International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Breast Cancer Detection using Machine Learning Algorithm (ANN)

***Hiba Masood*[1] , Mr Vinayak V. Palmur[2]**

[1] *Master of Engineering student, Department of Computer Science and Technology, V.V.P.I.E.T. Solapur, Maharashtra, India.*

[2] *HOD*, Department of Computer Science and Engineering, V.V.P.I.E.T. Solapur, Maharashtra, India.

Abstract - Breast cancer (BC) is one among the foremost common cancers among women worldwide, representing the bulk of latest cancer cases and cancer-related deaths consistent with global statistics, making it a significant public health problem in today's society. For identifying and earlier diagnosis of Brest cancer we require oncologists to examine the breast lesions for detection classifying of various stages of cancer. These manual examinations are very time consuming and some time we can say that it may inefficient in many cases. So there is a primary need of creating an efficient method for the diagnoses of those cancerous cells without the human interference with high accuracy.

The early diagnosis of breast cancer (BC) can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Some accurate classification of benign tumors can be preventing patients undergoing long-term treatments. Thus, the correct diagnosis of breast cancer (BC) and classification of patients into malignant or benign groups is the subject of much research. Because of its unique advantages in critical features detection from complex breast cancer (BC) datasets, machine learning (ML) is widely recognized as the methodology of choice in breast cancer (BC) pattern classification and forecast modeling. In this system, we aim to review machine learning (ML) techniques and their applications in breast cancer (BC) diagnosis and prognosis.

In this research, we uses different image processing techniques for developing the imaging biomarkers through the mammographic analysis and based on Machine Learning algorithms we are aiming to detect breast cancer in early stages to support the diagnosis and get fastest attention seeking of high-risk patients. For achieving this automatic classification of breast cancer based on mammograms, a generalized regression artificial neural network was actually trained and tested to separate the different types of tumors like malignant and benign tumors. And that reaching accuracy near about 95.83%. By using the biomarker and trained neural nets, a specific computer-aided diagnosis system is being designed.

Key Words*: *breast cancer detection, digital image processing, artificial neural networks, biomarkers, computer-aided diagnosis

1.INTRODUCTION

Momentarily, cancer is a massive widespread health problem throughout the globe. According to the International Agency for Research on Cancer (IARC), part of the World Health Organization (WHO), there were 9.2 million deaths caused by cancer in 2018, and 18.1 million new cases of this disease are required to occur until 2030. Cancer, medically defined as a malignant neoplasm, is a broad group of diseases involving unregulated cell growth. In breast cancer, a cell divides more quickly than any active cell, and these cells grow in an undisciplined form, riding nearby regions of the body. Cancer can spread to all regions of the body through the lymphatic systems or bloodstreams. Breast cancer cells are clearly identified by classifying them into malignant and benign.

1.1 Prediction Of Brest Cancer

Harmless tumors are those which remain in their primary spot without overrunning other sites of the body. Benign tumors cells are hardly lead to a person's death; yet, at some point, this benign tumor leads to a cancer tumor but they are not problematic they are going to easily get treated . On the other hand, a malignant tumor grows uncontrollably and spreads to remote spots of the body. They are capable of destroying the body tissue. So that's why malignant tumors are cancerous cells and they are responsible for patient death. That is why malignant tumors are more deliberate than benign cells. Breast cancer (BC) is the most generally occurred cancer type and it became the highest cause of death among women across the globe. Among the other cancer types, Breast Cancer (BC) is the second most common type of cancer which must be found in women, excluding skin cancer. Besides, the death rate of Breast Cancer(BC) is very high as it is compared with the other types of cancer. Breast Cancer(BC), equivalent to other cancers, starts with a quick and uncontrolled outgrowth and expansion of a part of the breast tissue, which relies on its potential harm, is divided into benign and malignant types. Generally, there are two types of Breast Cancer(BC) are in situ and invasive. DCIS(Ductal carcinoma) in situ is the presence of abnormal cells inside a milk duct in the breast side and it will not spread to other body parts even if it grows. Invasive breast cancer, on the opposite, is very aggressive and spreads to other nearby organs, and destroys them as well . It is very important to detect the cancerous cell before it spreads to other organs; thus, the survival rate for the patient will increase to more than 97% . A vital difficulty we found in our medical science is to a diagnosis of disease, which we found by performing various tests upon the patient. Judgments from specialists and evaluations of data taken from the patients are the necessary factors in the diagnosis of disease. So finding the correct results in the given time is difficult Breast cancer(BC) is the major difficulty in the medical field. As Breast Cancer(BC) is a dangerous disease, only early diagnosis of the disease can prevent life. Clinical diagnosis of Breast Cancer(BC) helps in identifying the malignant cases, timely diagnosis can increase the chances of a patient's life from 65 to 90%.BC has four early signs: micro-calcification, mass, architectural distortion, and breast asymmetries .

The different standard modes used for breast cancer diagnosis (BCD) are positron emission tomography (PET), magnetic resonance imaging (MRI), CT scan, X-ray, ultrasound, photoacoustic imaging, tomography, diffuse optical tomography, elastography, electrical impedance tomography,

optoacoustic imaging, ophthalmology, mammogram, etc. The results which we found from these procedures are very helpful in identifying their patterns, which seek to help the doctors for differentiating between malignant and benign cases. Nonetheless, manual classification of images is a hard and time-consuming job, being highly capable of interobserver variability and human errors, resulting in too low critical outcomes, thus markedly increasing the workload of radiologists because of their significant shortage in market. In addition, medical care prices that are relevant to imaging are fatly increasing. Therefore, we need new methods for diagnosis purposes. Nowadays, bio imaging quantification is an emerging method in the field of radiology with growing implantation in clinic centers. It delivers us the more relevant information that is not perceptible by the naked eye in orthodox radiological reading.

It consists of the generation of quantitative (numerical) data from images, mainly of the high solution, to provide knowledge on which to support a clinical examination.Biomarkers can be said to be the transition from radiology to the personalized antidote.

1.2 Breast Cancer Detection By Using Bio Image Markers

Bioimage markers allow to characterization and study of different diseases using some kind of information, such as genetic, histological, clinical imaging, etc. These kinds of biomarkers are responsible for detecting abnormalities in the genetic mutations which yield some type of disease and they are helpful or useful in the clinic for the diagnosis of patients with some types of disease. The application of measurement of bioimaging markers to aid in the recognition, treatment, and follow-up of pathologies delivers added value throughout the clinical practice process by delivering additional information to traditional diagnostic tests .From imaging tests processed in the right way, abnormalities in tissue are evidenced before they are perceptible in the reading of the radiologist, the fundamental objective of this type of biomarkers. In expansion, they allow the observation of the therapy results from a quantitative point of view.

As defined earlier, Breast Cancer(BC) is one of the highest causes of death, estimated for nearly one-third of the world's population. Recently, we found that clinical screening by using mammography is the most popular way in the early detection of this disease. Using analysis of mammograms obtained through X-rays allows radiologists to fantasize early symptoms of cancer, such as calcifications, masses, and architectural distortions among other early signs of cancer. However, this analysis is a regular, monotonous, and exhausting task and it will show that only 3–4% of the patients are carcinogenic.

It has been shown that because of these problems and other factors leads to cancer such as obscuration of abnormalities by fatty tissue, a radiologist can omit up to 30% of cancers. Moreover, because this type of analysis produces many false positives, the number of extreme biopsies is increased up to 35%, causing a high level of stress in the patient, and in turn saturating the health systems. Due to lots of problems presented by mammography screening, lots of efforts have been made to support the radiologist in the search for these types of injuries through Software algorithm or biomarker systems, these latest advanced effects are evolved very helpful for radiologists in computer vision and their manipulation in digital format. Nowadays, one of the primary research topics in diagnostic radiology and medical imaging is computer-aided detection/diagnosis. Software algorithm systems allow the radiologist to use mammography to emphasize certain parts that would otherwise be difficult to visualize. One of the most used techniques is the improvement of contrast, which allows highlighting objects in areas of low intensity. To date, software algorithm is a more useful method for diagnosing cancer in computed tomography, X-ray, MRI, or mammogram images. The software algorithm system is an effective intermediate between input images and the radiologist. The output from software algorithm is not considered as a result; nevertheless, the result is used as a reference concerning additional testing in the related field .The software algorithm approach is becoming extremely helpful for doctors to diagnose diseases with a higher degree of efficiency and minimize examination time and cost. Also, helpful in avoiding unnecessary biopsy procedures. However, software algorithm systems not only allow better visualization of mammograms but also use different digital image processing (DIP), knowledge discovery from data (KDD), artificial intelligence (AI) techniques such as artificial neural networks (ANN) allow preselecting specific regions of stakes (ROIs) for more delinquent examination by the radiologist .Type of histopathology images into separate histopathology patterns, interrelated to the cancerous and noncancerous condition of the examined tissue, is usually the primordial goal in image analysis systems for cancer by using automatic-aided diagnosis applications. Recent advances in DIP, KDD, and AI techniques allow to building software algorithm that can assist pathologists to be more productive, objective, and consistent in diagnosis. The major challenge of such procedures is dealing with the inherent complexness of histopathological images. This research aims to use advanced DIP to investigate and develop specific imaging biomarkers for Mexican patients through the quantitative mammography analysis and with this information to develop technology-based on advanced KDD and AI techniques, aiming to detect breast cancer in the early stages to support the diagnosis and prioritization of high-risk patients.

This research aims to use advanced DIP to investigate and develop specific imaging biomarkers for Mexican patients through the quantitative mammography analysis and with this information to develop technology-based on advanced KDD and AI techniques, aiming to detect breast cancer in the early stages to support the diagnosis and prioritization of high-risk patients.

2. LITERATURE REVIEW

Ali et al. work on Big health data it was concluded that machine learning algorithm is best suited for detection of specific problem and it is possible to predict individual cancer risk via deep learning based solely on personal health informatics. [2]

Gayathri and others, observed the actual performance of different machine learning algorithms like Support Vector Machine (SVM) and Relevance Vector Machine (RVM). Overall they found many researchers have applied the algorithm of neural networks for predicting cancers, especially the breast cancer. If studies on Relevance Vector Machine (RVM) continue, then it is likely that the use of

Relevance Vector Machine (RVM) will become much more useful in diagnosing breast cancer. [7]A research done by Habib Dhahri and other solved the problem of detection of breast cancer using a machine learning algorithm and this is done automatically by conducting three different experiments on the breast cancer dataset. In the first test, they proved that the three most popular evolutionary algorithms can achieve the same performance after effective configuration. The second experiment focused on the fact that combining features selection methods improves the accuracy performance. Important results were derived from the proposed method by evaluating an ensemble of approaches from an exhaustive machine learning technique; they encountered a significantly higher time consumption rate. The proposed model looks naturally suited for control parameter setting of the machine learning algorithms in one side and automated breast cancer diagnosis on the other side. [8]

Youness Khourdifi and Mohamed Bahaj used five learning algorithms: SVM (support vector machine), Random Forest, Naive Bayes, and K-NN (K-Nearest Neighbors), applied to the breast cancer dataset, and tried to compare them according to many criteria: accuracy, turnaround time, sensitivity, and specificity. In their work SVM (support vector machine) has proven its performance on several levels by the lowest error rate, and shortest turnaround time. [9]

Ebru Aydındag Bayrak and others from Department of Computer Engineering Istanbul University Turkey discussed two popular machine learning techniques for Wisconsin Breast Cancer classification. Artificial Neural Network and Support Vector Machine were used as ML (machine learning) techniques for the classification of WBC (Original) dataset in WEKA tool. Based on the performance metrics of the applied machine learning techniques, SVM (Sequential Minimal Optimization Algorithm) showed the best performance in the accuracy of 96, 9957 % for the diagnosis and prediction from WBC dataset. [10]

Abdelghani Bellaachia and others interpreted that the preliminary results are promising for the application of the data mining methods into the survivability prediction problem in medical databases. Their approach takes into consideration, the Survival Time Recode (STR), the Vital Status Recode (VSR) and Cause of Death (COD). [11]

Anusha Bharat and others from Institute of Technology Bangalore, India submit a research on different algorithm performs in a various way and that is completely depending upon the dataset & the parameter selection. They concluded SVM (Support vector machine) is a strong technique for predictive analysis it was concluded that SVM (support vector machines) using Gaussian kernel is the most suited technique for recurrence/non-recurrence prediction of breast cancer. The SVM (Support vector machine) used in the analysis was only applicable when the number of class variable was binary. To solve this problem scientist came up with multiclass SVM (Support Vector Machine).[12]

Muhammet Fatih Aslan and others in their work on Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis. This dataset was different from other datasets in terms of feature type. The significance of data in breast cancer detection was investigated by ML(machine learning) methods. Analysis was performed with four different ML (machine learning) methods. Interfaces for a Artificial Neural Networks (ANN) and were Extreme Learning Machine (ELM) developed. In addition, the hyper parameter values giving the least errors for Artificial Neural Networks (ANN), Extreme Learning Machine (ELM), K-Nearest Neighbour (KNN) and SVM (support vector machines) methods were determined using hyper parameter optimization technique. Accuracy rates and training times were obtained according to these values. The results indicated highest accuracy rate and the lowest training period by Standard Extreme Learning Machine (ELM). They proved, the use of Standard Extreme Learning Machine (ELM) is more advantageous in terms of time when there are a high number of samples.[13]

In the review published by Wenbin and others from China have shown for several decades Artificial Neural Networks (ANN)s have dominated Breast cancer diagnosis and prognosis, recently alternative ML(machine learning) methods are applied to intelligent healthcare systems to provide a variety of options. Lots of algorithms achieved very high accuracy in WBCD (Wisconsin Breast Cancer dataset). Classification accuracy is very important assessment criterion but it is not the only one. ML (machine learning) techniques have shown their remarkable ability to improve classification and prediction accuracy. [15]

Bibhuprasad Sahu, and other published a research article on A Hybrid Approach for Breast Cancer Classification and Diagnosis proposed predictive model for diagnosis of cancer. They incorporated Multivariate statistical and machine learning techniques for better accuracy. They measured performance of different classifier techniques. Their study result reveals Artificial Neural Networks (ANN) plays major factor for detection of cancer diagnosis to save the human life from the dangerous disease. [16]

Smita Jhajharia, and others review article on risk factors, susceptibility, and machine learning techniques for cancer prediction suggested the multitude of various general and miscellaneous risk factors have not been comprehensively taken into account for the modeling of a predictive tool. There is a need for a robust mathematical model incorporating all that have been left. They have highlighted various Absolute Risk Prediction Models like the Gail model and BRCAPRO model. [17]

Shubham Sharma, and others from Amity University Uttar Pradesh did a comparative study of different machine learning algorithms, for the detection of breast cancer. Performance comparison of the machine learning algorithms techniques was carried out using the Wisconsin Diagnosis Breast Cancer data set. They concluded that K- Nearest Neighbour (KNN)is the most effective in detection of the breast cancer as it had the best accuracy, precision and F1 score over the other algorithms. [18]

Naresh Khuriwal and Dr Nidhi Mishra proposed the deep learning method, convolutional neural network that is mostly used for classification of images dataset. After the implementation this method they achieved 99.67% accuracy. [19]

Sangeetha D, N. And others published their work in the International Journal for Research in Applied Science & Engineering Technology (IJRASET) in 2018 on Predicting Cancer using Machine Learning Algorithms. [20]

David A. and others have deduced the following conclusion in 2019. They analysed WDBC (Wisconsin Diagnostic Breast

Cancer) dataset using dimensionality reduction techniques and three popular ML (machine learning) algorithms to classify malignant and benign tumors. Their experimental work proves that classification performance is dependent on the ML (machine learning) classification technique chosen. Nevertheless, SVM-LDA (support vector machine-linear discriminant analysis) is chosen over ANN-LDA (Artificial Neural Networks-linear discriminant analysis) because it takes a longer computational time. They proposed an intelligent approach which integrates linear discriminant analysis and support vector machine (with RBF kernel) for breast cancer diagnosis. This chosen approach showed good and promising results over the validation dataset. It obtained a classification accuracy of 98.82%, sensitivity of 98.41%, specificity of 99.07% and area under the receiver operating characteristic curve of 0.9994. [21]

Shelly Gupta et al. have observed that the accuracy for the diagnosis analysis of various applied data mining classification techniques is highly acceptable and can help the medical professionals in decision making for early diagnosis and to avoid biopsy. The prognostic problem was mainly analysed under Artificial Neural Networks (ANN) s and its accuracy came higher in comparison to other classification techniques applied for the same. But more efficient models can also be provided for prognosis problem like by inheriting the best features of defined models [22]

B Nithya, published her research work which shows that machine learning tools and techniques are essential in numerous disease predictions. There are lot of open problems and future challenges in dealing with massive amounts of heterogeneous, distributed, diverse, highly dynamic data sets and increasingly large amounts of unstructured and non-standardized information with respect to varied types of diseases. Machine Learning techniques are momentous in various industrial applications. She recommended efficient machine learning approaches becomes essential in the health care industry to address these challenges. [23]

Margaret C.in her review article claimed, neural networks are one of many different computational techniques that may be applied to cancer diagnostics and treatment. The final conclusion was cancer research literature supports the claim that Artificial Neural Networks (ANN)s are effective tools in cancer diagnosis and treatment, and suggests that there is an expanding role for computer technologies in the future of medicine.[24]

3. PROPOSED SYSTEM

Presently, there are no public databases of Breast Cancer(BC) in India. Therefore, at the first stage, different public mammography databases were used for developing and validating digital image processing algorithms capable to select ROIs from mammograms to extract image features used to prepare a generalized regression artificial neural network (GRANN).

The aim was to generate a methodology for the characterization of mammograms and their association with risk factors in BC patients as well as to integrate and develop the technological instruments for mammography analysis for BCD using AI technology.

The main goal of the second part is the generation of an anonymous database of random patients for free use which is provided by the scientific community for the study of Breast Cancer(BC) and to check the methodologies developed in association with the different General Hospitals.

In the third stage is sought to generate a long-term prospective protocol, which will allow the creation of a database with different risk factors associated with the development of BC. This protocol will allow the collection of clinical data of patients with both high and low probabilities of developing cancer. These data will be capable to validate the processes of cancer detection by the scientific community.

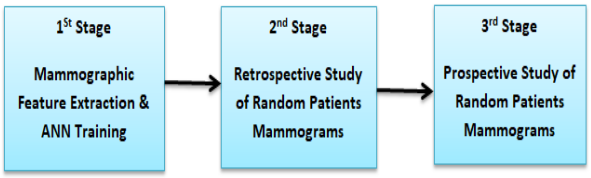

Fig -1: *Main Stages of Implementation*

The generation of a forthcoming protocol will allow the development of the database for the analysis of breast cancer in random patients. Unlike the retrospective protocol, the forthcoming protocol seeks to include clinical data, risk factors, and mammograms, among others.

4. IMPLEMENTATION

Patient prioritization plays a significant role in the reach of health services, where not all have access to these technological oncology services.

Therefore, this analysis aims the study Breast Cancer (BC) by developing a method that allows the detection of patients with a high possibility of BC in earlier stage.

In this work, a new technology to develop mammographic biomarkers and algorithm to diagnosis breast cancer by analyze Digital Image Mammograms (DIM).

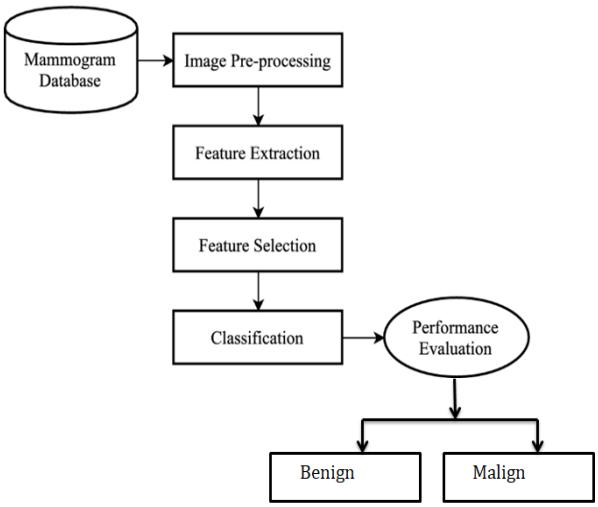

Fig -2: *Mammographic feature extraction Process and artificial neural network training*

4.1 Mammogram Database

The expansions of algorithm systems use their generation and validation by using mammograms got from clinical studies or using public databases

Regardless, at the international level, there are only a few databases available for the research community to examine. As mentioned, we found no public database currently available for BC in India for performing this kind of study. Hence, different public mammography databases are used for developing and validating digital image processing algorithms at the very first stage to select ROIs from mammograms and dragging image features used to train a GRANN capable to diagnose Breast Cancer (BC) as an aid for radiologists. Databases were used to develop and validate the evolution of a biomarker, an artificial neural network approach with gradual learning and with both, the creation of a methodology carried out in a general scope.

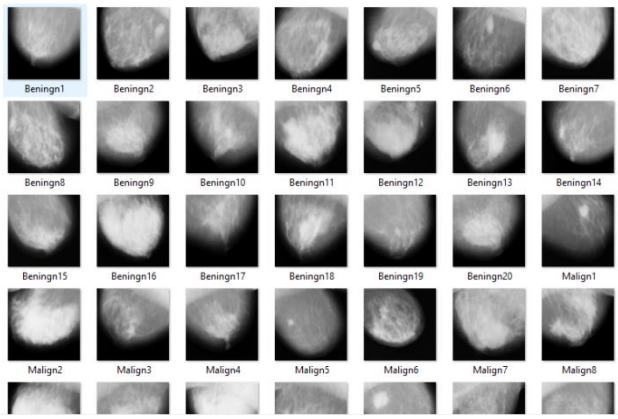

Fig -3: *Mammographic Images Database*

Moreover, it has been certified by the scientific community, Breast Cancer (BC) varies widely between different etiologies and may prove that methods created for a population may not work for a further population in the way they were supposed. In the second and third stages of this research, the distilled for its operation and the designed methodology will be focused on patients. In this section, the database of random patients presented the results which are obtained from it. The database contains some useful information for each type of mammogram such as gender: masculine or feminine; segmentation of mammogram, marked in red pixels the ROI that contains the lesion found by the radiologist; the age of the patient; breast density, i.e., the percentage of breast density according to Breast Imaging Reporting and Data System (BI-RADS) standard expressed as a percentage of glandular and fibrous tissue; breast localization, depending on the location of breast of the RIO with the lesions; mammography, the type of lesion found by the mammographic image expert and BI-RADS classification of the lesion is done after that finally, intensity and shape descriptors of ROI. Nevertheless, it is very important to mention that, these descriptors were not used to train the neural network for this type of research. Rather, a set of computer algorithms were developed to extract image descriptors of ROI of mammograms as explained in a later section.

In this project we will be using adaptive mean filter to remove noise from images, since it is better among all the spatial filters andit also distinguishes fine details from noise. The Adaptive Median Filter classifies pixels as noise by comparing each pixel in the image to its surrounding neighbor pixels.

The size of the neighborhood is adjustable, as well as there is threshold for the comparison. A pixel that is different from a other neighbors pixels, not actually structurally aligned with these similar pixels, which is labeled as impulse noise. After that Noise Labeling test is conducted on pixels and noise pixels replaced by the median pixel value of the pixels in their neighborhood.

4.2 Artifact Removing & Segmentation(Image Preprocessing)

A mammogram is the image that can be considered as a representation of the X-ray radiation density that reflects the tissue of the breast.
A breast cancer patient is recognized by the white region which appears on the mammogram image, which is considered a risk aspect in breast cancer patients which means a high tissue density appears that may be considered abnormal. A breast abbreviation is commonly known as an ROI technique. According to DIP, doing the segmentation of breast abnormalities on a mammogram is a crucial step in algorithm designs. Sometimes it is a very challenging task to identify the edges of a suspicious mass because these types of medical images have very low-intensity contrast. The methodology is used in this research, in only lateral mammographic images are taken from the database were used. In all determined images, digital mammographic images of the database can be accessed in two forms the first one from films(photographic films) and a second one from digital images taken from films n the X-ray system which we considered as a bengin or malignant when there is a lesion exist in the image films images are used by radiologists to identify the left and right breasts and this film's images require the design of digital image processing algorithms to eliminate relics which we identify by red pixels and pre noise such as labels. Contrarily, digital mammography images only need the design of algorithms for clearing red pixels. The film's method improves digital mammography images by increasing the high frequency and eliminating the noise and undesirable relics in the ROI. at preprocessing stage, software was designed to automate the preprocessing of film digital mammographic images (FDMIs). All FDMIs are ministered to destroy image relics such as background, noise, and image labels. In the FDMIs, the region of the breast and other regions with labels as well as artifacts on the mammography all are used as a common threshold for application.

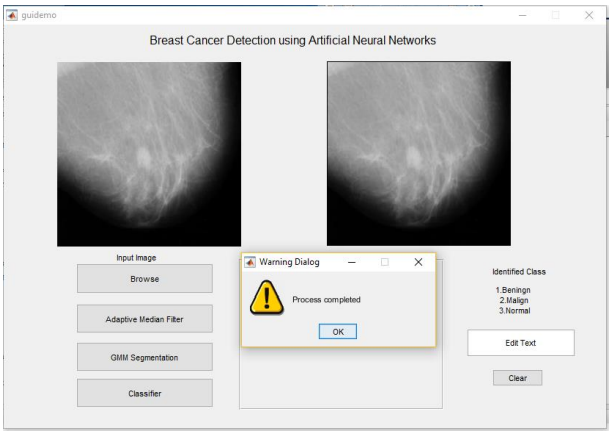

Fig -4: *Preprocessing on mammographic image*

By using the designed Software algorithm when we create the logical image so we try to remove all small regions (less than 10,000 pixels) that are destroyed.

Then, the Breast Cancer (BC) image is created by using a mask image instead of the actual image without relics and labels. Figure 3 of the digital mammography, the image shows the preprocessing method to remove noise and labels.
Transforming a greyscale image to a digital or logical idea is a standard task of digital image processing.

There are many methods for estimating the threshold value for creating analytic images. As shown in figure 4, in this work, the threshold value was calculated by converting the Non-Zero pixels values to 1. By converting the gray tones into white levels we created a logical image that contains the ROI and the pectoral muscle. By destroying the white region which is connected to the border of the binary image we can remove the pectoral region in the logical image, as shown in figure 4. This remaining white region represents the ROI detected in the mammography image. After a cleaned image, the next part is the segmentation process.
On another side, the digital mammographic technique works as follows:

After the processing stage, the image is ready for segmentation. Because the pixels are in red turned into black. When the black pixels are used then the ROI is separated from the rest of the breast image for making a segmentation of the ROI as shown in figure 6.
For the segmentation of the ROI, we create a binary and logical image with a very high binarization threshold where we found that gray levels become white.

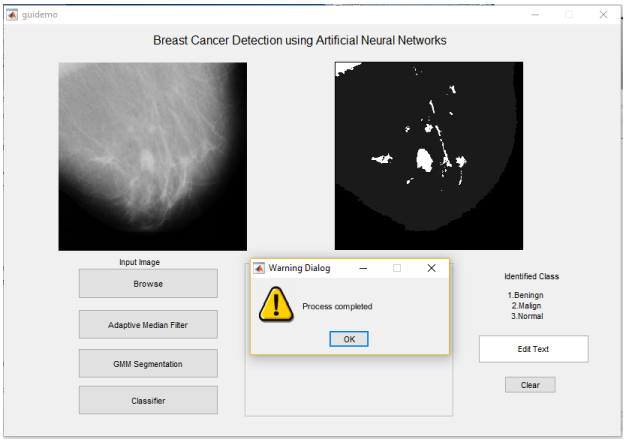

Fig -5: *Brest region removing process.*

This approach is very helpful in considering most of the gray pixels of the image looking not to lose many pixels from ROI. After that, the white logical region which is connected to the edge of the mammographic image is removed.

Some white pixels about the outline of the breast are scrapped when the pixels in the image with a smallish area are terminated. Finally, the white region with a large number of pixels is dragged, which would be viewed as neoplasia.
Next step, using the ROI which is obtained in the segmentation stage a binary image is created.

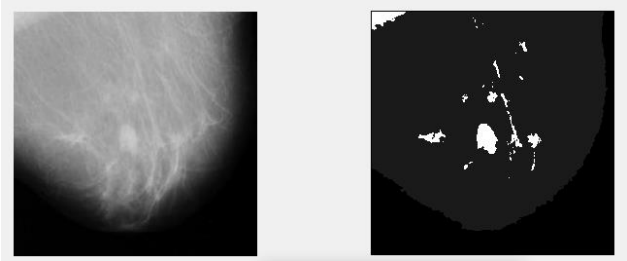

Fig -6: *Binary mask and ROI in tones of grays.*

Now, we will get all the ROI in the shades of gray as is shown in fig -6. With the mask and the complete image together in shades of gray. The next step is the feature extraction of the ROI in the operation of regular software systems. The feature extraction process is defined as the assumption and quantifies the parameters that characterize the object will be studied.

The feature extraction donates to the research of the ROI. It is possible to quantify and analyze different factors like the shape, texture, size, border, and other tissue parameters so that these features can contribute to the accurate diagnosis and detection of a Brest cancer with its risk factor. As is shown in Figure 7, in this work, shape, intensity, and texture features were extracted to create a biomarker for BCD using a software algorithm system that uses AI technology.

The image features of all Digital Image Mammography (DIM) of the database were dragged and utilized to produce a biomarker to instruct an ANN. The digital images are generally available in RGB and gray-level digitalized in JEPG formats. The RGB mammograms are used to show the red stated area by a radiologist to delimit the discovered abnormality.

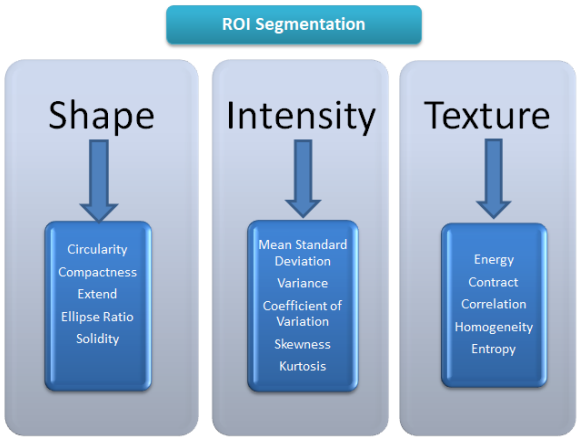

Fig -7: *Image features extraction Process*

The segmentation process uses the red section in the RGB mediolateral oblique view mammograms to obtain the ROI.
In the RGB mammograms, all the types of pixels such as red pixels in the image and the pixels which are outside the original red region were removed. Finally, the remaining pixels in the gray-level mammogram are helpful to get the ROI that is used for the calculation of the feature.

4.3 Features Shape, Intensity,& Texture (Feature Extraction)

Image processing is one of the most important regions where the feature extraction method is applied, in this, we used different mathematical algorithms to detect and separate various desired portions or forms, parts, of digitalized images or video streams, and it has equal importance in the area of design recognition or symbol recognition. The feature extraction method we used to measure the physical parameters is visualized in a segmented region of the given image. The main purpose of feature extraction for the image information is to find out the mathematical way, which has equal importance, for a computational task solving. In BCD, these types of features are very helpful in determining the tumor which is detected in a mammogram image.
The choosing right feature has a crucial impact on the accuracy of the various factors such as classification, the time taken for the classification, the number of examples which is needed for our learning, and the cost which is required in performing the classifications.
In breast abnormalities, classification of the differences in mass between benign and malignant on mammography can be distinguished from their shape, textures, and the intensity in the image. In this analysis, we use automated Software that is designed to calculate the different shapes, intensity as well as texture features from the ROI which is extracted from Database mammograms. The shape features of MDI use the border and the inside pixels of the ROI.
These types of descriptors has a valid meaning in binary or logical images and we use some simple shape features to describe a ratio between some geometrical figures; such as extend, ellipse ratio, and solidity.
The most common shape features are applied when the ROI size is invariant which are area and the perimeter. However, the area and perimeter can be used to create a relationship between circularity and compactness.

The intensity features use the shape intensity histogram used to get information that helps describe the image; i.e., these features have used the probability and statistics of value this is based on pixels in that image. The standard is the intermediate intensity level.
The standard is used to quantify the amount of divergence of the set of assertiveness grades. The conflict guides to the variation of the intensities around the mean value. The coefficient of divergence is a standardized calculation of diffusion in the values.
Texture features attempt to capture features of the intensity which changes between groups of neighboring pixels.
The human eye is very sensitive to texture. In this type of analysis, we use various types of relations such as energy, contrast, correlation, homogeneity as well as entropy is used. The image is a measure of the textural uniformity of that energy. The contrast is referred as difference in luminance in the ROI. Correlation texture estimates the reliance of gray tiers on those of adjacent pixels. Homogeneity calculates the similarity of values in the ROI. Entropy calculates the infection of value pixels of an image. As defined early, medical diagnosis plays a very important and complicated task that has to be executed perfectly and effectively.

At present, new methods based on data mining, KDD, and AI in healthcare are being used primarily for indicating different diseases as well as helping doctors in diagnosis in their clinical decision.

4.4 Classification and Performance Evaluation

For automatic classification of BC on DIM, an ANN Algorithm is used to separate malignant and benign tumors. ANN drops into the type of probabilistic neural networks (PNN). ANN is one-step only learning that can solve any function approximation problem of neural network architecture.

The learning process is comparable to encountering a surface in a multidimensional space that supplies the best fit for the training data. During the training process, we just try to store training data, and then later we will try to use it for predictions. This neural net is very helpful in the comparisons of system performance as well as to perform predictions. In ANN architecture, after the network is trained there are no training parameters just a smoothing factor (σ). The selection of this aspect is very.
In this research, as shown in figure 8, an ANN used a data set of random mammograms which is extracted from the public database and is trained and tested. For each mammogram, using automated Software specifically we calculated some image descriptors. These types of image features are used to train the neural net to classify benign and malignant BC for decision-making in BDC. As shown in fig 8, these image features were used as entrance data, and these malignant (cancerous) and benign (noncancerous) samples were used as output data. To train any type of cancer network, the dataset was unsystematically divided into the form of two subsets, in which one with 80% of the instances for training and other with the 20% of instances for testing. After training the number of network pieces, a smoothing factor equal to 1e4 was also calculated. This smoothing training value is used for training the neural network which is reaching an accuracy of 96%. The result we find from this work shows that ANN is a favorable and rich system for BCD.

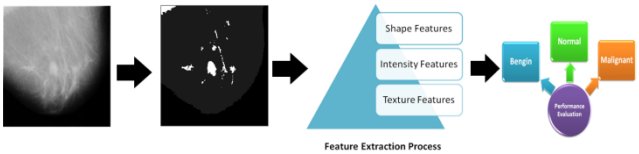

Fig -8: *Training with Database Images and Evaluating result by feature extraction*

There are two main stages in the Our ANN Based software algorithm Diagnosis system. The first part represents the results for the radiologist of the suspicious region which is available with high sensitivity. The preprocessing algorithm which is based on DIP techniques is used to designed to reduce the noise acquired in the image and its improvement, and then it executes the segmentation process of different ROIs designed which is helpful in high suspicion of some signs of cancer. This information obtained from the segmentation process is helpful in the classification of

positive or negative predictions of Breast Cancer (BC) which is obtained through ANN.
After finishing the evolution of this software algorithm, we will use it at real workplaces such as GHz, making verification of the prediction which is obtained by the neural network it is also compared with predictions made by specialized oncologists.

The main purpose of developing this software algorithm system for BC patients is to expand the understanding of BCD with image mammograms information, clinical data, risk factors, biopsy results, genomic information, etc., as is shown in fig 9

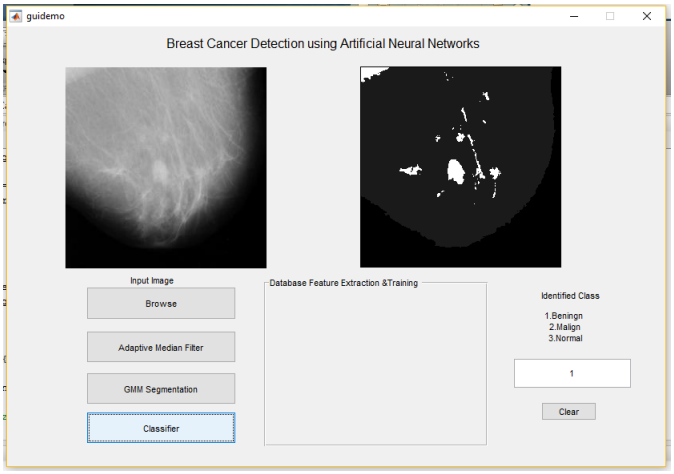

Fig 9 -: *system developed based on ANN Algorithm*

5. CONCLUSION

Worldwide the reports of BCD and classification is performed or analyzed by a clinician or a pathologist only using the traditionally by observing stained biopsy images under the microscope. However, this method is time consuming and also can lead to erroneous results. Therefore, there is a need to develop intelligent and automated technology for it. By considering this is an base problem we try to design this system which uses cross-dimensional, KDD uses algorithms and techniques from a vast array of fields like soft computing, pattern recognition, machine learning statistics, artificial intelligence (AI), and get the maximum accurate result like Benign or Malignant.

REFERENCES

[1] S. Willium, Paper named as "Network Security and Communication", IEEE Transaction, Vol.31**,** Issue.**4**, pp.**123-141**, in **2012**. For Journal

[2] R. Solanki, Article "Principle of Data Mining" & McGraw-Hill Publication, **India**, pp. 386-398, in 1998. For Book

[3] M. Mohammad, Published Research on "Performance Impact of Addressing Modes on Encryption Algorithms", In 2001 IEEE International Conference on Computer Design (ICCD 2001), Indore, **USA**, pp.**542-545**, in **2001.** For Conference

[4] S.K. Sharma, Published Article on "Performance Analysis of Reactive and Proactive Routing Protocols for Mobile Ad-hoc –Networks", In International Journal of Scientific Research Under Network Security and Communication, Vol.**1,** No.**5**, pp.**1-4**, In the Year of **2013**.

[5] S.L. Mewada, Published Research on "Exploration of Efficient Symmetric AES Algorithm", International Journal of Computer Sciences and Engineering, Vol.**4**, Issue.**11**,pp.**111-117**, in the year **2015**.

[6] A. Mardin, T. Anwar, B. Anwer, Published Research on "Image Compression: Combination of Discrete Transformation and Matrix Reduction", In International Journal of Computer Sciences and Engineering, Vol.**5,** Issue.**1**, pp.**1-6**, In the year **2017**.

[7] H.R. Singh, Published Research on "Randomly Generated Algorithms and Dynamic Connections", In International Journal of Scientific Research in Network Security and Communication, Vol.**2,** Issue.**1**, pp.**231-238**, In the year of **2014.**

[8] Thomas L., Published Research on "A Scheme to Eliminate Redundant Rebroadcast and Reduce Transmission Delay Using Binary Exponential Algorithm in Ad-Hoc Wireless Networks", In International Journal of Computer Sciences and Engineering, Vol.**3,** Issue.**8**, pp.**1-6**, In the year of **2017**.

[9] C.T. Lee, A. Girgensohn, J. Zhang, Published Research on "Browsers to support awareness and Social Interaction," In Computer Graphics and Applications, Journal of IEEE Access , Vol.**24,** Issue.**10**, pp.**66-75**, In the year of **2012.**

[10] Lin C., Lee B., Published Research on "Exploration of Routing Protocols in Wireless Mesh Network", In 2015 IEEE Symposium on Colossal Big Data Analysis and Networking Security, Canada, pp.**111-117**, In the year of **2015**.

[11] S. Tamilarasan, P.K. Sharma, Published Article on "A Survey on Dynamic Resource Allocation in MIMO Heterogeneous Cognitive Radio Networks based on Priority Scheduling" International Journal of Computer Sciences and Engineering, Vol.**5,** No.**1**,pp.**53-59**, In the year of **2017**.

[12] Jun Deng, PhD Professor Department of Therapeutic Radiology in Yale University School of Medicine In the year of November 4, 2017, Published Research, under International Journal of Advances in Science Engineering and Technology.

[13] Medisetty Hari krishna, Dr. Kunjam. Nageswara Rao M. Tech Student, ,Professor Department of Computer Science and Systems Engineering in Andhra University College of Engineering, Published Research Visakhapatnam, AP, India in International Journal of Distributed and Parallel Systems (IJDPS) Vol.4, No.3, In the year of May 2013.

[14] Hindawi Journal of Healthcare Engineering Volume In the year of 2019, Article ID 4253641, kram GAGAOUA Tolga Published Research under Computer Engineering, Istanbul University, Istanbul, Turkey. Meriem AMRANE1 Saliha OUKID Computer Science Department, LRDSI Laboratory, University of Blida 1, Blida, Algeria

[15] IEEE Third International Conference on Circuits, Control, Communication and Computing

[16] International Journal of Intelligent Systems and Applications in Engineering Advanced Technology and ScienceISSN:2147-67992147-6799

[17] Joseph A. Cruz, David S. Wishart Published Research under Departments of Biological Science and Computing Science, University of Alberta Edmonton, AB, Canada T6G 2E8

[18] Department of Computer Science, Brunel University London, Uxbridge, Middlesex UB8 3PH, UK; 2 School of Mathematics, Southeast University, Nanjing 210096, China; hwchen_seu@163.com Published Research on 9 May 2018.

[19] Bibhuprasad Sahu1, Sachi Nandan Mohanty and Saroj Kumar Rout Published Research under Research Scholar, North Orissa University, From Gandhi Institute for Technology, Bhubaneswar, Odisha , India.

[20] Published Research under International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009 Volume-5, Issue-3, In the Year of Jul.-2017

[21] Ali Al Bataineh. Published Research under International Journal of Machine Learning and Computing, Vol. 9, No. 3, In the Year of June 2019.

[22] Published Research under International Journal for Research in Applied Science & Engineering Technology (IJRASET)

[23] Saveetha School of Engineering, Published Research under UG Scholar, Saveetha Institute of Medical and Technical Sciences, Thandalam, Chennai, Tamilnadu ISSN: 0193-4120Page No. 6667- 6670Volume 82 Page Number: 6667 - 6670 Publication Issue: In the Year January-February2020.

[24] Procedia Computer Science 83 (2016) 1064 – 1069 1877-0509 © In the Year 2016 The Authors.