A Review of AI Methods for the Diagnosis of Breast Cancer by Image Processing

Mansi Yadav[1], Mr. Sarad Nigam[2]

[1]Research Scholar, SHEAT College of Engineering, Varanasi, India
[2]Professor, SHEAT College of Engineering, Varanasi, India

A R T I C L E I N F O

A B S T R A C T

Breast cancer is the most common cancer among women around the world. Despite enormous medical progress, breast cancer has still remained the second leading cause of death worldwide; thus, its early diagnosis has a significant impact on reducing mortality. However, it is often difficult to diagnose breast abnormalities. Different tools such as mammography, ultrasound, and thermography have been developed to screen breast cancer. In this way, the computer helps radiologists identify chest abnormalities more efficiently using image processing and artificial intelligence (AI) tools. This article examined various methods of AI using image processing to diagnose breast cancer. The results were provided in tables to demonstrate different techniques and their results over recent years. In this study, 18,651 articles were extracted from 2007 to 2017. Among them, those that used similar techniques and reported similar results were excluded and 40 articles were finally examined. Since each of the articles used image processing, a list of features related to the image used in each article was also provided. The results showed that support vector machines had the highest accuracy percentage for different types of images (ultrasound =95.85%, mammography =93.069%, thermography =100%). Computerized diagnosis of breast cancer has greatly contributed to the development of medicine, is constantly being used by radiologists, and is clear in ethical and medical fields with regard to its effects. Computer-assisted methods increase diagnosis accuracy by reducing false positives.

Keywords: Breast Cancer, Breast Cancer Screening Techniques, Artificial Intelligence Techniques, Medical Image Processing

I. INTRODUCTION

Breast cancer is the most common cancer and the second leading cause of death among women around the world.1 Breast cancer occurs when the cell tissues of the breast become abnormal and uncontrollably divided. These abnormal cells form a large lump of tissues, which consequently becomes a tumor.2 It was reported that 1.7 million cases of breast cancer were identified in the world in 2012. Breast cancer is the second cause of cancer death with the standardized mortality rate of 12.9 per 100,000, and its incidence has increased over the years.3–5

Breast cancer could be successfully treated if detected early. Thus, it is of importance to have appropriate methods for screening the earliest signs of breast cancer.6 There are various imaging methods available for the screening and diagnosis of breast cancer, the most important of which are mammography, ultrasound, and thermography.7 Mammography is one of the most important early diagnosis methods for breast cancer. Since mammography is not very successful for dense breasts, ultrasound or diagnostic sonography techniques are recommended.8 Taking into account that small masses may pass radiations of radiography, thermography can be more powerful in diagnosing smaller cancerous masses than the ultrasound technique.9

There is no doubt that evaluation of patient data and expert judgment is the most important factor in image-based diagnosis; however, there are many factors affecting this type of diagnosis. Among the factors affecting image-based diagnosis are the presence of noise in images, the radiologist's visual perception ability, inadequate clarity, poor contrast, and the less experienced radiologist.10

Due to inherent problems associated with an image, including poor contrast, noise, and lack of recognition with the eye, tools have been developed to create and develop image processing. Currently, medical image processing is one of the fastest growing areas in the health care sector. The purpose of the image processing is to use techniques for making proper images of the human body, which are reliable for use in the diagnosis and treatment processes.6

In the early 1980s, there was an increment in the use of neural networks in the field of image and signal processing. Since diagnosis of breast cancer is very difficult, statistical methods and artificial intelligence (AI) techniques can be very important in this regard.11 The AI is said to be an artificial intelligent machine in various situations. In other words, these are systems that can respond to similar conditions such as an intelligent human, including understanding complex situations, simulating thinking processes and human reasoning methods, and demonstrating accurate responding, learning and ability to acquire knowledge, and reasoning for solving problems.12 For example, Dheeba and Selvi26 used a particle swarm-optimized wavelet neural network (PSOWNN) to identify breast cancer on the mammogram. This method was applied with real data and had a sensitivity and accuracy of 94% and 92%, respectively. Moreover, the result showed that the area under the receiver operating characteristic (ROC) curve of the proposed algorithm was 0.96, indicating the excellent performance of the system. In addition, new tools including image processing tools have been developed to facilitate the diagnosis of breast cancer masses. Image processing methods identify abnormal features in medical images more easily. By using image processing, pattern recognition, and AI, researchers have been able to provide techniques that accurately detect masses.13

This article examined various AI techniques that use medical images to detect breast cancer. The findings are presented in the form of tables with the aim of demonstrating different methods and results of using each method in recent years.

II. MATERIALS AND METHODS

The search strategy was performed by searching the databases such as Medical Literature Analysis and Retrieval System Online (MEDLINE) via PubMed,

Springer, IEEE, ScienceDirect, and Gray Literature (including Google Scholar, articles published in conferences, government technical reports, and other materials not controlled by scientific publishers) for relevant publications from 2007 to 2017.

Keyword searches based on Mesh included "breast cancer", "breast cancer screening techniques", "artificial intelligence techniques", and "medical image processing". The symbol "*" was also used to allow retrieving all variations with suffixes of the source words. The above terms were combined using the logical connectives "AND", "OR", and "NOT".

Among the articles searched, non-English articles as well as articles that used similar techniques and reported similar results were excluded. Moreover, articles whose full texts were not available were also excluded.

Two experts independently reviewed all potentially relevant studies. Disagreements were solved with discussion and by using the viewpoint of a third expert. Finally, 18,651 articles were extracted; however, majority of the extracted articles were deleted due to being repetitive and not having access to their full texts, and, 40 articles remained for review (Figure 1 shows the process.). The selected articles were examined in terms of the name of the method used, type of the image, advantages and limitations of each method, features used, and application results of each method, and all the results were presented in the form of tables.
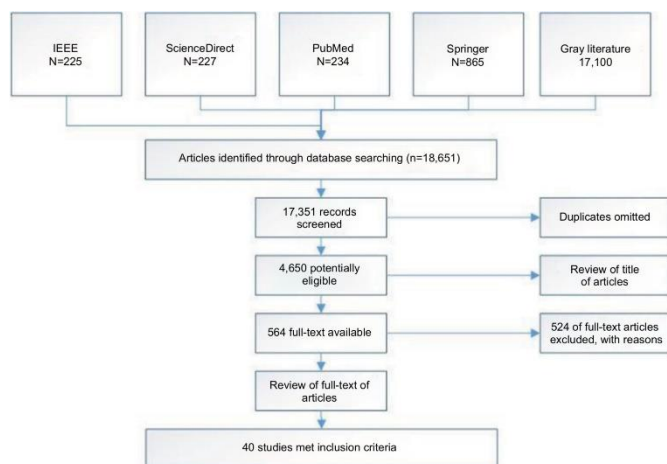

Figure 1 Stages of systematic review.

III. RESULTS

If we discard skin cancer, breast cancer is the largest cause of cancer among women, accounting for one-third of all the cancer types. Obtaining the best outcomes in breast cancer depends on early diagnosis. Therefore, imaging techniques have been developed to increase the likelihood of early diagnosis of breast cancer and reduce unnecessary biopsy.[14] The importance and necessity of processing digital images are examined in the following two directions: 1) to improve images for human interpretation and 2) to process images for automatic understanding and interpretation by the machine.[15]

In the field of diagnosis, first, medical images are collected and, then, preprocessing, segmentation, extraction of features, and eventually categorization are performed.[9,16] Figure 2 shows the steps involved.
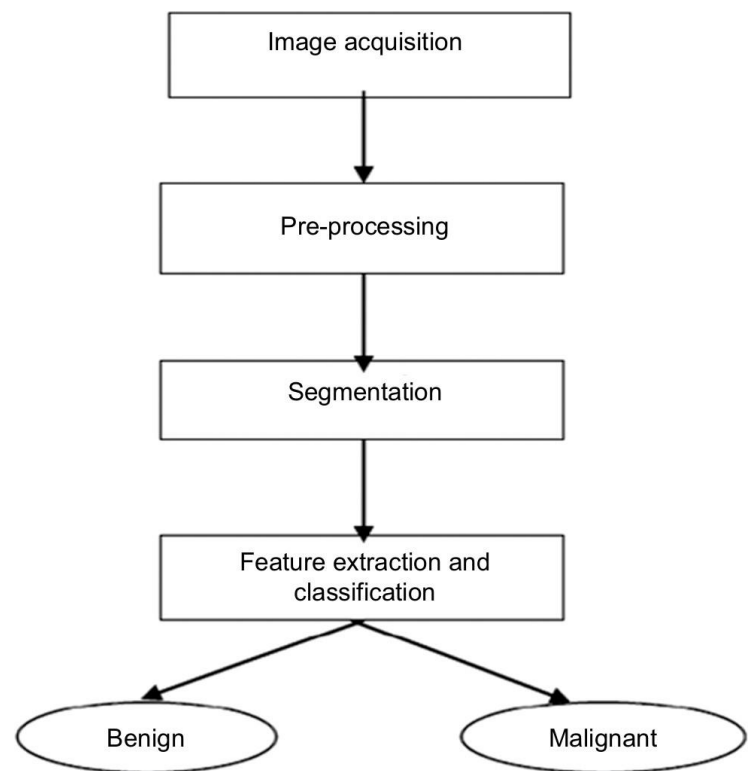

Figure 2 Stages of cancer detection by image processing.

Image acquisition

The first step in processing images is to capture an image. In this step, data are collected in the form of

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 5

196

digital images. The image format is usually a portable gray map, which is fixed format, and does not erase the image data when compressing images.[9,16]

Image preprocessing

The next step is to preprocess input images to improve their quality by eliminating noise. Preprocessing of images is done through the middle filter. Preprocessing, in addition to deleting or reducing noise, improves image quality through increased contrast.[17] Some of the most important preprocessing techniques are presented in Table 2.

Image segmentation

Segmentation refers to cutting an image and dividing it into its constructive parts. In image analysis and processing, the most important step is to fragment an image. Segmentation is actually dividing an image into sub-sections that can be similar or different depending on its features. The output depends largely on the accuracy of the measurements of the features.[18]

Feature extraction

The conversion of input data into a set of extracted features is called feature extraction. There are many ways to extract features. Some of the most commonly used methods are as follows: 1) spatial features, 2) transform feature, 3) edge and boundary features, 4) color features, 5) shape features, and 6) texture features.

Feature extraction techniques play a very important role in the diagnosis of disorders. Tissue features are very useful in differentiating the mass from the natural tissue of the breast. These categories of features are capable of separating natural and abnormal lesions of the masses or microcalcifications.[16,19] Table 4 refers to some of the most important features of the tissue.

IV. INTRODUCTION OF VARIOUS AI TECHNIQUES IN IMAGE PROCESSING

Support vector machine (SVM)

One of the most widely used techniques for diagnosing breast cancer is the SVM. This technique is a brilliant among the learning algorithms inspired by the statistical learning theory and has been incorporated in the machine learning set in recent decades. In this way, the overfitting problem in the training data is reduced and it is possible to identify a large training set with small subsets of training points. Moreover, this technique can operate on optional features with no need to create independent hypotheses. In the fuzzy SVM, each sample of {xi, yi} is weighted in the training data set using the fuzzy membership function. This set becomes {xi, yi, si}, where si is a fuzzy membership function, which indicates that the membership of this sample is assured by one or more classes. In other words, each training example {xi, yi} is associated with a fuzzy membership function ($0 \leq si \leq 1$).[20–22]

Cascade forward back-propagation network

The cascade forward back-propagation model uses the postpropagation algorithm to update weights such as back-propagation neural network; however, the main characteristic of this network is that each layer of neurons is linked to all previous neuron layers.[24]

Feed forward back-propagation network

This model includes input, output, and hidden layers. The back-propagation learning algorithm is used for learning in these networks. During the training, computation is performed from the input layer to the output layer and error values are released to the previous layer.[24]

k-nearest neighbor (k-NN)

The *k*-NN selects a group of K records from a training record set that has the closest records to the test record and decides about the class of the test record based on their rank or label superiority. Simply put, this method selects the rank with the highest number of records in the selected neighborhood.[25]

Genetic algorithm as optimizer

This algorithm can quickly scan a set of broad solutions and eliminate bad suggestions without negatively affecting the final result. The genetic algorithm works

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 5

197

on its own rules and therefore can be used for problems that are defined as irregular.[22,26]

Naive Bayes classifier

Bayes classifier is estimated by the covariance matrix and is a simple probability classifier based on the theory of Bayes with powerful independent assumptions. The advantage is that this model requires a small amount of training data to estimate the required classification parameters.[27]

Deep learning technology

A deep learning network contains more image processing layers than the conventional image feature-based machine learning classifiers. Each layer is a typical neural network, such as a convolution neural network. Instead of using a set of manually or automatically selected image features computed from images, the deep learning network utilizes the image itself as a single input.[28] Effective image features are automatically learned and extracted with lower layer networks. Accordingly, higher layer networks use the extracted feature patterns and classify images into different target categories.[29]

A number of AI techniques used to classify breast cancer into two malignant and benign forms using medical images are presented in Table 5. In this table, for each technique, the type of image, along with its strengths and weaknesses, is mentioned. Since each of the discussed articles used image processing, a list of image features was also provided. Investigations showed that the SVM classification method had a higher accuracy than the other methods. The results also indicated that the method had the highest accuracy for different types of images (98.58% ultrasound, 93.063% mammography, and 100% thermography).

The highest accuracy in the SVM method was observed in the results of a research, which used an appropriate segmentation method for obtaining the desired area in the image. The shape and intensity of the extracted features had the most effect in the classification. The combination of gray-level co-occurrence matrix (GLCM) and Pratio features along with morphological features resulted in the highest accuracy.

V. DISCUSSION

Computerized diagnosis of breast cancer has a number of great advantages; thus, it is constantly being used by radiologists and its impact on the field of medicine is clear. Computer-assisted methods increase diagnosis accuracy by reducing false positives. Advantages of using developed systems for diagnosing breast cancer include 1) helping radiologists in the process of interpretation and screening as a second interpreter after the radiologist; 2) reducing the number of false positives, which will eliminate the need for unnecessary biopsy and lead to cost savings; and 3) reducing the time of the patient's examination by reviewing and reporting the findings in a few seconds.

So far, creative methods have been developed to diagnose and classify breast cancer; however, none of the methods has been able to accurately classify all cancer cases. In recent years, application of the AI techniques along with image processing has yielded significant results. This article examined these techniques, which used image processing to diagnose breast cancer. The results of the study showed that the SVM had the highest accuracy in cancer diagnosis.

The SVM theory has some limitations. Typically, we know that each sample {xi, yi} belongs to one class in a training dataset. For example, the value of yi is set to only 1 or −1. In this way, in the learning process of the SVM, all instances in the training dataset will be grouped into a single class.

The SVM was invented by Vladimir Vapnik in 1963 in the field of statistical learning theory. This model has been able to succeed in various classifications and in predicting different problems. The model can be used in many regression estimation and pattern recognition problems and is also used to predict and build smart machines. This method is considered as a very efficient

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 5

198

method when it comes to solving classification problems with noise data. Two of the main reasons behind the reliability of the SVM classification engine include 1) choosing an optimal subset of context properties for learning and 2) appropriately regulating page parameters using the v-fold cross-validation approach.

The main advantage of the SVM is the ability to use it in a data set that has many features while only a few of them may be involved in the training process. In addition, the SVM has several other notable advantages, including 1) having a large number of nonlinear split pages that give it a higher degree of differentiation; 2) having high generalization ability for the classification of hidden data; and 3) having the ability to determine the optimal network structure (such as its hidden layers and neurons) without tuning external parameters.

As the popularity of the SVM is on the rise, various applications have been published to facilitate a broader view of the model in practical and academic terms. However, many scholars have cited some of its limitations including the following: 1) difficulty in choosing the kernel function for a problem; 2) low performance of the machine in training and testing; 3) lower covariance in the test stage; 4) difficulty in choosing the appropriate kernel parameters; 5) requiring a lot of memory space to run the model; 6) requiring to select one of two parametric and nonparametric methods to run the model; and 7) dependence of the function of each learning machine on the problem. The reason for this is that the function is based on multifactors such as the test dataset, selected optimal subset of the model, and method by which the data samples are divided between the test sets and training sets.

VI. CONCLUSION

Although the diagnosis of breast cancer can be highly accurate, it is not necessarily the same as the results obtained in other sets of images. Therefore, future research can be done to improve the system's

performance and validate it by conducting tests on a larger set of images. In addition, the following can be suggested as future work: 1) setting SVM parameters through using a genetic algorithm: traditionally, feature selection and parameter optimization are performed independently; independently performing these two problems may result in a loss of information related to the classification process; motivated by these views, the trend in recent years is to simultaneously select feature subsets and optimize parameters of SVMs[30]; genetic algorithms have the potential to generate both the optimal feature subset and SVM parameters simultaneously; and the most widely used genetic algorithm is proposed by Huang et al[31] and 2) evaluating other texture approaches: texture extraction methods are classified into three main categories: structural, statistical, and spectral.

First, in structural approaches, "texture primitive", as a basic element of texture, is used to construct more complex texture patterns with grammar rules that determine the generation of texture patterns. Second, in the statistical approaches, for the discrimination of various textures, items such as gray-level histogram moments and statistics based on the GLCM are calculated. Finally, in the spectral approaches, the textured image is converted into frequency domains. Then, the extraction of texture features can be done by analyzing the power spectrum. Markov random field models and texture spectrum are some other texture descriptors.[32] Then, the extraction of the texture features can be done by analyzing the power spectrum by methods such as the Markov model and the texture spectrum.

VII. REFERENCES

[1]. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–E386.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 5

199

[2]. Mehdy MM, Ng PY, Shair EF, Saleh NIM, Gomes C. Artificial neural networks in image processing for early detection of breast cancer. Comput Math Methods Med. 2017;2017:1–15.

[3]. Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and Mortality and Epidemiology of Breast Cancer in the World. Asian Pac J Cancer Prev. 2016;17(S3):43–46.

[4]. Oliver A, Freixenet J, Martí R, et al. A novel breast tissue density classification methodology. IEEE Trans Inf Technol Biomed. 2008;12(1):55–65.

[5]. Tarique M, Elzahra F, Hateem A, Mohammad M. Fourier Transform Based Early Detection of Breast Cancer by Mammogram Image Processing. J Biomed Eng Med Imaging. 2015;2(4):17.

[6]. American Cancer Society [webpage on the Internet]. How is Breast Cancer Diagnosed? 2014. Available from: http://www.cancer.org/cancer/breastcancer/deta iledguide/breast-cancer-diagnosis. Accessed September 20, 2017.

[7]. Hou MF, Chuang HY, Ou-Yang F, et al. Comparison of breast mammography, sonography and physical examination for screening women at high risk of breast cancer in Taiwan. Ultrasound Med Biol. 2002;28(4):415–420.

[8]. Haddadnia J, Hashemian M, Hassanpour K. Diagnosis of breast cancer using a combination of genetic algorithm and artificial neural network in medical infrared thermal imaging. Iran J Med Phys. 2012;9(4):265–274.

[9]. Giger ML. Computer-Aided Diagnosis in Radiology. Acad Radiol. 2002;9(1):1–3.

[10]. Pendharkar P, Rodger J, Yaverbaum G, Herman N, Benner M. Association, statistical, mathematical and neural approaches for mining breast cancer patterns. Expert Syst Appl. 1999;17(3):223–232.

[11]. Poole DL, Mackworth AK. Artificial Intelligence: Foundations of Computational Agents: New York: Cambridge University Press; 2010.

[12]. Quintanilla-Dominguez J, Cortina-Januchs MG, Ojeda-magaña B, Jevtić A, Vega-Corona A, Andina D, editors. Microcalcification detection applying artificial neural networks and mathematical morphology in digital mammograms. In: 2010 World Automation Congress. New Jersey: IEEE; 2010.

[13]. Jalalian A, Mashohor SB, Mahmud HR, Saripan MI, Ramli AR, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clin Imaging. 2013;37(3):420–426.

[14]. Kumar A, Shaik F. Image Processing in Diabetic Related Causes. Singapore: Springer; 2015.

[15]. Pradeep N, Girisha H, Sreepathi B, Karibasappa K. Feature extraction of mammograms. Int J Bioinform Res. 2012;4(1):241–244.

[16]. Ibarra-Castanedo C, Bendada A, Maldague X. Thermographic image processing for NDT. In: Paper presented at: IV Conferencia Panamericana de END; 2007; Buenos Aires, Argentina.

[17]. Sonawane M, Dhawale C. A brief survey on image segmentation methods. In: IJCA Proceedings on National Conference on Digital Image and Signal Processing. New York: IJCA Journal; 2015.

[18]. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8–17.

[19]. Saini S, Vijay R. Mammogram analysis using feed-forward back propagation and cascade-forward back propagation artificial neural network. In: Proceedings; 2015 Fifth International Conference on Communication Systems and Network Technologies. New York: IEEE; 2015:1170–1180.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 5

200

[20]. Medjahed SA, Saadi TA, Benyettou A. Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. Int J Computer Applications. 2013;62(1):1–5.

[21]. Karabatak M. A new classifier for breast cancer detection based on Naïve Bayesian. Measurement. 2015;72:32–36.

[22]. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE; 2015:1–9.

[23]. Abdel-Zaher AM, Eldeib AM. Breast cancer classification using deep belief networks. Expert Syst Appl. 2016;46:139–144.

[24]. Danenas P, Garsva G. Selection of Support Vector Machines based classifiers for credit risk domain. Expert Syst Appl. 2015;42(6):3194–3204.

[25]. Xu H, Chen T, Lv J, Guo J. A combined parallel genetic algorithm and support vector machine model for breast cancer detection. J Comput Method Sci Eng. 2016;16(4):773–785.

[26]. Francis SV, Sasikala M, Saranya S. Detection of breast abnormality from thermograms using curvelet transform based feature extraction. J Med Syst. 2014;38(4):23.

[27]. Prabusankarlal KM, Thirumoorthy P, Manavalan R. Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound. Human Centr Comput Inform Sci. 2015;5(1):12.

[28]. Saini S, Vijay R. Mammogram analysis using feed-forward back propagation and cascade-forward back propagation artificial neural network. In: Proceedings; 2015 Fifth International Conference on Communication Systems and Network Technologies. New York: IEEE; 2015:1177–1180.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 5

201