

MACHINE LEARNING METHODS FOR THE DETECTION OF BREAST CANCER: A REVIEW

Amandeep Kaur¹, Dr. Rakesh Kumar², Parekh Sharma³
M.Tech (Scholar), Principal, Assistant Professor
Department of Computer Science and Engineering
Sachdeva Engineering College for Girls, Gharuan, Mohali (Punjab)

Abstract:- A universal kind of disease known as breast cancer that try to be a huge challenge for both doctors and researchers. A survey was completed that intends that approximately 80% population of females was distress from (BC) breast cancer in the year 2012. The diagnosis and the initial detection of disease are required. Hence, to complete the need, the machine learning techniques are utilized for both detection and prevention of diseases and especially image processing carried out for capturing the image from the body's internal tissues. In this research, the detailed study is completed on introduction of breast cancer, its classification, sign and common causes of diseases. Screening is a crucial phase in the detection of cancer related diseases that extracted the affected tissues in the body and represent it through the image. Basically, screening is another kind of x-ray images and the best for the detection of breast cancer is mammography. These screening systems rely on the algorithms and theories of (ML) machine learning methods. Further, the description of various machine learning techniques is presented as namely as Naïve Bayes, K nearest neighbor, wavelet transformation and Contour-let transformation. The present work is related to overview of breast cancer and how to detect and predict the breast cancer disease.

Keywords:- Breast Cancer, NB (Naïve Bayes), KNN (K-Nearest Neighbor), WT (Wavelet Transformation), CT (Contour-let Transformation).

I. INTRODUCTION

Human cancers are the world wide diseases that affect the majority of people. It is a production of cells that grow in malignant cells and in uncontrollable ways. The cancer is classified in various categories as liver cancer, brain tumor, ovarian cancer, breast cancer and so on. In cancer disease the cells are partitioned into overpowering manner and generate the tumor in the major organs of body [1]. The history of cancer cells is described in the following figure 1.

When some changes or the mutation of DNA (Deoxyribonucleic Acid) occurred in the blood cells and the

changes are uncontrollable and abnormal, they are being a cause to generate the cancer cells. The cells are segmented and distributed themselves in the entire body they created their copies as called cell doubling [2].

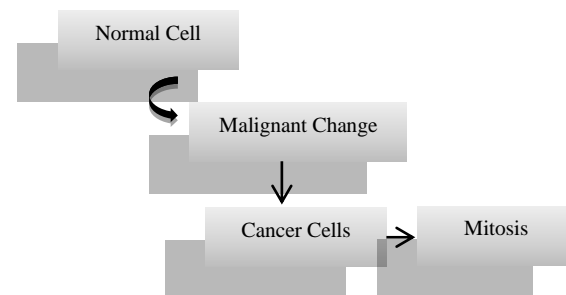


Fig 1. Formation of Cancer Cells [1]

Breast cancer is a huge challenge these days and the most dangerous form of cancer disease for women. It affects the majority of female's world widely. It is noticed from a survey that occurred in 2017 give information about approximately 255,180 females and 2,470 men were affected by this disease in the USA. Therefore, the chances type of cancer (Breast Cancer) in women is more as compared to the men. Basically, BC (breast cancer) is a complicated and heterogeneous disease that needs to be diagnosed in the early stages due to the long survival [3]. BC (Breast cancer) is the most communal reason of deaths in females. The detection of disease is difficult due to the unknown symptoms in the early stages [4].

The present research on breast cancer is partitioned into five portions. The first is about the basic information about the cancer and the production of cancer cells and the breast cancer. A second portion is related to the background of breast cancer, its classification and common causes. A third portion is the deep study of previous work done on the discovery of BC (breast cancer). Fourth portion contains the ML (machine learning) methods and the last part six is the summary of research with the scope of machine learning methods in future for the discovery of diseases.

II. HISTORY OF BREAST CANCER

The basic information about the BC is described in this section.

A. Breast Cancer Formation

The body cell in the breast grows in continuously in breast tissues that are abnormal and uncontrollable and being a cause of tumor and felt like the lump in the chest. The cancer cells are commonly produced on the inner side of the milk duct and the labels. Other cells are generated in the labels. Once the cells are produced they slowly reached into the additional portions of the body and create the more effects in the entire body [5] [6].

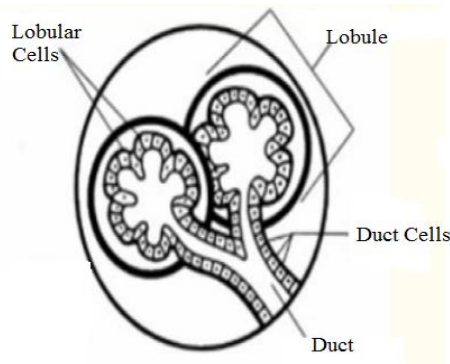


Fig 2. Structure of Breast [5]

B. Classification of Breast Cancer

The BC classification method is described as follows:-

- 1) *Non-Invasive Cancer*: These are the cancer cells that utilized to the duct and which are not covered by fatty materials of the breast.
- 2) *Invasive Cancer*: The breast cancer cells are partitioned by the duct and the label walls that covered by the fatty tissues. The tumor moves to the lymph-nodes and other organs without spreading.
- 3) *Lobular Carcinoma*: In this form, the breast cancer cells are not distributed in the area of its formation and increased the count of cells in the milk glands.
- 4) *Ductal Carcinoma*: The most commonly occurred breast cancers that formed in the ducts and called the distal come docarcinoma [5].
- 5) *Inflammatory Cancer*: The uncommon breast disease that changed the color of skin as red and the skin become warm after that, the covering changed to more thickness and pitted.

- 6) *Nipple Disease*: The tumor cells are moved to the nipples and create the dark circles around the covering of the nipples.
- 7) *Phyllodes Cancer*: The breast cancer obtained in the stroma of the breast.
- 8) *Angiosarcoma*: The cells are different kinds and initialized the spreading from the line lymph and go to the blood vessels [7].

C. Signs and Symptoms

The signs and symptoms in breast cancer are realized in the initial stages that required to be known. There are various symptoms are abdominal pain, lump in breast tissues, texture of breast, size changes, skin conditions, swelling and nipple discharges.

D. Basic Causes of Breast Cancer

The common causes of breast cancer are given below-

- 1) *Family Background*: The patients whose family suffered from the cancer like diseases have more risk factors of cancer produced in breast tissues.
- 2) *Genetic Reasons*: As mentioned above family history matters in case of cancer production. The genes of both maternal and paternal relatives are crucial and the chances of the rise of breast cancer are more in the younger age. There are more chances of genes as mother, sister and daughter.
- 3) *Hormone Changes*: If the continuous alterations occurred in the hormones then, the changes produced the cancer cells. Its most common occurrence is at the beginning and ending of the menstrual cycle, for instance, pregnancy at an early age, hormone shuffling and the high usage of oral pills.
- 4) *Dietary System*: The diet of also matters. If the diet is not healthy and also full of fatty food, then being a cause of cancer diseases.
- 5) *Environmental Issues*: Environment effects matter for women who had no enough radiation for a long-time period [5] [8].

III. LITERATURE SURVEY

Amrane, M., et al., (2018) [9] proposed a classification and detection of breast cancer on the basis of machine learning methods. Basically, breast cancer was a form of cancer disease that occurred in the breast tissues. These were obtained via the obesity, family history and the reproductive factors. These days, there were majority of women that tackled the breast cancer disease. The detection of breast cancer was identified by its signs such sustain pain, mutation of genes, alterations in size, skin texture and redness on the breast skin. The

classification of disease was crucial for searching a systematic and the aim full predictive. The common method of classification was binary and malign cancer. In the current research, the machine learning methods were assigned for the detection and classification of breast cancer. It was utilized to give high accuracy and effective treatment capabilities. Generally, the two approaches of machine learning were trained as Naïve Bayes and the K nearest neighbor (KNN). There was a comparison was performed between these methods and the accuracy of KNN was higher as compared to Naïve Bayes with the lowest error rate. The accuracy was recorded at 97.51%. **Jafarpisheh, N., et al., (2018) [10]** performed the relapse prognosis on the breast cancer disease through the combination of basic and modern structure of MLA (Machine Learning Algorithms). It was assumed that, the current most harmful disease was cancer disease and breast cancer was a sub category of this. The majority of women's deaths were recorded to be due to the breast cancer (BC). Therefore, MLA was assigned to detect and diagnose BC. The purpose of these machine learning methods to be sure about the breast cancer was relapsed or not. The present work was linked to the predictions of BC through MLPN (MultiLayer Perceptron Network) along with its two different outputs. First one was DNN (Deep Neural Network) which worked as a feature extraction in which MLPN was a classifier. On the other side, the second one was RNN (Rough Neural Network), finally SVM (Support Vector Machine) was trained. The comparison of performance of each method demonstrated that the RNN worked well on the generation of two outputs and gained the higher percentage of accuracy in the shortest variety of structures. The obtained accuracy of the proposed method by RNN was 92.77% with variance at 12.2333. **Hela, B., et al., (2013) [11]** proposed a work on the detection of breast cancer and the approach was based on the mammograms. Generally, the mammogram was a form of x ray specifically utilized for detection of nodules that represents the occurrence of abnormal cells in the breast and shows the existence of breast cancer. It was considered by researchers and the doctors that, the early detection of cancer disease was easier to detect and to diagnose rather than the detection of higher stages of disease. It would decline the long survival of the patients. The research was described the initial stages of breast disease in tiny lesion boost prognosis and conveyed it to expel the mortality. Hence, the mammography becomes best from other techniques for better diagnosing of the disease by screening. Although the procedure of mammograms was not a simple task because of the presence of different tissues in the image and it was specifically accessed the dense breast. The work was all about the study of the early detection methods of BC through the analysis and the analysis given access to the radiologists to understand the type of images to detect the disease. The early detection was useful to almost decline the mortality rate by 25%. **Gayathri,**

B. K., et al., (2016) [12] utilized the image segmentation based approaches for the early detection of breast cancer (BC). Basically, digital image processing becomes an interesting area of the research almost in every field and had a lot of attention in the medical field for its vast range of application of detection diseases. Image segmentation process was an approach to detect and screening of images of diseases such as cancer disease. The image segmentation was briefly described in this work and it included various kinds of segmentation methods as thresholding, edge based that further categorized as gray and gradient based, region based segmentation that was a collection of various methods as Sobel, Prewitt, Laplacian and canny approach. The advantages and disadvantages of each of the methods were represented in this work. The procedure of segmentation was to partition the images according to their relevant data which obtained via a region growing segmentation and give access to the quick seed and to choose the early seeds. **Kashyap, K. L., et al., (2015) [13]** represented the digital mammograms specifically for the detection of breast cancer. Mammography was a fascinating approach for the investigation of breast cancers. It was performed well in the early stages of BC and in case when recovery chances were more. The aim was to find out the abnormalities and locate the exact location in the image. The current research depicts a technique that had the tendency to detect the abnormalities in mammograms by also using ROI (Region of Interest). The un-sharp masking also approached for the improvement in the process of the mammogram and the reduction of noise was declined by filtration. The proposed methods as tamura features and moment invariants worked together. It obtained the higher accuracy and valid on the Mini MIAS datasets. The parameters and the performance of proposed method were sensitivity at 97.14%, specificity at 96.67% and the accuracy was noticed at 96.92%. The higher accuracy of the research was being a way to enhance the feature extraction more in a suspicious region and the classification of benign and malignant become easier. The results of moment based were better overall as compared to the region based and tamura based approach. **Paramkusham, S., Rao, et al., (2013) [14]** proffered the dynamic image processing approach by implementation in matlab and lab view particularly for the breast cancer. The detection of BC at the initial stages was solved by using mammographic images. The low contrast properties of the images make difficulties in the way to detect the signs of micro calcifications and masses. In the research, the objective was to utilize a primitive method that was highly dynamic for this purpose of detection through following image processing methods. The main reason behind the execution of primitive method was to implement in the mass region particularly for feature extraction from the large masses and for the supervision of the boundary simply. The common extraction of features was mean, standard deviation, entropy and kurtosis. It was demonstrated that, the majority of

the features were utilized to sort out the benign and malignant masses. Therefore, some predictions were created for future that, it would be easy to identify and classify the breast masses in the beginning of cancer disease.

TABLE I. LITERATURE COMPARISON

Author's Name and Year	Used Technique	Performance parameter s	Research Gap
Amrane, M., et al., (2018) [9]	Naïve Bayes KNN	Accuracy Error Rate	Classification method was only on binary and malign cancer
Jafarpisheh, N., et al., (2018) [10]	MLPNN DNN SVM	Accuracy Variance	Cancer recurrence or not
Hela, B., et al., (2013) [11]	Mammograms	Mortality rate	Long survival
Gayathri, B. K., et al., (2016) [12]	Image segmentation	-	Processing of images Not proper investigation of disease
Kashyap, K. L., et al., (2015) [13]	Digital mammograms DWT ROI Un-sharp masking	Accuracy Specificity Sensitivity	Low recovery changes
Paramkusham, S., Rao, et al., (2013) [14]	Feature Extraction	Mean Standard deviation Smoothness Skewness	Detection of signs

IV. MACHINE LEARNING METHODS FOR THE DETECTION OF BREAST CANCER

Basically machine learning is the arrangement of different kinds of techniques that accessed for the creation and determination of different algorithms. The purpose is to describe the process of prediction, detection, recognition and classification of various diseases. Generally machine learning is relied on the four fundamental phases as the collection of data, selection of the perfect model, training phase and finally the testing phase. In the medical field, machine learning played out a crucial role for the detection and treatment of certain diseases. In this work, the objective is to detect the breast cancer disease. There are various techniques come under the machine learning such as SVM (Support Vector Machine), MLP (MultiLayer Perceptron), NN (Neural Networks), Naive Bayes, KNN (K-Nearest Neighbor), WT

(Wavelet Transformation), CT (Contourlet Transformation) and so on. For the detection of breast cancer KNN and Naïve Bayes are used.

A. Naïve Bayes method

The naïve Bayes algorithm is an approach that gives output in the form of probabilities and the statistics. It describes a design of decisions. The variables that are used in Naive Bayes are independent and impact on the determination of a particular model. The training compounds are active (A) and in-active (I) provides the description (D), hence the condition probability distribution are evaluated as-

$$\text{Probability Distribution (P)} = \frac{D}{A} \text{ and } \frac{D}{I}$$

The classifiers of theorem are also utilized for the ranking of the databases of the probability activities. The Bayesian theorem is as-

$$P(I/A) = \frac{P\left(\frac{1}{A}\right)P(I)}{P(A)} \dots\dots(i)$$

In equation (i), P(I) is considered as the prior probability on the event I, P(A) is the probability of training data and the conditional probability is the $P\left(\frac{1}{A}\right)$. The decision theorem also combined to determine the given data y_i belong to R_i where R_i is the class.

$$P(y/R_i) P(R_i) > P(y/R_j) P(R_j) \dots\dots(ii)$$

Equation (ii) represents the classes and if this equation is not equal to I then it shows the two different classes present and shows that x belong to R_i .

B. K- Nearest Neighbor (KNN)

The main motive of KNN is to regulate the classes and the properties of a particular data. KNN come under the classification techniques and extremely utilized for its simple implementation and the effectiveness. The discovery of KNN was made by T.M. Cover and P.E. hart. KNN is being the best approach for the detection and better diagnosis of cancer disease. The process is dependent on the instance approach in which the model is not executed while the training period that raises the cost and time consumption. It just stores the training tuples and the class labels. The classifiers are created at the classify time. Euclidean distance is used to determine the difference between the training and testing phase of the model. The Euclidean distance ED (y_i, y_j) is represented as in equation (iii).

$$ED (y_i, y_j) = \sqrt{\sum (b_s(y_i) - b_s(y_j))^2} \dots\dots (iii)$$

The instance is described by using an attribute vector $\langle b_1, b_2, b_3 \dots b_n \rangle$ the b_j is the value of the attribute b of x . the irrelevant attributes are increased and to sort out the problem weights are assigned with the ED as shown in equation (iv).

$$ED (y_i, y_j) = \sqrt{\sum Wt_j (b_s(y_i) - b_s(y_j))^2} \dots\dots (iv)$$

Where Wt_j is the weight assigned to the different attributes and it varies as $Wt_j (j= 1, 2, \dots n)$ [9] [15].

C. Wavelet Transformation (WT)

Generally, Gabor transformation utilized due to the ability to perform the task of describing information related to the frequency and the time signals.

Although some of the challenges were occurring due to the lack of the sustainable window. Therefore, wavelet transformation is considered to sort out the challenges. Basically, it is a window time analyzer that worked in the variable scale. The continuous WT function $Funct (T)$ is represented as the integral of $P2 (S)$ mentioned in below equation (i)

$$WT^* Funct (v,r) = \int_{-\infty}^{+\infty} Funct (T)_{\beta_{v,r}} (T) dt \dots\dots(i)$$

The function of WT is modified with the transformation and the wavelet scale changed as shown in equation (ii).

$$\beta_{v,r} = \frac{1}{\sqrt{r}} \beta \left(\frac{T-v}{r} \right) \dots\dots (ii)$$

Where r is considered as the parameter of scale and linked to the frequency information with the transfer parameter (v). The major objective behind the creation of this transformation is to apply it for the production of 2 dimensional signals or the images. Further, it is also applicable for the discrete signals.

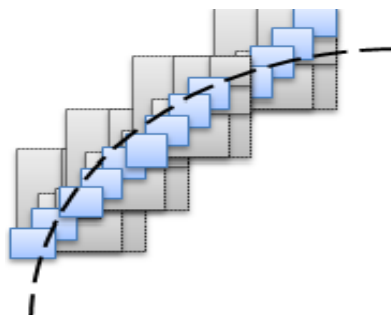


Fig3. Wavelet Transformation [16]

There are enormous benefits of wavelet transformer that applicable in the vast usage of image processing methods. Sometimes WT called as the mother wavelet because it is a collaboration of different operators. The basic two categories of DWT are Haar and Daubechies [16].

D. Contourlet Transform (CT)

The contourlet is another enhanced version of wavelet transformation that is a two dimensional direction based transformation assist in the image processing. The main reason of accessing CT is to represent the curves and the other important details of the image. The implementation of this transformation relies on the core functions that used to show the different directions towards the shape, curve and scale of the images. The common features of CT are to smooth the contours, extraction of geometrical features, use filter bank, defines the edges.

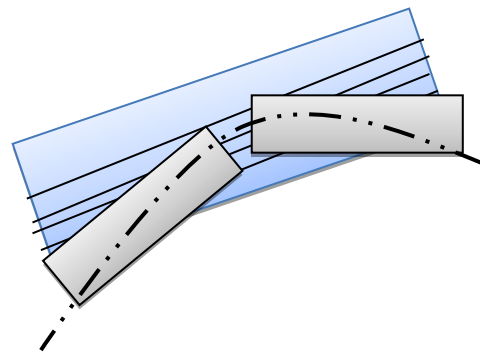


Fig 4. Contourlet Transformation [17]

When wavelet transformation is combined with the direction based filters and it described the spots in a particular image and formatted into a line, the procedure is called as contour transformation.

V. CONCLUSION AND FUTURE SCOPE

To summarize, BC is a leading cancer disease and there are certain steps that accessed can easily reduce the effect of cancer, such as the early detection by medical tests as ultrasounds, mammography and others. The medical tests are also helpful to expel the long survival of patients. In this work, the basic information about the breast cancer is represented and the description of KNN, Naïve Bayes, CT and WT is also discussed.

The current research intends that in future these machine learning methods are implemented and enhanced the diagnosis system with its unique characteristics. Therefore, the usage of

BC (breast cancer) converted from a challenging task to a simple solution.

REFERENCES

1. Idikio, H. A. (2011). Human cancer classification: a systems biology-based model integrating morphology, cancer stem cells, proteomics, and genomics. *Journal of Cancer*, 2, 107.
2. Bio-Organic Chemistry of Natural Eneidyne Anticancer Antibiotics. NPTEL – Chemistry – Bio-Organic Chemistry of Natural Eneidyne Anticancer Antibiotics Joint. Joint initiative of IITs and IISc – Funded by MHRD.
3. Coleman, C. (2017, May). Early detection and screening for breast cancer. In *Seminars in oncology nursing* (Vol. 33, No. 2, pp. 141-155). WB Saunders.
4. Guzmán-Cabrera, R., Guzmán-Sepúlveda, J. R., Torres-Cisneros, M., May-Arrijoja, D. A., Ruiz-Pinales, J., Ibarra-Manzano, O. G., and Parada, A. G. (2013). Digital image processing technique for breast cancer detection. *International Journal of Thermophysics*, 34(8-9), 1519-1531.
5. Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., & Sharma, K. K. (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research*, 1(2), 109.
6. Pareek, A., and Arora, S. M. (2017). Breast cancer detection techniques using medical image processing. *Breast cancer*, 2(3).
7. Pushpalatha, M. S., Kayalvizhi, S., Nansi, N. M. D., and Kumar, D. N. A Survey on Breast Cancer Detection Techniques. *International journal of Computer Science and Information Technolgoeis*, Vol. 7 (6), 2445-2447.
8. Hider, P., and Nicholas, B. (1999). *The Early Detection and Diagnosis of Breast Cancer: A Literature Review: an Update*. New Zealand Health Technology Assessment Clearing House.
9. Amrane, M., Oukid, S., Gagaoua, I., and Ensari, T. (2018, April). Breast cancer classification using machine learning. In *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)* (pp. 1-4). IEEE.
10. Jafaripisheh, N., Nafisi, N., and Teshnehlab, M. (2018, February). Breast cancer relapse prognosis by classic and modern structures of machine learning algorithms. In *Fuzzy and Intelligent Systems (CFIS), 2018 6th Iranian Joint Congress on* (pp. 120-122). IEEE.
11. Hela, B., Hela, M., Kamel, H., Sana, B., and Najla, M. (2013, March). Breast cancer detection: A review on mammograms analysis techniques. In *Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on* (pp. 1-6). IEEE.
12. Gayathri, B. K., and Raajan, P. (2016, January). A survey of breast cancer detection based on image segmentation techniques. In *Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on* (pp. 1-5). IEEE.
13. Kashyap, K. L., Bajpai, M. K., and Khanna, P. (2015, September). Breast cancer detection in digital mammograms. In *Imaging Systems and Techniques (IST), 2015 IEEE International Conference on* (pp. 1-6). IEEE.
14. Paramkusham, S., Rao, K. M., and Rao, B. P. (2013, September). Early stage detection of breast cancer using novel image processing techniques, Matlab and Labview implementation. In *Advanced Computing Technologies (ICACT), 2013 15th International Conference on* (pp. 1-5). IEEE.
15. Taneja, S., Gupta, C., Goyal, K., and Gureja, D. (2014, February). An enhanced k-nearest neighbor algorithm using information gain and clustering. In *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on* (pp. 325-329). IEEE.
16. Nasiri, Y., Hariri, M., and Afzali, M. (2015, November). Breast cancer detection in mammograms using Wavelet and Contourlet transformations. In *Knowledge-Based Engineering and Innovation (KBEI), 2015 2nd International Conference on* (pp. 923-926). IEEE.
17. Do, M. N., and Vetterli, M. (2005). The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 14(12), 2091-2106.