



Review

Methods for Detecting Cervical Cancer via Machine Learning with k-NN and Artificial Neural Networks

Priyanka K. Malli*; Suvarna Nandyal

Department of Computer Science and Engineering, PDA Engineering College Kalaburagi, Karnataka, India

*Corresponding author

Priyanka K. Malli

Department of Computer Science and Engineering, PDA Engineering College Kalaburagi, Karnataka, India

Article information

Received: January 3rd, 2023; Revised: February 19th, 2023; Accepted: March 22nd, 2023; Published: April 12th, 2023

Cite this article

Malli PK, Nandyal S. Methods for detecting cervical cancer via machine learning with k-NN and artificial neural networks. 2023; 2(1).

doi: <https://doi.org/10.70705/ppp.bioai.2023.v02.i01.pp23-26>

ABSTRACT

Among women worldwide, cervical cancer and micro classification are two of the most common types of cancer. When cervical cancer develops, the cells in the cervix either lose their nucleus or undergo morphological changes. Some of the characteristics of these cells include an abnormal number of nucleuses, damaged or absent cytoplasm, or dissolving cytoplasm, among others. Cervical cancer cells do not differ much in texture or color from normal cells, making it very difficult to detect the disease in a microscopic smear test (cervix fluid studied under a microscope). Hence, in order to detect anomalies in cancer detection systems involving human cells, advanced digital image processing techniques are necessary. Thus, this study proposes an automated, all-encompassing machine learning method. The suggested method reveals the cervix cell's nucleus and cytoplasm's form and color. The cell's nucleus and cytoplasm are isolated using a state-of-the-art fuzzy-based method. This method involves training a KNN or Neural Network using the form and color attributes of the segmented cell units. Then, unknown cervical cell samples may be categorized using these networks. While ANN achieved a classification accuracy of 54%, KNN achieved an impressive 88.04%. Including more classifiers may improve the performance of the suggested system even further.

Keywords

Pap smear images; Feature extraction; KNN; ANN.

INTRODUCTION

After breast cancer, cervical cancer is the most frequent malignancy affecting women. Whether or whether the anticipated rise in cervical cancer risk and cancer mortality rate is acceptable. Developed nations have seen a decline in cervical cancer mortality rates due to widespread screening programs. Although it is curable if caught early, cervical cancer is among the worst diseases in the world.

The Pap smear test was first suggested in 1943 by Dr. George Papanicolaou, a scientist. It involves using a spoon-like wooden device to remove a smear, which can then be examined under a microscope. This test has been used as a screening tool for cervical cancer ever since. As a preliminary step, our method separates out individual cells in the micrograph. Cancer may develop in one or more cells, none of the cells, or just a single cell among many.

malignant, or maybe every single cell in the micrograph is malignant. Accordingly, because regression-based techniques fail to distinguish between cancerous and noncancerous cells, our method must suc-

cessfully isolate each and every cell in the picture before extracting any characteristics. In order to determine if a cervical smear sample contains cancer, we use the morphology descriptor for each cell to feed that data into a machine learning system. The system then uses that information to classify the cells as normal or cancerous.

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.

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 from 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.

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 which can be removed by performing erosion. 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 thresholding 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

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 two pixels on the object's border, and forms a line L as shown in figure 4 for the cytoplasm. The names are N_{long} and C_{long} 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 by the object. This distance is approximated by the sum of the two lines s_1 & s_2 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 are N_{short} and C_{short} 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. 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 $N_{roundness}$ and $C_{roundness}$ respectively.

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

Here the position is given by the centrum (X_N, Y_N) and (X_C, Y_C) 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.

3. CLASSIFICATION

Training and testing are the two main phases of classification. When we train a classifier, we walk it through the process of picture classification. During testing, we test the class by providing it with an input.



A trained matrix is what is used to extract features in this case. There will be a training feature library. Both k-NN and ANN classifiers are used to categorize the learned picture. The trained picture is used to classify cancer kinds based on feature extraction matches. F, C, and INFO are the three primary parameters that make up a trained matrix. In this case, we are calculating the size of F, which includes all twenty features that were trained before. "C" includes every name of a class and INFO every name of the file. Using the classifiers for classification follows the completion of all these parameters. Here, we use two classifiers—K-nearest neighbor and neural network—to make our classifications.

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.

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

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

From above graph it is absorbed 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 moderate_dyplastic cancer types.

From above graph it is absorbed 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_superficial cancer types.

Total Accuracy of the work

5. Conclusion and future works

The findings of a cytoplasmic or pap smear test are analyzed by

looking at cells under a microscope. First, our method separates the nucleus and cytoplasm of each cell. Then, using a machine learning-based approach, we can determine whether a cell or cells are malignant. Despite several previous proposals, no method has proven to be very accessible in terms of accuracy in this area. This study presents a method for properly segmenting and classifying cells as normal or malignant, as well as their kinds, by combining machine learning with efficient fuzzy based segmentation. The average classification results for k-NN in our study were 88.04% accurate, whereas ANN only managed 54%.

Including more classifiers may improve the performance of the suggested system even further. Additionally, it has room for improvement by simultaneously identifying the various cell types and cancer stages, such as early or advanced malignant. There is a strong correlation between early cancer identification and a much reduced risk of mortality for patients.

REFERENCES

[1] Using Artificial Neural Networks for Cervical Cancer Classification [1] M. Anousouya Devi, S. Ravi, J. Vasishanavi, and S. Putri Volume 12, Issue 4, Pages 465–472, 2016 International Conference on Multi-Conference on Information Processing

1. "Classification of MR Images of Cervical Cancer Using SVM and ANN" by Dr. Ashwini B. Phatak and Aabha Phatak Pp. 3-6 in the 2015 edition of the International Journal of Scientific Research

Rajashekar said Deepak Usha and Ramakrishnan Chandran Pournami, Pundluvalu Nataraju, Rajesh Kumar, and Nindothadakhil Balakrishnan Byju "Automated Pap Smear Analyzer with Computer Assistance for Quantitative Microscopy-Based Cervical Cancer Screening" Journal of Cytology and Histology(2015), pp.1-6.

[4] In: Informatics, volume XX, issue XX, September 2015, authors Zhi Lu, Carnino, Andrew P. Bradley, Daniel Ushizima, Masoud S. Nosrati, and Andrew The essay "Evaluation of

Cervical cancer cell overlap segmentation using three different algorithms Scientific and Medical Journal -11

[5] "Automatic detection of cervical cancer cells by a Two level cascade classification" (Jio su, Xuan Xu, Yongjun He, and Jinming Song) Analytical cellular pathology, volume 20016, article ID9535027, Hindwai Publishing Corporation, pp. 1-11.

"Comparison of classification techniques for pap smear diagnosis" published in the IOSR Journal of Computer Engineering, page 6, was written by P. Soorya Prabha and R. Priya. 2278-8727, pages 11–18

The article "semi Automatic Segmentation and Classification of Pap Smear Cells" was published in the January 2014 issue of the IEEE Journal of Biomedical and Health Informatics and was



co-authored by 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.

[8]Dr. Aruna, B. Ashok “SVM classifier for feature selection in cervical cancer diagnosis: a comparison of methods” Volume 6, Issue 1, Part 1 of the Journal of Engineering Research Applications (ISSN: 2248-9622) Pages 94–99, January 2016

“Automated detection of cell nuclei in Pap smear images utilizing morphological reconstruction and clustering” [9] by Marina R. Plissiti, Christophoros Nikou, Member, and Antonia Charchinti Sections 233–241 of the March 2011 issue of the IEEE Journal of Transactions on Information Technology in Biomedicine

Dr. Seema Singh and Chandrababha R.A concise overview and

future directions for artificial intelligence systems used in cervical cancer diagnosisJournal of the Most Recent Findings in Engineering and Technology, International Serial Number: 2454-5031, Pages: 38–41 (2016)

Using an adaptive FUZZY moving K-MEANS algorithm and a fuzzy min-max neural network, Anas Quteishat, Mohammad Al-Batah, and Anwar Al-Moffeh developed a cervical cancer diagnosis system.Publication date: November 10th, volume 57, issue 1, pages 48–53, Journal of Theoretical and Applied Information Technology

“Classification of Pap smear images” by Erik Martin [12]. The Danmark Technical University