International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

A Novel Approach for Breast Cancer Prediction

Ms. Komal R. Hausalmal and Prof. Jyoti P. Kshirsagar[2]

Department of Computer Engineering JSPM's Rajarshi Shahu College of Engineering Pune,India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

The classification of breast cancer has been the subject of interest in the fields of healthcare and bioinformatics, because it is the second main reason of cancer-related deaths in women. Breast cancer can be analyzed using a biopsy where tissue is eliminated and studied under microscope. The identification of problem is based on the qualification and experienced of the histopathologists, who will attention for abnormal cells. However, if the histopathologist is not well-trained or experienced, this may lead to wrong diagnosis. With the recent proposition in image processing and machine learning domain, there is an interest in experiment to develop a strong pattern recognition based framework to improve the quality of diagnosis. In this work,the image feature extraction approach and machine learning approach is used for the classification of breast cancer using histology images into malignant. The preprocessing on the image is done using histopathological image after that apply feature extraction and classify the final result using SVM and Naive Bayes Classification techniques.

Keywords: *Histopathological image classification, breast cancer diagnose, feature extraction, SVM classification, NaiveBayes Classification;*

Introduction

Breast cancer is the most common and dangerous intrusive cancer in women and the second main effect of cancer death in women, after lung cancer. The International Agency for Research on Cancer (IARC), which is part of the World Health Organization (WHO), the numbers of deaths reasoned by cancer in the year of 2012 only come to around 8.2 million. The number of new cases is expected to growth to more than 27 million by 2030.Finding breast cancer quick and getting state-of-the-art cancer treatment are the key plan of action to avoid deaths from breast cancer. In existing, it is a widely-used way to identification of breast cancer by identifying hematoxylin and eosin (HE) stained histological slide preparations that are checked under a high powered microscope of the changed area of the breast. In medical practice, classification of breast cancer biopsy result into different plans (e.g. cancerous and noncancerous) is manually driven by experienced pathologists. Come out machine learning approaches and enlarging image volume developed automatic system for breast cancer classification possible and can help pathologists to obtain precise identification of problem more efficient. Breast cancer can be find or identified using medical images testing using histology and radiology images. The radiology images search can help to find the areas where the difference is located. However, they cannot be used to find or identified whether the area is cancerous. The biopsy, where a tissue is gives as input and processed under a microscope to see if cancer is present, is the only sure way to find if an area is cancerous. After completing the biopsy, the identification of problem will be based on the qualification of the histopathologists, who will analyze the tissue under a microscope, looking for exceptional or cancerous cells. The histology images allow us to differentiate the cell nuclei types and their flowchart according to a specific pattern. Histopathologists particularly examine the consistency of cell shapes and tissue distributions and decided the cancerous regions and malignancy degree. If the histopathologists are not well-trained, this may lead to an incorrect identification of problem. Also, there is a lack of specialists, which maintain the tissue sample on hold for up to two months. There is also the issue of reproducibility, as histopathology is a subjective science. This is right especially between non-specialized pathologists, where we can get a different identification of problem on the same sample. Therefore, there is an insistent demand for computer-assisted identification of problem.

Review of Literature

Breast cancer (BC) is a savage disease, executing a huge number of individuals consistently. Creating robotized dangerous BC recognition framework connected on patient's symbolism can assist managing this issue all the more effectively, making diagnosis more versatile and less inclined to mistakes. DeCAF (or deep) highlights comprise of an in the middle of arrangement it depends on reusing a formerly trained CNN just as highlight vectors, which is then utilized as contribution for a classifier prepared just for the new order assignment. In the light of this, they display an assessment of DeCaf highlights for BC recognition, with a specific end goal to all the more likely see how they contrast with alternate methodologies [1].

This work proposes to classify breast cancer histopathology images independent of their magnifications using convolutional neural networks (CNNs). They propose two different architectures; single task CNN is used to predict malignancy and multi-task CNN is used to predict both malignancy and image magnification level simultaneously. Evaluations and comparisons with previous results are carried out on BreaKHis dataset [2].

The reason for this work is to create an insightful remote discovery and finding approach for breast disease in light of cytological pictures. Initially, this work exhibits a completely mechanized methodology for cell nuclei recognition and division in bosom cytological pictures. The areas of the cell cores in the picture were identified with roundabout Hough change. The expulsion of false-positive (FP) discoveries (loud circles and platelets) was achieve utilizing Otsu's thresholding procedure and fluffy c-implies grouping strategy. The division of the nuclei limits was proficient with the utilization of the marker-controlled watershed change. Next, an astute breast malignancy grouping framework was created [3].

The effectiveness of the treatment of breast cancer depends on its timely detection. An early advance in the finding is the cytological examination of breast material acquired straightforwardly from the tumor. This work gives in PC supported breast growth recognizable proof of issue in light of the examination of cytological pictures of fine needle biopsies to recognize this biopsy as either benevolent or harmful. Rather than confer on the exact division of cell nuclei, the nuclei are finding by circles utilizing the roundabout Hough change system. The result circles are then sifted to keep just astounding estimations for additionally think about by a help vector machine which groups identified circles as right or wrong utilizing surface highlights and the level of cores pixels as per a cores veil acquired utilizing Otsu's thresholding system [4].

This work direct some fundamental examinations utilizing the deep learning way to deal with arrange breast cancer histopathological pictures from BreaKHis, an openly dataset accessible at http://webinf.ufpr.brivri/bosom malignancy database. They propose a strategy in view of the extraction of picture patches for preparing the CNN and the mix of these patches for definite grouping. This strategy means to permit utilizing the high-goals histopathological pictures from BreaKHis as contribution to existing CNN, maintaining a strategic distance from adjustments of the model that can prompt a more unpredictable and computationally exorbitant engineering [5].

Current methodologies depend on handcraft highlight portrayal, for example, shading, surface, and Local Binary Patterns (LBP) in arranging two areas. Contrasted with carefully assembled include based methodologies, which include undertaking subordinate portrayal, DCNN is a conclusion to-end highlight extractor that might be straightforwardly gained from the crude pixel force estimation of EP and ST tissues in an information driven mold. These abnormal state highlights add to the development of a directed classifier for separating the two kinds of tissues [6].

The test turns out to be the means by which to cleverly join fix level arrangement results and model the way that not all patches will be discriminative. They propose to prepare a choice combination model to total fix level forecasts given by fix level CNNs, which to the best of our insight has not been appeared previously. They apply the technique to the grouping of glioma and non-little cell lung carcinoma cases into subtypes [7].

Computerized atomic identification is a basic advance for various PC helped pathology related picture examination calculations, for example, for mechanized evaluating of breast disease tissue examples. Be that as it may, computerized core location is muddled by (1) the huge number of nuclei and the measure of high goals digitized pathology pictures, and (2) the inconstancy in estimate, shape, appearance, and surface of the individual nuclei. As of late there has been enthusiasm for the utilization of "Profound Learning" techniques for order and investigation of enormous picture information [8].

This work present a dataset of 7,909 breast tumor (BC) histopathology pictures procured on 82 patients, that is currently openly accessible from http://web.inf.ufpr.br/vri breast-cancer-database. The dataset incorporates both benign and malignant pictures. The undertaking related to this dataset is the robotized classification of these pictures in two classes, which would be an important PC helped finding instrument for the clinician. So as to evaluate the trouble of this undertaking, we demonstrate some primer outcomes acquired with state-of-the-art image classification systems [9].

There are a few issues still exist in conventional individual Breast Cancer Diagnosis. To take care of the issues, an individual credit appraisal display in view of help vector order technique is proposed. Utilizing SPSS Clementine information mining device, the individual credit information is bunching investigation by Support Vector Machine. It is investigated in detail with the distinctive part capacities and parameters of Support vector machine. Bolster vector machine could be utilized to enhance crafted by medicinal specialists in the determination of breast growth [10].

Proposed Methodology

Classifying breast cancer histopathological images automatically is an important task in computer assisted pathology analysis. However, extracting informative and nonredundant features for histopathological image classification is challenging.

In our proposed work using Histopathological image, firstly we will apply image pre-processing technique to remove the noise of an image. After that we will apply the feature extraction process. The feature-based approaches consist of the features extraction phase and then classification phase. This approach focuses on extracting the feature of image and classify them using machine learning classification technique. The extracted features are trained using support vector machines and Naive Bayes Classification technique. Finally, we compared the performance using the existing classification methods.

Advantages of Proposed System:

1. Work could be beneficial to obtain fast and precisequantification, reduce observer variability, and increase objectivity.
2. Cell nuclei detection using image thresholding and imageedge detection.
3. We can measure accurate cell features.
4. This application can be used by physicians from theirhomes or any other place.
5. This work will be suitable for images with a high degreeof noise and blood cells and cell overlapping, as it can successfully detect the cell nuclei.

A. Architecture

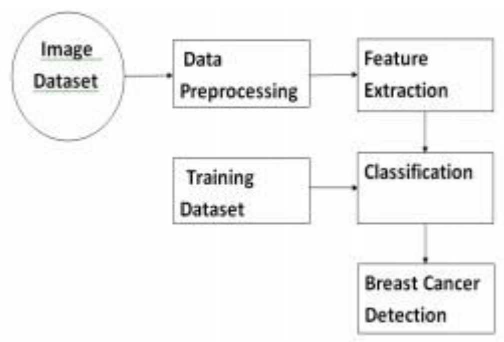

Fig. 1. Proposed System Architecture

Explanation:

Input of system:

In this proposed system, we take histopathological breast image as an input for processing.

Image Pre-processing:

In this step, check the size of input image and then the input image is converted into grayscale image. Also, we remove the noise of image using noise reduction technique that i.e.

here we use the median filter for noise reduction.

Feature Extraction:

In this step, after image preprocessing, we extracts all feature of preprocessed image i.e. infected and healthy cell nuclei.

Classification:

In this step, after image feature extraction, we classify the infected and healthy cell nuclei using support vector machine and also na¨ıve bays classification technique. Result:

This step displays the final breast cancer result.

B. Algorithms

1. Support Vector Machine:

Support Vector Machine (SVM) is used to classify the fruit quality. SVM Support vector machines are mainly two class classifiers, linear or non-linear class boundaries.

The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.

The task is to train the machine with known data and then SVM find the optimal hyper plane which gives maximum distance to the nearest training data points of any class Steps:

Step 1: Read the test image features and trained features. Step 2: Check the all test features of image and also get all train features.

Step 3: Consider the kernel.

Step 4: Train the SVM using both features and show the output.

Step 5: Classify an observation using a Trained SVM Classifier.

2. Na¨ıve Bays Classification:

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.

The Naive Bayes algorithm is called “naive” because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

The Naive Bayesian classifier is based on Bayes’ theorem with the independence guess between predictors.

A Naive Bayesian model is easy to form, with no critical iterative parameter computation which makes it particularly useful for very large datasets.

Regardless of its simplicity, the Naive Bayesian classifier often does particularly well and is widely used because it often outperforms more experienced classification methods. C. Mathematical Model

1. Mathematical Equations of Support Vector Machine:

We have k sub-spaces so that there are k classification results of sub-space to classifying breast cancer cells, called CL SS1,CL SS2, ..., CL SSk. Thus the problem is how to integrate all of those results. The simple integrating way is to calculate the mean value:

$$CL=\frac{1}{k}\sum_{i=1}^{k} CL\ SS_i \qquad (1)$$

Or weighted mean value:

$$CL=\frac{1}{k}\sum_{i=1}^{k} W_i CL\_SS_i \qquad (2)$$

Where W_i is the weight of classification result of subspace, i.e. breast cancer cells result , SS_i and satisfies:

R

P $W_i = 1$ ———————————-(3)

i=1

The centroid is calculated as follows:

$$\bar{X} = \frac{\sum_{i=0}^{k} X_i}{k}, \bar{Y} = \frac{\sum_{i=0}^{k} Y_i}{k} \text{ —————-(4)}$$

Where $(\bar{X}, \bar{Y})$ represents the centroid of the hand, X_i and Y_i are x and y coordinates of the i^{th} pixel in the hand region and k denotes the number of histopathological image pixels that represent only the hand portion.

In the next step, the distance between the centroid and the pixel value was calculated. For distance, the following Euclidean distance was used:

Distance=$\sqrt{(x2-x1)^2(y2-y1)^2}$ —–(5)

Where (x1, x2) and (y1, y2) represent the two co-ordinate values of histopathological image pixel.

2. Mathematical equation in Naive-Bayes Classification: It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events. Here we will use this technique for breast cancer classification. More formally, Bayes' Theorem is stated as the following equation:

$$P(\frac{A}{B}) = \frac{P(\frac{B}{A})P(A)}{P(B)} \text{ —————(6)}$$

Let us understand the statement first and then we will look at the proof of the statement. The components of the above statement are:

$P(\frac{A}{B})$: Probability (conditional probability) of occurrence of event A given the event B is true

P(A) and P(B): Probabilities of the occurrence of event A and B respectively

$P(\frac{B}{A})$: Probability of the occurrence of event B given the event A is true

A. Dataset

This proposed system use Breast cancer UCI machine learning dataset.Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. A few of the images can be found at [Web Link] Separating plane described above was obtained using Multisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree Construction Via Linear Programming." Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97-101, 1992], a classification method which uses linear programming to construct a decision tree. Relevant features were selected using an exhaustive search in the space of 1-4 features and 1-3 separating planes.

Result and Discussion

Experiments will be done by a personal computer with a configuration: Intel (R) Core (TM) i5-6700HQ CPU @ 2.60GHz, 16GB memory, Windows 8, MySQl Server 5.1 and Jdk 1.8. Some functions used in the algorithm are provided by Opencv2.4.7.

We will use the histopathological breast image data set which consist of total 500 images, collected from the 500 different patient of both healthy and infected. This data set collected from city hospital.

Use of mean and Standard deviation:

Standard deviation (SD) is a widely used measurement of variability used in statistics. It shows how much variation there is from the average (mean). A low SD indicates that the data points tend to be close to the mean, whereas a high SD indicates that the data are spread out over a large range of values.

The mean and standard deviation values of the input image are computed in each spectral channel as the feature. We let n be the number of pixels in the input image, and vij denotes the jth band value of the ith pixel in a image. The mean ($mean_j$) and standard deviation (std_j) of the patch are calculated according to

$$Mean_J = \frac{\sum_{i=1}^{n} V_{ij}}{n} \text{ ———————————(7)}$$

$$Std_j = \sqrt{\frac{\sum_{i=1}^{n} V_{ij} - mean_j^Z}{n}} \text{ —————————(8)}$$

Table I is a summary of classification accuracies among different classifiers based on the feature for classifiers. Note that the Support Vector Machine and Na¨ıve Bayes-based classifier outperform other classifiers. The classification accuracy for SVM and NB is 77.5% and 77.2% on average, respectively.

Table 1 Mean And Deviation Table

Classifier	Mean (Exis. System)	Standard Dev. (Exis. System)	Mean (Prop. System)	Standard Dev. (Prop. System)
Support Vector Machine	72.1	5.8	77.5	7.4
Na¨ıve Bayes Classification	70.3	5.0	77.2	7.1

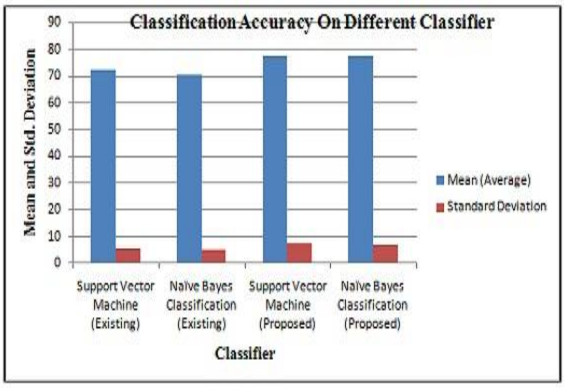

Fig. 2. Mean and Deviation Graph

A. Performance Analysis

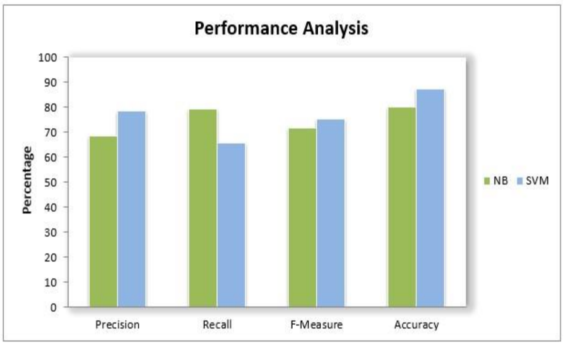

Fig. 3. Accuracy Graph

Table 2 Accuracy Table

	Naïve Bayes	SVM
Precision	68.45	78.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	80.29	87.26

Conclusion

This proposed system work on histopathological images by using Support Vector Machine (SVM) and Naive Bayes Classification with various configurations for the classification of breast cancer histology images into malignant. The designed SVM topology and Naive Bayes Classification worked well on histopathological images features in classification tasks. However, the performance of the SVM classification and Naive Bayes Classification are better compared to the one of the existing classification methods. SVM have become state-of-the-art, demonstrating an ability to solve challenging classification tasks. This proposed work successfully classifies using breast cancer histology images into malignant.

References

[1]. F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, “Deep features for breast cancer histopathological image classification,” 2017 IEEE International Conference

[2]. N. Bayramoglu, J. Kannala, and J. Heikkila, “Deep learning for magnification independent breast cancer histopathology image classification,” 2016 23rd International Conference on Pattern Recognition (ICPR), 2016.

[3]. Y. M. George, H. H. Zayed, M. I. Roushdy, and B. M. Elbagoury, “Remote computer-aided breast cancer detection and diagnosis system based on cytological images,” IEEE Systems Journal, vol. 8, no. 3, pp. 949–964, Sept 2014.

[4]. P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, “Computeraided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies,” IEEE Transactions on Medical Imaging, vol. 32, no. 12, pp. 2169–2178, 2013.

[5]. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks,” in International Joint Conference on Neural Networks, Vancouver, BC, Canada, July 2016, pp. 2560–2567.

[6]. J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, “A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images,” Neurocomputing, vol. 191, pp. 214–223, 2016.

[7]. L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, “Patch-based convolutional neural network for whole slide tissue image classification,” in IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, June 2016, pp.2424–2433.

[8]. J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, “Stacked sparse autoencoder (SSAE) for nuclei detection of breast cancer histopathology images,” IEEE transactions on medical imaging, vol. 35, no. 1, pp. 119–130, 2016.

[9]. F. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, “A dataset for breast cancer histopathological image classification,” IEEE Transactions on Biomedical Engineering, vol. 63, no. 7, pp. 1455–1462, 2016.

[10]. Shang Gao Hongmei Li, “Breast Cancer Diagnosis Based on Support Vector Machine”. 2012 International Conference on Uncertainty Reasoning and Knowledge Engineering

[11]. Su H, Liu F, Xie Y, Xing F, Meyyappan S, Yang L. “Region segmentation in histopathological breast cancer images using deep convolutional neural network.” In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on 2015 Apr 16 (pp. 55-58).

[12]. Wan, S., Huang, X., Lee, H.C., Fujimoto, J.G. and Zhou, C., 2015, April. “Spoke-LBP and ring-LBP: New texture features for tissue classification”. In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on (pp. 195-199). IEEE.

[13]. Levi, Gil, and Tal Hassner. “Age and gender classification using convolutional neural networks.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34-42. 2015.

[14]. M. Veta et al., “Breast cancer histopathology imge analysis: A review,” IEEE Transactions on Biomedical Engineering, vol. 61, no. 5, p. 1400, 1411 2014.

[15]. C. Desir et al., “Classification of endomicroscopic images of the lung based on random subwindows and extra-trees,” IEEE Transaction on Biomedical Engineering, vol. 59, no. 9, pp. 2677–2683, 2012