A Survey on Different Neural Network Techniques on Breast Cancer Diagnosis

Yogitha.M,
Department Of BCA,
RNS First Grade College,Bangalore
yogithamudoor@gmail.com

Ravi.K
Department Of BCA,
RNS First Grade College,Bangalore
janravi@yahoo.com

***Abstract*-- Breast cancer is the most commonly occurring cancer in women and the second most common cancer overall. There were over 2 million new cases in 2018.**

Breast cancer diagnosis has been approached by various machine learning techniques for many years. The early diagnosis of BC can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Further accurate classification of benign tumours can prevent patients undergoing unnecessary treatments.

Early detection of this disease can greatly enhance the chances of long-term survival of breast cancer victims. Artificial Neural Networks (ANN) have been widely used for cancer prediction and prognosis. This paper studies various techniques used for the diagnosis of breast cancer using ANN. Different methods for breast cancer detection are explored and their accuracies are compared.

***Keywords*-- Machine learning, Neural networks, Cancer prevention, Breast cancer, Wisconsin breast cancer dataset(WBCD).**

I.INTRODUCTION

Cancer is the second leading cause of death globally, and is responsible for an estimated 9.6 million deaths in 2018. Globally, about 1 in 6 deaths is due to cancer.Approximately 70% of deaths from cancer occur in low- and middle-income countries.[1]
Breast cancer is the most frequently diagnosed cancer among women in 140 of 184 countries worldwide.
Globally, breast cancer now represents one in four of all cancers in women. [2]Breast cancer mortality rates have fallen by 10 percent in five years, according to the latest analysis released by Cancer Research UK[3]
"Cancer Statistics, 2018," published in the American Cancer Society's journal CA: A Cancer Journal for Clinicians, estimates the numbers of new cancer cases and deaths expected in the US this year. The estimates are some of the most widely quoted cancer statistics in the world. The information is also released in a companion report, Cancer Facts and Figures 2018,available on the interactive website, the Cancer Statistics Center. A total of 1,735,350 new cancer cases and 609,640 deaths from cancer are projected to occur in the US in 2018.

The drop in cancer mortality is mostly due to steady reductions in smoking and advances in early detection and treatment. "This new report reiterates where cancer control efforts have worked, particularly the impact of tobacco control," said Otis W. Brawley, M.D., chief medical officer of the American Cancer Society. "A decline in consumption of cigarettes is credited with being the most important factor in the drop in cancer death rates. Strikingly though, tobacco remains by far the leading cause of cancer deaths today, responsible for nearly 3 in 10 cancer deaths." [4]

Most breast cancer cases occur in women aged 40 and above but certain women with high-risk characteristics may develop breast cancer at a younger age[18].Cancer is a disease in which cells become quirky and form more cells in an immoderate way. With breast cancer, the cancer begins in the tissues that make up the breasts. The cancer cells may form a mass called a tumor. They may also invade nearby tissue and spread to lymph nodes and other parts of the body.

The most common kinds of breast cancer are carcinomas, and are named based on where they form and how far they have spread. Ductal carcinoma in situ (DCIS; also known as *intraductal carcinoma*) is a non-invasive or pre-invasive breast cancer. Lobular carcinoma in situ (LCIS) may also be called lobular neoplasia. This breast change is not a cancer, though the name can be confusing. In LCIS, cells that look like cancer cells are growing in the lobules of the milk-producing glands of the breast, but they don't grow through the wall of the lobules. Breast cancers that have spread into surrounding breast tissue are known as invasive breast cancer. There are many different kinds of invasive breast cancer, but the most common are called invasive ductal carcinoma and invasive lobular carcinoma.[6]
About 1 in 10 breast cancers are of this type[19]. There is much research on medical diagnosis of breast cancer with Wisconsin breast cancer database (WBCD) in neural network literature [20]. Existing/popular neural network techniques with WBCD data for the diagnosis of breast cancer are reviewed in this paper.

II. NEURAL NETWORK TECHNIQUES FOR DIAGONOSIS OF BREAST CANCER

Various artificial intelligence techniques have been used to improve the diagnose procedures and to aid the physician's efforts [5].

A.Multilayer perceptron(MLP):

The most common neural network model is the multilayer perceptron (MLP). This type of neural network is known as a supervised network because it requires a desired output in order to learn. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown.[14].

MLP is a class of feed forward neural networks which is trained in a supervised manner to become capable of outcome prediction for new data [7]. The structure of MLP is shown in fig 1. An MLP consists of a set of interconnected artificial neurons connected only in a forward manner to form layers. One input, one or more hidden and one output layer are the layers forming an MLP[10]. Artificial neuron is basic processing element of a neural network. It receives signal from other neurons, multiplies each signal by the corresponding connection strength, that is weight, sums up the weighted signals and passes them through an activation function and feeds the output to other neurons[8].

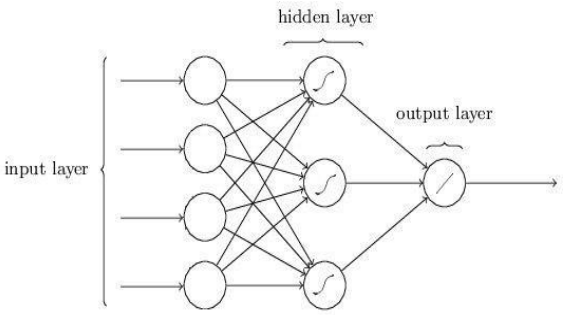

Fig.1 Structure of MLP

Neural classification of breast cancer data consists of two steps-training and testing. The classification accuracy depends on training[9]. A mapping between the input and output data could be established by assigning weights to the input numerical data during training[10]. The training requires a series of input and associated output vectors. During the training, the network is repeatedly presented with the training data and the weights and thresholds in the network are adjusted from time to time till the desired input output mapping occurs[8].

B.The Radial Basis Function Neural Network (RBFNN):

RBF is a different approach by viewing the design of a neural network as a curve-fitting problem in a high-dimensional space. According to this viewpoint, learning is equivalent to finding a surface in a multidimensional space that provides a best fit to the training data, with the criterion for "best fit" being measured in some statistical sense. [11]

RBF is a different approach by viewing the design of a neural network as a curve-fitting problem in a high-dimensional space. According to this viewpoint, learning is equivalent to finding a surface in a multidimensional space that provides a best fit to the training data, with the criterion for "best fit" being measured in some statistical sense.

The construction of a radial-basis function network in its most basic form involves three entirely different layers. The input layer is made up of source nodes. The second layer is a hidden layer of high enough dimension, which serves a different purpose from that in a multilayer perceptron. The output layer supplies the response of the network to the activation patterns applied to the input layer. The transformation from the input space to the hidden-unit space is nonlinear where as the transformation from the hidden-unit space to the output space is linear.

RBFNN is trained to perform a mapping from an m-dimensional input space to an n-dimensional output space.

An RBFNN consists of the m-dimensional input x being passed directly to a hidden layer. Suppose there are c neurons in the hidden layer. Each of the c neurons in the hidden layer applies an activation function, which is a function of the Euclidean distance between the input and an m-dimensional prototype vector. Each hidden neuron contains its own prototype vector as a parameter. The output of each hidden neuron is then weighted and passed to the output layer. The outputs of the network consist of sums of the weighted hidden layer neurons[12]. The transformation from the input space to the hidden-unit space is nonlinear where as the transformation from the hidden-unit space to the output space is linear[11].

RBFNN network performance depends on the number and location (in the input space) of the centers, the RBFNN functions shape at the hidden neurons, and the method used for determining the network weights. Some researchers have trained RBFNN networks by selecting the centers randomly from the training data[13].

C.Probabilistic neural network (PNN) :

Probabilistic neural network is a kind of radial basis network suitable for classification problems. PNN networks have three layers: input, pattern and summation. The input layer has as many elements as there are individual parameters needed to describe the samples to be classified. The pattern layer organizes the training set such a way that an individual processing element represents each input vector. The pattern units in the probabilistic network are used to store pattern examples, taken directly from the training data. The entire set of training data is used, and so the number of units in the first hidden layer is set equal to the number of training cases. The summation layer has as many processing elements as there are classes to be recognized and simply collects the outputs from all hidden neurons of each respective class. The products of the summation layer are forwarded to the output (one neuron for each data class), where the estimated probability of the new pattern being a member of that data class is computed. The transfer function is radial basis function for the first layer and is competitive function for the second layer. Only the first layer has biases. Training of the probabilistic neural network is much easier than with back-propagation. It can be simply finished by setting the weights of the network using the training set. The outputs of summary layer are binary values y i is larger than input of other neurons (which means that this input pattern belongs to class i), y i is set to 1, otherwise it is set to zero[15].

D. Generalized Regression Neural Networks (GRNNs):

The generalized regression neural networks (GRNNs) are the paradigms of radial basis function (RBF) networks, often used for function approximations. It's another term for Nadaraya-Watson kernel regression, and has the following form for the function mapping.

GRNNs share a special property, namely that they do not require iterative training; the hidden-to-output weights are just the target values t_k, so the output y(x), is simply a weighted average of the target values t_k of training cases x_k close to the given input case x. It can be viewed as a normalized RBF network in which there is a hidden unit centered at every training case.

is linear[11].

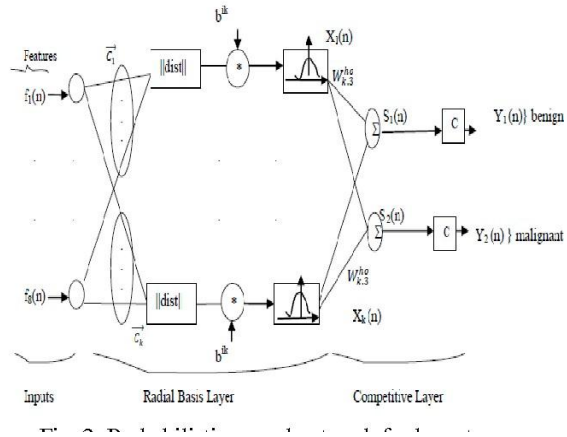

Fig 2. Probabilistic neural network for breast cancer diagnosis.

These RBF units are called "kernels" and are usually probability density functions such as the Gaussians considered in (3).

$$y(x) = \frac{\sum_k t_k \exp\left\{-\|x - x_k\|^2 / 2h^2\right\}}{\sum_k \exp\left\{-\|x - x_k\|^2 / 2h^2\right\}} \quad (3)$$

The only weights that need to be learned are the widths of the RBF units h. These widths (often a single width is used) are called "smoothing parameters" or "bandwidths" and are usually chosen by cross validation. GRNN is a universal approximator for smooth functions, so it should be able to solve any smooth function approximation problem given enough data. The main drawback of GRNNs is that, like kernel methods in general, they suffer seriously from the curse of dimensionality. GRNNs cannot ignore irrelevant inputs without major modifications to the basic algorithm[16].

E.Neuro-Fuzzy Systems:

Neural networks and fuzzy logic are two approaches that are widely used to solve classification and pattern recognition problems. The main advantage of neural networks is their learning capabilities and their ease of implementation. In the other hand , the non interpretability of their results is a major disadvantage (black boxes). The fuzzy inference systems can interpret their results using a knowledge base (rule base). The joint use of neural networks and fuzzy inference systems can exploit the advantages of both methods.

Table 1 below gives a comparative view between the two approaches:

Table 1: Comparison between neural networks and fuzzy inference systems [17]

Artificial neural network	**Fuzzy inference system**
Difficult to use prior rule knowledge	prior rule base can be incorporated
Based on Learning	Can not learn
Black box	Interpretable(If – then rules)
Complicated learning algorithms	Simple interpretation and implementation
Difficult to extract knowledge	Knowledge can be available

We can say that neuro-fuzzy systems are connectionist models that allow learning as artificial neural network , but their structure can be interpreted as a set of fuzzy rules .Fuzzy logic and neural networks form the basis of the majority aided diagnostic intelligent systems.It would be interesting to combine the two approaches to exploit both advantages.

F.Genetic algorithms

The idea of applying the biological principle of natural evolution to artificial systems, introduced more than four decades ago, has seen impressive growth in the past few years. Usually grouped under the term e6olutionary algorithms or e6olutionary computation, we find the domains of genetic algorithms, evolution strategies, evolutionary programming, and genetic programming [6,15,23]. Such algorithms are common nowadays, having been successfully applied to numerous problems from different domains, including optimization, automatic programming, machine learning, economics, medicine, ecology, population genetics, and hardware design. In this paper we consider the evolutionary methodology known as genetic algorithms. A genetic algorithm is an iterative procedure that involves a population of individuals, each one represented by a finite string of symbols, known as the genome, encoding a possible solution in a given problem space. This space, referred to as the search space, comprises all possible solutions to the problem at hand. Genetic algorithms are usually applied to spaces which are too large to be exhaustively searched. The symbol alphabet used is often binary, though this has been extended in recent years to include character-based encodings, real-valued encodings, tree representations, and other representations [23]. The standard genetic algorithm proceeds as follows: an initial population of individuals is generated at random or heuristically. Every evolutionary step, known as a generation, the individuals in the current population are decoded and evaluated according to some predefined quality criterion, referred to as the fitness, or fitness function. To form a new population (the next generation), individuals are selected according to their fitness. Many selection procedures are currently in use, one of the simplest being fitness-proportionate selection, where individuals are selected with a probability proportional to their relative fitness. This ensures that the expected number of times an individual is chosen is approximately proportional to its relative performance in the population. Thus, high-fitness (‘good’) individuals Stand a better chance of ‘reproducing’, while low-fitness ones are more likely to disappear. Selection alone cannot introduce any new individuals into the population, i.e. it cannot find new points in the search space. These are generated by genetically inspired operators, of which the most well known are crossover and mutation. Crossover is performed with probability pc (the ‘crossover probability’ or ‘crossover rate’) between two selected individuals, called parents, by exchanging parts of their genomes (i.e. encodings) to form two new individuals, called offspring. In its simplest form, substrings are exchanged after a randomly selected crossover point. This operator tends to enable the evolutionary process to move toward ‘promising’ regions of the search space. The mutation operator is introduced to prevent premature convergence to local optima by randomly sampling new points in the search space. It is carried out by flipping bits at random, with some (usually small) probability pm. Genetic algorithms are stochastic iterative processes that are not guaranteed to converge. The termination condition may be specified as some fixed, maximal number of generations or as the attainment of an acceptable fitness level.

III. DESCRIPTION OF WISCONSIN BREAST CANCER DATABASE (WBCD)

The database used for detection of breast cancer by artificial neural networks is publicly available in the Internet[21]. Wisconsin breast cancer database(WBCD) have been collected by Dr.William H. Wolberg(1989–1991) at the University of Wisconsin–Madison Hospitals.

There are 699 records in this database. Each record in the database has nine attributes. The nine attributes detailed in Table1 are graded on an interval scale from a normal state of 1–10, with 1being the most abnormal state. In this database, 241(65.5%) records are malignant and 458(34.5%) records are benign.(25). A simple example of how ANN is trained to predict the diagnostic outcome from six inputs and one hidden layer with 8 neurons is shown in Figure 3.

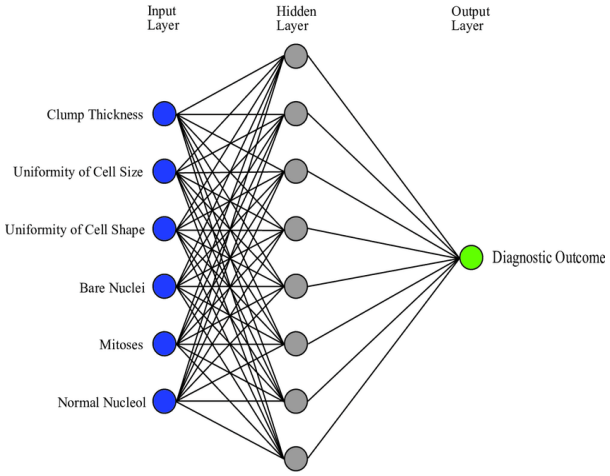

Fig .3 A simple example of how ANN is trained to predict the diagnostic outcome from six inputs and one hidden layer with 8 neurons.

Table 1 provides the attribute information.

Table 1:

	Min.	Max.	Mean	Std.
clump_thickness	1	10	4.44	2.82
cell_size_uniformity	1	10	3.15	3.07
cell_shape_uniformity	1	10	3.22	2.99
marginal_adhesion	1	10	2.83	2.86
single_epithelial_cell_size	1	10	3.23	2.22
bare_nuclei	1	10	3.54	3.64
bland_chromatin	1	10	3.45	2.45
normal_nucleoli	1	10	2.87	3.05
mitoses	1	10	1.6	1.73
diagnose	0	1	0.35	0.477

The original data can be presented in the form of analog values with values ranging from 0-10. Conversion of the given data sets into binary can be done based on certain ranges, which are defined for each attribute[22].

IV. COMPARISON TABLE

In [23] four different neural network structure, MLP,RBFNN,PNN,GRNN and Fuzzy were applied to WBCD to show the performance of statistical neural networks on breast cancer data. The spread value of RBF, PNN and GRNN was chosen 4.4, 1 and 3, respectively. In MLP, learning rate was 0.6. According to results, RBF and PNN gives the best classification accuracy with 342 correct classifications while GRNN has the lowest accuracy with 330 correct classifications for the training set. MLP has 335 correct classifications. Accuracy comparison for popular neural network techniques with WBCD data for the diagnosis of breast cancer is shown by table 2.

Table 2: Accuracy comparison for test data classification

Network Type	Accuracy	References
Radial Basis Function Neural Network (RBFNN)	96.18%	[23]
Probabilistic Neural Network (PNN)	97.0%	[23]
Multilayer Perceptorn (MLP)	95.74%	[23]
Generalized Regression Neural Network (GRNN)	98.8%	[23]
Fuzzy	96.71%	[24]

V. CONCLUSION

The last decade has witnessed major advancements in the methods of the diagnosis of breast cancer. It was found that the use of ANN increases the accuracy of most of the methods and reduces the need of the human expert.

In this paper, we have provided explanations of different ML approaches and their applications in BC diagnosis and prognosis used to analyse the data in the benchmark database WBCD. ML techniques have shown their remarkable ability to improve classification and prediction accuracy. Lots of algorithms have achieved very high accuracy in WBCD, the development of improved algorithms is still necessary. Classification accuracy is a very important assessment criteria but it is not the only one.

Different algorithms consider different aspects, and have different mechanisms. Although for several decades ANNs have dominated BC diagnosis and prognosis, it is clear that more recently alternative ML methods have been applied to intelligent healthcare systems to provide a variety of options to physicians.

VI. REFERENCES

[1].https://www.who.int/news-room/fact-sheets/detail/cancer
[2].https://www.bcrf.org/breast-cancer-statistics
[3]https://medicalxpress.com/news/2018- 02.breast-cancer-death-years.html
[4].https://www.cancer.org/latest-news/facts-and-figures-2018-rate-of-deaths-from-cancer-continues-decline.html

[5].D. B. Fogel , E. C. Wasson , E. M. Boughton, and V. W. Porto, "A step toward computer-assisted mammography using evolutionary programming and neural networks."
[6].https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types-of-breast-cancer.html
[7].S. S. Haykin, Neural Networks and Learning Machines, 3rd ed. New York: Prentice Hall, 2009, pp. 906.
[8] F. Paulin and A. Santhakumaran, "Classification of breast cancer by comparing back propagation training algorithms", Int. J. on Comput.Sci. and Eng. (IJCSE), vol. 3, no. 1, Jan. 2011.
[9].R. Nithya and B. Santhi, "Classification of Normal and Abnormal Patterns in Digital Mammograms for Diagnosis of Breast Cancer" Int.J. of Comput. Applicat., vol. 28, no.6, pp. 0975 – 8887, Aug. 2011.
[10]. A. S. Mojarad, S. S. Dlay, W. L. Woo, and G. V. Sherbet "Cross Validation Evaluation for Breast Cancer Prediction Using Multilayer Perceptron Neural Networks" American J. of Eng. and Applied Sci., vol. 5, no 1 pp 42-51, 2012.
[11].S. Haykin, Neural Networks: A Comp rehensive Foundation, Mac Millan College Publishing Company, 1994.
[12].M. R. Senapati, P. K .Routray, P. K. Dask "Breast Cancer Detection using Recursive Least Square and Modified Radial Basis Functional Neural Network", Int. Conf. [ICCT-2010], IJCCT vol. 2, no. 2, 3, 4, 3rd-5th December 2010.
[13].D. Broomhead and D.Lowe, "Multivariable Functional Interpolation
[14].http://www.ijtrd.com/papers/IJTRD8579.pdf
[15].http://www.cancerjournal.net/article.asp?issn=09731482;year=2013;volume=9;issue=3;spage=456;epage=466;aulast=Kalteh#top
[16].https://www.researchgate.net/publication/228759310_Breast_cancer_diagnosis_using_statistical_neural_networks
[17].Nadia Nedjah,Luiza de Macedo Mourelle . ''Fuzzy systems engineering: theory and practice" , éditeur :Springer, (2005).and adaptive networks", complex systems, vol. 2, pp. 321-355, 1998.
[18].Tawam's 2012 Breast Cancer Awareness Campaign[online].
Available: http://www.ameinfo.com/tawams-2012-breast-cancer-awareness-campaign-312992.
[19].Breast Cancer [online]. Available: http://www.womenshealth.gov/breast-cancer/what-is-breast-cancer.
[20]. W. Land, and E. Veheggen, "Experiments using an evolutionary programmed neural network with adaptive boosting for computer aided Diagnosis of breast cancer", in Proc. IEEE Int. Workshop on Soft Computing in Ind. Application, 2003 (SMCia-2003), Finland,
23rd-25th June, pp. 167-172.
[21].Breast Cancer Wisconsin Data [online]. Available:http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data.
[22].Dr. K. U. Rani, "Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique" Int. J. of Comput. Applicat., vol.10, no. 3, pp. 0975 – 8887, Nov. 2010
[23].T. Kiyan, T Yildirim. "Breast cancer diagnosis using statistical neural networks", J. of Elect. & Electron. Eng., vol .4-2, 2004, pp. 1149- 1153.
[24].P. Meesad and G. Yen., "Combined numerical and linguistic knowledge representation and Its application to medical diagnosis", IEEE transactions on Syst., Man and Cybernetics, Part A: Syst. and Humans, vol. 3, no.2, pp. 206-222.
[25].https://www.researchgate.net/publication/325732902_Breast_cancer_tumor_type_recognition_using_graph_feature_selection_technique_and_radial_basis_function_neural_network_with_optimal_structure.