result have shown an accuracy of 88.04% for KNN and 54% for ANN. The proposed system work can be further enhanced by taking other classifiers.

Keywords: Pap smear Images, Feature Extraction, KNN, ANN

1. INTRODUCTION

Cervical cancer is the second most common type of cancer that affects women, ranked after the breast cancer .The acceptability of increased cervical cancer risk and cause of cancer death projected. Cervical screening program has minimize the rate of death in developed countries. Cervical cancer is the one of the deadliest disease, it can be cured if detected in early stage.

Pap Smear Test this was introduced as a screening test for cervical cancer in the year of 1943 by Dr. George Papanicolauname of the scientist who proposed a method where a spoon like wooden instrument is used to take out a smear to use we need a microscope. Our technique first segments each independent cells in the microscopic image .One or more cells may cancerous cells or non of the cells may be cancerous or only one out of the many cells may be cancerous or all the cells present in the microscopic image may be cancerous. Therefore regression based technique can't separate the cancers cells from the normal cells therefore our technique should effectively separate every cell present in the image and then extract the features .And the morphology descriptor for each of the cell presents that information to a machine learning system which farther classifies the normalcy or the presence of cancer from each of the cells if one or more cells present in the cervical smear sample has cancer then we say the presence of the cervical cancer.

Fig 1a: Normal cervix cell

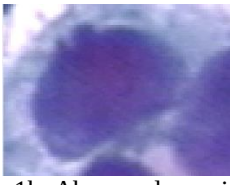

Fig 1b. Abnormal cervix cell.

2.RELATED WORK

The work related to problem are outlined as follows
[1] Authors have made a survey and analysis on the different types of detection of cervical cancer which is useful for the classification of normal and abnormal cervical cells [2]automated diagnostic system using image processing techniques have been proposed which proves to be an aid to radiologist and play a key role in early detection of cancer using MRI of cervical cancer using SVM and ANN. [3] The authors proposed the computer assisted pap smear Analyzer for cervical cancer screening using quantitative microscopy for detection of normal and abnormal cancer cells .[4] In this the cytology images are taken to segment the nucleus and cytoplasm for detection of cervical cancer [5] also two level cascade classification is used to classify the images captured from thin liquid based cytology slides. [6]And author shown that comparison of classification techniques for pap smear Diagnosis, The classification for pap smear diagnosis aims at classifying the pap smear cells whether it is affected or not.[7] The author proposed semi-automatic segmentation and classification of pap smear cells. The proposed system provides a feasible and effective tool in evaluating cytological specimens. [8]Diagnosis method and multi-thresholding ,SFS ,SFFS is used for detecting cervical cancer .[9]fully automated method is used for the author

uses FCM and SVM classification technique and clustering algorithm. [10] gives the revive of various algorithms and technique for diagnosis of essential features for the classification of cervical cancer in to different stage of Normal, LSIL, HSIL. [11] Using Fuzzy moving and K-Means algorithm and Fuzzy Min-Max neural network based system for classification of normal, LSIL, HSIL.[12] And Pap smear Classification is Explained

3.MATERIALS AND METHODS

The proposed system is works based on MATLAB 2016 .And it deals with implantation of an automated system which helps us to detect the cervical cancer at the cell level using Digital Image Processing technique

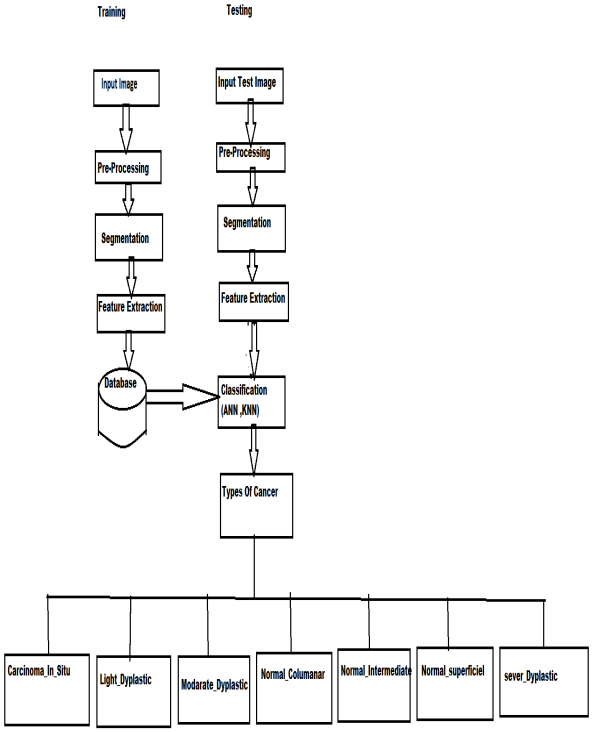

Figure 2: Block diagram of the Proposed system for detecting cervical cancer

Initially the process includes color conversion for RGB to Gray scale , Enhancement and scaling, segmentation, detecting the cell boundary and detecting the overlapping cytoplasm and nuclei, feature extraction and classification of cells based on the features. Most discriminating set of features were selected and used as input to an artificial neural network (ANN) and KNN. The shape features such as roundness, area, width, height ,texture features and color features are extracted. Both features are trained by ANN and KNN classifier.

2.1. Segmentation

The segmentation is based on the Fuzzy c mean clustering technique.

Fuzzy c- mean clustering

This is one of the data clustering technique which allows one piece of data to belong to two or more clusters. There are two clusters nucleus and cytoplasm .Segmentation is the fundamental step of image analysis it separates the objects form the background to obtain the region of **interest** so as to get the features .Many types of methods are there among those we are using thresholding method in this work .

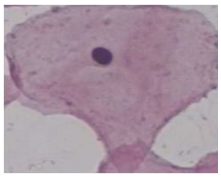

Fig.3.a: Normal cell

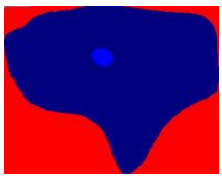

fig.3.b: Segmented cell

2.2. Feature Extraction

Feature extraction is transforming the input data into the set of features .In this Work there are many features are taken in to consideration for detecting the cancerous cell from the sample pap smear image.

Nucleus and Cytoplasm Morphology

In this we are taking color features and shape features of the nucleus and cytoplasm. The features of nucleus and cytoplasm such as area, centroid ,size of the nucleus ,shape of the nucleus ,nucleus cell distribution are extracted using region properties. The area of the nucleus is the number of pixels inside the nucleus .In the training phase the area of the nucleus of the normal cells and abnormal cells are calculated . The small areas remaining in the picture which might be looking like a nucleus are removed first .negative of the segmented image is more suitable as it aids in removing unwanted clusters using morphological erosion faster than the true image .The resultant image has nucleus region along with small suspicious objects with can be removed by performing erosion can be removed by performing erosion operation. The eroded image is subtracted from the negative image to get counter of the segmented objects as shown in the figure. After erosion the resultant image consists of nucleus region along with other regions. It is required to select only the nucleus region by some criteria. From this it is observed that the nucleus is bigger region than the other regions. Hence the next step is the labeling the objects followed by selecting the bigger cluster of pixels among the various clusters. After extracting features from the sample image, we apply thesholding for the extracted features. The threshold is calculated using these following color features and shape features

Nucleus area and Cytoplasm area : Calculated by counting the corresponding pixels of the segmented picture. A pixels area is

$$(0.201\mu m)^2$$

N/C ratio : Tells how small the nucleus area is compared to the area of the cytoplasm. It is given by:

$$N/C = \frac{Nucl_{area}}{Nucl_{area} + cyto_{area}}$$

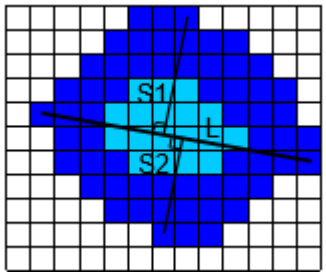

Figure 4: A binary picture of a cell with background(white),cytoplasm(dark blue) and nucleus(light blue). Longest diameter line(L) and shortest diameter lines(s1 & s2) is shown.

Nucleus and Cytoplasm brightness: Nucleus and Cytoplasm brightness is calculated as the average perceived brightness, that is a function of the colors wavelength. In our case it is calculated as

$$Y = 0.299.Red_{\mu} + 0.587.Green_{\mu} + 0.114.Blue_{\mu}$$

intensity for each of the colors. They are weighted by the perceived brightness of the human eye.

Nucleus and Cytoplasm longest diameter: This is the shortest diameter a circle can have, when surrounding the whole object. It is measured as the biggest distance between to pixels on the objects border, and forms a line L as show in figure 4 for the cytoplasm. The names is Nlong and Clong for nucleus and cytoplasm, respectively.

Nucleus and Cytoplasm shortest diameter: This is the biggest diameter a circle can have, when the circle is totally encircled of the object. This distance is approximated by the sum of the two lines s1 & s2 shown in figure 4. for the cytoplasm. They are perpendicular to the line L, and is the longest line to each side sitting inside the object. The names of the features is Nshort and Cshort for nucleus and cytoplasm, respectively.

Nucleus and Cytoplasm elongation: The elongation is calculated as the ratio between the shortest diameter and the longest diameter of the object.

$$N_{elong} = N_{short}/N_{long}$$
$$C_{elong} = C_{short}/C_{long}$$

Nucleus and Cytoplasm roundness: The roundness is calculated as the ratio between the actual area and the area bound by the circle given by the longest diameter of the object. They are here named Nroundness and Croundness respectively:

$$N_{circle} = \frac{\pi}{4} \cdot N_{long}^2 \Rightarrow N_{roundness} = N_{area}/N_{circle}$$
$$C_{circle} = \frac{\pi}{4} \cdot C_{long}^2 \Rightarrow C_{roundness} = C_{area}/C_{circle}$$

Nucleus and Cytoplasm perimeter: The length of the perimeter around the object.

Nucleus position: This is a measure of how well the nucleus is centered in the cytoplasm. It is calculated by nding the distance between the nucleus center and the center of the cytoplasm :

$$N_{Pos} = \frac{2a\sqrt{(x_N - x_C)^2 + (y_N - y_C)^2}}{C_{long}}$$

Here the position is given by the centrum (X_N,Y_N) and (Xc,Yc) for the nucleus and cytoplasm respectively.

Nucleus and Cytoplasm Maxima/Minima: This is a count of how many pixels is a maximum/minimum value inside of a 3 pixel radius.

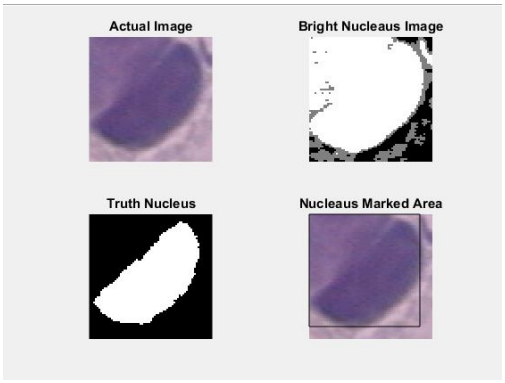

Fig 5: Image segmentation and feature extraction

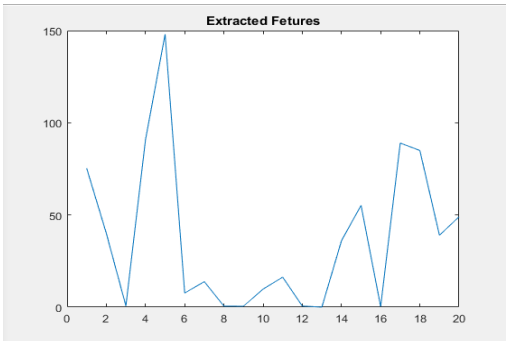

Fig 6: Thresholding the sample image

The threshold is calculated using 20 features.

3 .CLASSIFICATION

Classification is divided in to two steps training and testing .Training is the one where we tell the classifier what are the steps to classify images .In testing we give an input and test the class. Here whatever the features are extracting it consist of a trained matrix .After training feature library is created. The trained image is classified using k-NN and ANN classifier. Based on feature extraction matches the trained image is classified for cancer types.Trained matrix consist of mainly 3 parameters F,C,INFO .Here we are taking the size of F (it consist of all 20 features which we previously trained) 'C ' contains all the class names and

INFO all the file names .After taking all these parameters we are use the classifiers to classify .In this work the classification is made using two classifiers K-nearest neighbor and Neural network methods

3.1 K-nearest neighbor classification

It is the pattern recognition technique used for classification. Here we load the train matrix .Here we have to classify the result of feature of the current image. Against the F and C. It will be going to give us detected class of C If the loaded image shape features and color features are matches with the trained matrix then it shows the type of Cancer.

3.2 Artificial Neural Network

The neural network is simply trained by loading the train matrix and then it expects the classes based on vectorized form and based on feature vectors Once we train the neural network .here we are classifying the features which are of 20 types. we taking the features as an input layer and the number of classes are taking as the output layers. and the hidden layer consists of number of images trained and it will be mapped with number of layers .As many as we train the images the hidden layers also increases.

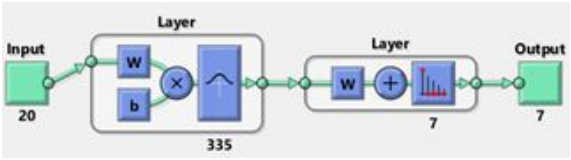

Fig 7: Neural Network Training

In order to test the neural network .we have to load our pre-stored network and we are going to calculate the result by simulating current set of networks. Likewise we are performing the neural network operations

4.Performance measures

Working of segmentation and feature extraction

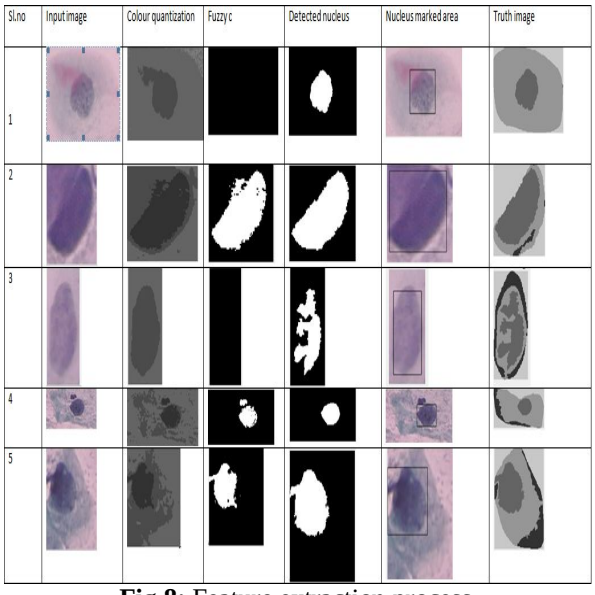

Fig 8: Feature extraction process

4.1 Performance comparison of sample images using k-NN and ANN

Cancer Types	no.images	no.images trainned	no.images tested	K-NN classifier			ANN classifier		
				correct	incorrect	accuracy	correct	incorrect	accuracy
carcinoma_in situ	301	30	40	40	0	100.00%	38	2	95%
Light dyplastic	365	35	65	58	7	89.23%	22	43	34%
modarate dyplastic	295	30	55	40	15	72.73%	34	21	62%
normal columanar	197	30	46	40	6	86.96%	23	23	50%
normal intermediate	141	30	33	30	3	90.91%	22	11	67%
normal superficiel	191	33	56	50	5	89.29%	14	42	25%
sever dyplastic	395	34	70	61	9	87.14%	33	37	47%

Fig 9: performance comparison KNN and ANN

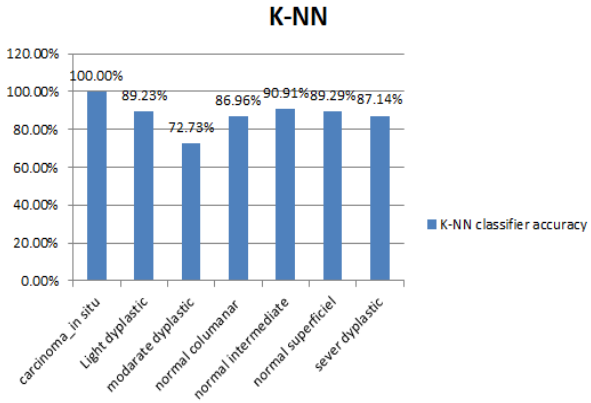

Fig 10: Classification accuracy of Cancer types using k-NN classifier

From above graph it is absorved that based on analysis of shape and color features of images the maximum accuracy of 100% is obtained for carcinoma_in suit cancer types and minimum accuracy of 72.73% is obtained for modarate_dyplastic cancer types.

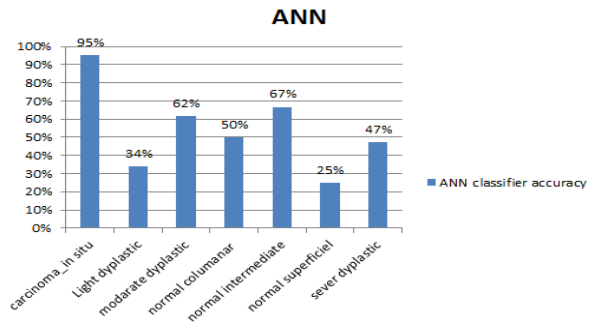

Fig 11: Classification accuracy of Cancer types using NeuralNetwork classifier

From above graph it is absorved that based on analysis of shape and color features of images the maximum accuracy of 95% is obtained for carcinoma_in suit cancer types and minimum accuracy of 25% is obtained for normal_superficiel cancer types.

Total Accuracy of the work

Classifiers	Accuracy
k-NN	88.04%
ANN	54%

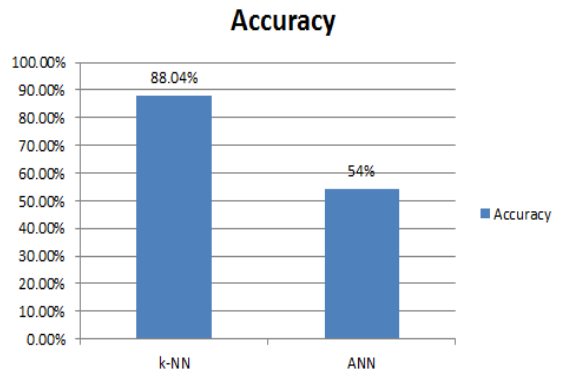

Fig 12: Avarage Performance Accuracy of the work using k-NN and ANN

5. Conclusion and future works

In order to analysis whether pap smear test or cytoplasmic test the results analysis is a cells .which is observed in microscope .our technique first segments each independent cells components such as nucleus and cytoplasm and then detect whether one or more cells are cancerous or not through machine learning based technique. There are several technique which are been proposed in the past towards this direction there accuracy has not been found to be significantly accessible .In this work we have developed a machine learning based technique coupled with efficient Fuzzy based segmentation to accurately segment and classify the cell in to normal or cancerous cell and its types . In our work the Average classification result have shown an accuracy of 88.04% for k-NN and 54% for ANN.

The proposed system work can be further enhanced by taking other classifiers. And also it can be further improved by not only classifying the different types of cells but also at the same time that classify the stage of the cancer such as beginning or malignant .Any detection of the cancer at the beginning stage can help significantly to reduce the risk of life by the patient

REFERENCES

[1] M Anousouya Devi , S.Ravi , J.Vaishanavi and S.Punit ,"Classification of cervical cancer using Artificial Neural networks" Twelfth international Multi conference on information Processing(2016) pp. 465-472

[2]Aabha Phatak , Dr Ashwini.B "Classification of MR Images of Cervical Cancer Using SVM and ANN" International Journal of scientific research(2015) pp.3-6

[3]Rajashekar Usha Deepak,Ramakrishnan Rajesh Kumar, Nindothadakhal Balakrishnan Byju, Pundluvalu Nataraju, Chandran Pournami "Computer assisted Pap Smear Analyzer for Cervical cancer screening Using Quantitative Microscopy" Journal of cytology and Histology(2015) pp.1-6

[4]Zhi Lu,Carnio,Andrew P Bradley,Daniel Ushizima, Masoud S.Nosrati,Andrew Informatics vol. XX,NO.XX, September 2015 pp.4 G.Cbainchi, C.M.Carneiro and Ghassan Hamarneh, "Evaluation of Three Algorithms for the segmentation of overlapping Cervical cancer cells "Journal of Biomedical and Health -11

[5]Jio su,Xuan Xu,Yongjun He,and Jiniming Song "Automatic detection of cervical cancer cells by a Two level cascade classification "Hindwai Publishing corporation Analytic cellular Pathology volume 20016,Article ID9535027 pp.1-11

6]P.Soorya Prabha,R.Priya "Comparison of classification techniques for pap smear Diagnosis" IOSR Journal of Computer Engineering e-ISS:2278-0661,p-ISSN:2278-8727 ,pp11-18

[7]Yung-Fu Chen ,Po-Chi Huang, Ker-cheng Lin,Hsuan-Hung Line,Lin-en Wang,Chung-Chung Cheng,Tsung-Po cheng,Yung-Kuan Chan,and Jhan Y.Chiang "semi Automatic Segmentation and Classification of Pap Smear Cells"IEEE Journal of Biomedical and Health Informatics ,vol 18,No.1January 2014 pp.94-108

[8]B.Ashok, Dr.Aruna "comparison of Feature selection methods for Diagnosis of Cervical cancer using SVM classifier" Journal of Engineering Research Applications IISN:2248-9622,vol.6,Issue 1,(Part-1) January2016 pp.94-99

[9]Marina R,Plissiti, Christophoros Nikou, Member, and Antonia Charchinti "Automated detection of cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering IEEE Journal of Transaction on Information Technology in Biomedicine ,vol.15,No 2 March 2011 pp.233-241

[10].Chandraprabha R,Dr.Seema Singh"Artificial Intelligent system for Diagnosis of cervical cancer :a brief review and future"International Journals of Latest Research in Engineering and Technology ISSN:2454-5031 pp.38-41 (2016)

[11]Anas Quteishat,Mohammad Al-Batah,Anwar Al-Mofleh, " Cervical cancer diagnostic system using adaptive FUZZY moving K- MEANS Algorithm and Fuzzy Min-Max neural network"Journal of Theoretical and Applied Information Technology 10th November vol.57 no.1 pp.48-53

[12]Erik Martin "Classification of Pap smear images" Technical University of Denmark

AUTHORS

Priyanka K Malli Recieved B.E degree in computer scince and engineering from Viswesvaraya Technological University Belgaum Karnataka India in the year of 2014. And at present studying M.Tech in computer science and engineering from PDA Engineering college kalaburagi,karnataka ,India.

Dr.Suvarna Nandyal Recieved B.E and M.Tech degree In Computer science and Engineering from visvesvaraya Technological University Belagaum,Karnataka India.and Recieveed Ph.D from Jawaharlal Nehru Technolgical University Hyderabad. At present working as Head of the Department at PDA Engineering college Kalaburagi Karnataka,India