Volume of Breast Cancer Microarray Image Early Disease Prediction using Machine Learning

Menda Sravani[1] and Dr. R. N. V. Jagan Mohan[2]

Assistant Professor[1] and Associate Professor[2]

ISTS Womens Engineering College[1] and SRKR Engineering College[2]

Abstract - **To store the pixels data that is enormous in gauge is the word gigantic pixels data is used. Gigantic data normally utilized in the analysis of diseases today in healthcare. The kind of cancer detected in a woman is breast cancer. Breast cancer through machine learning methods related to cancer diagnosis. Ensuing a short-lived gestalt of machine learning concepts of data pre-processing methods, feature assortment methods and classification procedures, we designate three specific case studies on the risk of cancer, cancer recurrence and cancer survival based on the popular approach to the machine Learning. Evidently, there are a huge number of machine learning studies; it can provide precise result to concern the exact predictive cancer results. However, identifying potential defects are including experimental design, collection of appropriate data samples, and validation of classified results is critical to the extraction of clinical decisions. This paper uses statistical methods with machine learning approach is on Breast Cancer for early disease prediction using Expectation Maximization with Gaussian Mixture Model.**

Keywords - Breast Cancer, Expectation Maximization, Gaussian Mixture Model, Naivy Bayes.

INTRODUCTION

Over the past periods, there has been a continuous development related to cancer study. Research scientists have used various methods such as early stage screening to detect cancer kinds before causing indications. In addition, they have developed new strategies for predicting the outcome of cancer treatment. With the coming on of new technology in the medical field, large amounts of cancer data have been collected and made available to the medical research community. However, accurate assessment of the disease outcome is a highly interesting and challenging task for clinicians. As a result, machine-learning techniques have become a popular tool for medical researchers. These methods can detect and identify patterns and relations between them from difficult datasets, but they can also effectively predict future results of the cancer kind.

One of the cancer kinds is Breast cancer is a disease in which the cells in the breast grow un controllably[1]. There are different kinds of breast cancer. The kind of breast cancer depends on which cells in the breast become cancerous. Breast cancer begins in different parts of the breast[3]. The breast is made up of three main parts namely as the lobules, the vessels, and the connective tissue. Lobules are milk-producing glands. The vessels are the tubes that carry milk to the nipple[4]. The connective tissue, which contains fibrous and adipose tissue surrounds and holds everything together. Most breast cancers begin in the vessels or lobules. Breast cancer spreads outside the breast through blood vessels and lymphatic vessels. When breast cancer has spread to other parts of the body, it is said to be have metastasized[5,6]. There are two kinds of Breast Cancer namely as invasive ductal carcinoma and invasive lobular carcinoma. Invasive ductal carcinoma is cancer cells grow in other parts of the breast tissue outside the vessels. Invasive cancer cells can spread to other parts of the body or even metastasize. Invasive lobular carcinoma is cancer cells spread from the lobules to the breast tissue. These invasive cancer cells can spread to other parts of the body[7].

This paper is organized is as follows are categorized in below the section-1 deals with Breast Cancer using Naivy Bayes Classification. Dissection of Microarray Breast Cancer image in section-2. The statistical measures with expectation maximization utilized the objects of microarray image in section-3 and section-4. Section-5 procedure of early detection is used. Finally experimental results are in section-6.

1. NaiveBayes of Microarray Brest Cancer Image Classification: Naïve Bayes is one of the simplest classifiers and works surprisingly well for Microarray image processing particularly those involving text classification [1]. Given a microarray image pixels record point M to classify to the general approach is to output that class C_i whose

probability of occurrence P(C_i|M) is maximum. To predict the disease value of P(C_i|M), this classifier naively take on that the pixel attributes of M are sovereign of each other, therefore it is recognized as Naïve Bayes[9]. As soon as individuality has been expected that the origin be used to calculate P(C_i|M) as follows:

P (C_i|M) = P (M ∩ C_i)/ P(M)

$\qquad$ = P(M | C_i) P(C_i) /P(M)

$\qquad$ α P(M | C_i) P(C_i)

$\qquad$ α P(P_1 = m_1| C_i) P(P_2 = m_2| C_i) P(P_2 = m_2| C_i) … P(P_k

$\qquad$ = m_j| C_i) P(C_i)

At this point, the pixels record M covers attributes P_k with values m_j.The denominator P(M) is overlooked for the reason that, it is common for all the classes.The last line of the derivation is obtained by assuming independence between the Pixel attributes.

It is intended for the classification values of P (P_k= m_j| C_i) are pre-calculated and stored for all possible pixel attribute values and classes of the image. At this time of classification then these probability values are used to approximation P (C_i|M) according to the above derivation and the microarray image class with the maximum probability of incidence is amount produced.

2. Microarray Breast Cancer Image Dissection: Microarray image dissection is the process of dividing a microarray image into many segments that is sets of pixels, broadly recognized as image objects. The intention of separation is to simplify and/or change the depiction of a microarray image into rather that is more meaningful and easier to analyze. Microarray image dissection is classically used to locate objects and boundaries like lines, curves in microarray images[8]. It is more exactly of microarray image dissection is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. The consequence of image dissection is a set of slices that collectively cover the whole microarray image, or a set of shapes extracted from the microarray image. Each of the pixels in an area is similar with respect to some characteristic or computed property such as color, intensity, or texture [10].

3. Gaussian Mixture Model: To classify the microarray image inside which each part is a pixel. The value of the microarray image pixel is a number that shows intensity, color of the microarray image, or shape of the microarray image. Let X is an arbitrary variable that takes these values. For likelihood model determination that we can what if to have mixture of Gaussian distribution as the following form are as follows

$$f(x) = \sum_{i=1}^{c} p_i \, N \left(x \mid \mu_i, \sigma_i^2 \right)$$

Where c is the number of components or regions and $p_i > 0$ are weights such that

$$\sum_{i=1}^{c} p_i = 1,$$

$$N(\mu_i, \sigma_i^2) = \frac{1}{\sigma\sqrt{2\pi}} \, exp\left(\frac{-(x - \mu_i)^2}{2\sigma_i^2} \right)$$

where μ_i, σ_i^2, these two are mean and standard deviation of are class i. For an assumed mircoarray image I, the lattice data are the values of pixels and MoG is our pixel-based model. However, the parameters are $\theta = (p_1, …, p_k, \mu_1, …, \mu_k, \sigma_1^2, …, \sigma_k^2)$ and we can deduction the number of regions in MoG by histogram of lattice data.

4. Expectation Maimization Procedure: This procedure is changed and name again to Expectation Maimization with Gaussian Mixture Model. The process of Expectation of Mapping microarrays pixels data Procedure and Maximization Procedure stages clearly defined:

1.Input: Observed MicroArray Image in a vector x_j, j=1,2,…,n and i ε {1,2,…,k} label set.

2. $\qquad\qquad$ Make $\qquad\qquad$ ready:

$$\theta^{(0)} = \left(p_1^{(0)}, …, p_k^{(0)}, \mu_1^{(0)}, …, \mu_k^{(0)}, \sigma_1^{(0)}, …, \sigma_k^{(0)} \right)$$

3. E-Step:

$$p_{ij}^{(r+1)} = p^{(r+1)}\left(i \mid x_j \right) = \frac{p_i^{(r)} N\left(x_j \mid \mu_i^{(r)}, \sigma_i^{2(r)} \right)}{f(x_j)}$$

4. M-Step:

$$\hat{p}_{ij}^{(r+1)} = \frac{1}{n} \sum_{j=1}^{n} p_{ij}^{(r)}$$

$$\hat{\mu}_i^{(r+1)} = \frac{\sum_{j=1}^{n} p_{ij}^{(r)} x_j}{n\hat{p}_i^{(r+1)}}$$

$$\hat{p}_i^{2(r+1)} = \frac{\sum_{j=1}^{n} p_{ij}^{(r+1)} \left(x_j - \hat{\mu}_i^{(r+1)} \right)^2}{n\hat{p}_i^{(r+1)}}$$

5. Repeat step-2 and step3 until an random error i.e., $\sum_i e_i^2 < \varepsilon$.

6. Calculate the $p_{ij} = Arg\ Max_i\ p_{ij}^{(final)}$ j=1, 2… n.

7. Build labeled microarray image corresponding of each one microarray image.

This Expectation Maimization Procedure is a pixel labeling method such that the labeled microarray image shows each fragment or object by not the same kind of labels.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICRADL - 2021 Conference Proceedings

5. Procedure of Microarray Breast Cancer Image: The process of Microarray breast cancer image process in relations of three tactics namely Normalization, Feature Extraction and Disease Detection are showing in the disease identification of Brest Cancer practice as follows. Intially, Normalization is the first stage for all Disease Detection system. In this process, initially image area is detected. Receives the input microarray image of size N x N and is compared to the size of database image; if the input image and database image are not equal, the input image is to be resized to get the equal size of the database image. If the shape of the selected microarray image is required to obtain the database microarray image, the object of microarray shape image with until it matches with the database image. In the feature extraction, the feature is defined as a function of one or more measurements; each of them specifies some quantifiable property of an object. It is calculated with some significant physiognomies of the object. All features can be classified into low-level and high-level features. Low-level features are extracted from the original microarray image, whereas the extraction of the high-level features must be based on low-level features. Once the normalized microarray image is achieved, it can be compared to other microarray images, under the same nominal size, shape and illumination conditions. This evaluation is based on features extracted in transformation method[2]. One such popular transformation is DCT (Discrete Cosine Transform) is a feature extraction step in various studies on Micriarray image disease idenification. The input microarray images are divided into N x N blocks to define the local regions of processing. The N X N, two-dimensional Discrete Cosine Transform is used to transform the data into the frequency domain. After that, statistical operators that calculate various functions of spatial frequency in the blocks are used to produce a block-level DCT coefficient. The most recent part of disease identification. In this process, in order to identify a particular input image, various distant and nearest area classifiers by Duda and Fart, 1973 are used to compare the input image feature vector with the database feature vectors. After obtaining the distances for N x N matrix, one may need to find the averages of each column of the matrix. With the overall average is zero then there is a match between the input microarray image and database microarray image object.

6. Experimental Result: The microarray image pixels measures is calculated using TP and TN, which are true positive and negative tuples classified by classifier. FP and FN are positive and negative tuples, which are incorrectly classified. For Naive Bayes Classifier in view of DCT reduced data result is as shown in

Table-I

Real Class	Foreseen class	
	Is Benign	Is Malignant
Is Benign	93	1
Is malignant	3	43
Total =140	96	44

The correctness of naivy bayes classification got is 96.7 % for the entire data as the maximum in the statistical machine learning.

7. Conclusion: As of the analysis, we can conclude that the statistal with expecation maximization model is useful in predicting the breast cancer from microarrays data; there is too possibility for study using naivy classifier and dimensionality reduction methods in which may help in better sympathetic of gigantic data sets with several topographies in nearby upcoming.

8. REFERENCES

[1] Animesh Hazra et.al: Study and Analysis of Breast Cancer Cell Detection using Naive Bayes, SVM and Ensemble Algorithms, International Journal of Computer Applications (0975 8887) Volume 145 No.2, July 2016.

[2] Aalaei S., Shahraki H., Rowhanimanesh A., Eslami S: Feature selection using genetic algorithm for breast cancer. 16 Computational and Mathematical Methods in Medicine diagnosis: An experiment on three different datasets. Iran. J. Basic Med. Sci. 2016;19:476.

[3] Bazazeh D., Shubair R. Comparative study of machine learning algorithms for breast cancer detection and diagnosis; Proceedings of the fifth International Conference on Electronic Devices, Systems and Applications (ICEDSA); Ras Al Khaimah, UAE. 6–8 December 2016; pp. 1–4.

[4] Dai H., Cheng Z., Baii J.: Breast cancer cell line classification and its relevance with breast tumor subtyping. J. Cancer. 2017; 8:3131–3141. doi: 10.7150/jca.18457.

[5] Kathija.A and Shajun Nisha: Breast Cancer Data Classification Using SVM and Naive Bayes Techniques International Journal of Innovative Research in Computer and Communication Engineering.Vol.4, Issue12, December 2016.

[6] Oyewola D., Hakimi D., Adeboye K., Shehu M.: Using five machine learning for breast cancer biopsy predictions based on mammographic diagnosis. Int. J. Eng. Technol. IJET. 2017;2:142–145. doi: 10.19072/ijet.280563.

[7] Pratiwi P.S.: Development of intelligent breast cancer prediction using extreme learning machine in Java. Int. J. Comput. Commun. Instrum. Eng. 2016;3 doi: 10.15242/ijccie.er0116114.

[8] Sadhana and Sankareswari: A Proportional Learning of Classifiers Using Breast Cancer Datasets, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol- 3, Issue-11, pg-223 232, November 2014.

[9] Shukla N., Hagenbuchner M., Win K.T., Yang J.: Breast cancer data analysis for survivability studies and prediction. Computational Methods Programs Biomed. 2018; 155:199–208. Doi: 10.1016/j.cmpb.2017.12.011.

[10] Wang H., Yoon W.S.: Breast cancer prediction using data mining method; Proceedings of the 2015 Industrial and Systems Engineering Research Conference; Nashville, TN, USA. 30 May–2 June 2015.