IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0903116|

An Accurate Breast Cancer Detection and Classification using Image Processing

R.Jeeva, S.Dhanasekar, A.Harshathunnisa, V.Eshwin, Amit Karn

Department of ECE, Sri Eshwar College of Engineering, Coimbatore, India

ABSTRACT: In this research article, accurate breast cancer detection and classification are analyzed using image processing. An image processing includes pre-processing of the image which is used to remove the redundancy present in captured images without affecting the input images. We have proposed a Support Vector Machine (SVM) algorithm for Breast cancer detection. It constructs a group of hyper planes during a high- or infinite-dimensional space, which may be used for classification, regression, or other tasks like outliers detection. SVM gives 85% of accuracy in Breast cancer detection whereas ANN gives 70%. Discrete Wavelet Transform (DWT) coefficients has feature vector and it may be a powerful mathematical tool for feature extraction. Machine learning becomes the solution to all or any chaos by proposing clever alternatives to analysing huge volumes of knowledge. Machine learning can produce accurate results and analysis by developing fast and efficient algorithms and data-driven models for real-time processing of data.

KEYWORDS*:* DWT, Image processing, Gray scale image, SVM

I. INTRODUCTION

Breast cancer is a serious threat to women's life and health, and the morbidity and mortality of breast cancer are ranked second out of all female diseases. Early detection of lumps can effectively reduce the mortality rate of breast cancer It is not possible for the doctor to be available with the patient and monitor them. Hence this system helps to monitor the patient continuously. We have used DDSM Database (Digital Database for Screening Mammography) for the images which contain approximately 2700 cases and are being used worldwide for cancer research. The mammogram is widely used in the early screening of breast cancer due to its relatively low expense and high sensitivity to minor lesions. In the actual diagnosis process, however, the accuracy can be negatively affected by many factors, such as radiologist fatigue and distraction, the complexity of the breast structure, and the subtle characteristics of the early-stage disease. Early detection of lumps can effectively reduce the mortality rate of breast cancer.

For an early examination of the tissues, the affected areas have discernible differences from a benign tumour. Such differences include texture and morphological abnormalities in specific clusters thus allowing us to identify, analyse, discriminate, and extract the malignant region. Improvements in the medicinal field and more accurately the association of computing offers a novel diagnosis tool known as Medical Image Processing, which has beneficial not only for cancer detection, also it finds the applications for the detection of several types of diseases and also one of the improvements in the classification of breast cancer subtype by miRNA/mRNA expression modules [1 -2].

Image Processing includes various algorithms and extracts the Region of Interest (ROI) to detect the accuracy and the drawback of using those methods are linear in nature [3]. A feature extraction method was introduced for mammograms and their class. SVM (Support Vector Machine) and the seven features of GLCM were offered for classification. GLCM's classifications named Wavelet-CT1, Wavelet-CT2, and ST-GLCM. The specificity of GLCM is 96.88%, sensitivity is 98.43% and accuracy is 97.91% [4].

Ultrasound (US) imaging plays a vital role within the diagnosis of liver, neck, heart, abdomen, chest prostate, and kidney to its non-invasiveness, secure and practicability in real-time. In spite of its advantages, the main issue in US image diagnosis is that the existence of speckle noise. Diagnosis isn't a simple task for physician. The speckle noise has got to be far away from the US images for better analysis and this noise removal process is mentioned as a despeckling technique [5]. Wavelet transform with appropriate thresholding removes noise in signal and images. The main benefit of wavelet transform is to separate process into the low and high-frequency components of an image [6]. The reconstruction of the image and the redundancy of noise are done better by compressive sensing and its performance is ranked both computably and conditionally [7 -8].

Image processing techniques are mainly utilized in the medicinal sector and they will provide additional information by eliminating the issues like noise, resolution, and appearances of image. To classify the features of ROI regions as positive or negative, the extracted features are normally fed into a classifier. The image segmentation process has simplest representation in varied regions of the breast, background, artifacts, and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0903116|

labels are often chosen by a person's. The image processing also bears some fuzziness in nature which has two parameters- ambiguity and vagueness within the interpretation of the low-level image processing. Digitization has almost been engaged in all disciplines of engineering out of which creates an impact on image processing which is used for locating and detecting many diseases more effectively [9 – 10]. Based on tumor detection and the classification of tumor from the medicinal images has been studied [11 -14]. With the tremendous development of wideband wireless communication, image capture, and transmission has been widely developed for many applications [15 – 16].

This paper methodized the Artificial Neural Network in Section II. Section III discusses a proposed method of processing of images. The results of Breast cancer detection and classification are discussed in Section IV. Finally, a conclusion is presented in Section V.

II. ARTIFICIAL NEURAL NETWORK

An ANN (Artificial Neural Network) is based on a collection of connected units or nodes called artificial neurons that loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain will transmit a signal to other neurons. An artificial neuron that receives a signal processes it and can make the signal neurons connected to it. The signal which is at a connection is called a real number and the output of each neuron is computed by some non-linear function of the sum of its given inputs. The connections are called edges. Neurons and edges basically have a weight that adjusts as learning proceeds further.

The strength of the signal at a connection gets increased or decreased by the weight. Neurons may have a threshold and considering that threshold, a signal will be sent only if the aggregate signal crosses that threshold level. Usually, neurons are aggregated into layers. Different transformations on their inputs will be performed by Different layers. Signals travel from the first layer to the last layer and possibly after traversing the layers multiple times.
The basic structure of Feed-Forward and Feed-Back mechanism of ANN is shown in Fig 1.1

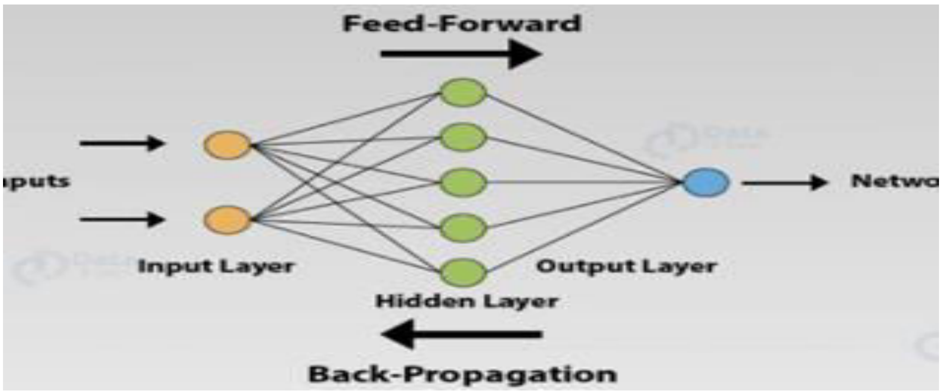


Fig1.1 Structure of Artificial Neural Network model

The Artificial Neural Network consists of artificial neurons which are conceptually derived from biological neurons. Each artificial neuron has inputs that will produce a single output which can be sent to multiple other neurons. The inputs will become the feature values of a sample of external data, such as images or documents, or they will become the outputs of other neurons. The output of the final output neurons of the neural net recognises an object in an image.

To find the output of the neuron, first we have to take the weighted sum of all the inputs which is weighted by the weights of the connections from the inputs to the neuron. We add a bias term to the sum. This weighted sum is called activation. This weighted sum is then passed through an activation function to produce the output. The initial inputs are external data which is nothing but images and documents. The ultimate output recognises an object in an image. The network consists of connections and these connections provide the output of one neuron as an input to another neuron. Each connection is assigned a weight that will represent its relative importance. Multiple input and output connections are there in neurons.

The propagation function computes the given input from the output of its predecessor neurons and also their connections as a weighted sum. A bias term can be added to the result of the propagation. The neurons are organized into multiple layers in deep learning. The neurons of one layer connect only to the neurons of the immediately

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 3, March 2021 ||
| DOI: 10.15680/IJIRCCE.2021. 0903116|

preceding and following layers. The input layer is the layer that receives the external data and the output layer produces the ultimate result. In between them, there are zero or more hidden layers. Single layer and unlayered networks are also used. Between two layers, multiple connection patterns are possible.

They will be fully connected with every neuron in one layer connecting to every neuron in the next layer. When a group of neurons in one layer connect to a single neuron in the next layer, they will be pooling, thereby reducing the number of neurons in that layer. Such connection forms a directed acyclic graph and they are called as feed forward networks. If the network allows connection between neurons in the same or previous layers, then they are known as recurrent networks.

Before the learning process begins, the value of hyper parameter which is a constant parameter is set. The values of the parameters are derived via learning. Examples of such hyper parameters include learning rate, the number of hidden layers and the batch size. The value of some hyper parameters can be dependent on those of other hyper parameters. For example, the size of some layers depends on the overall number of layers

Some of their drawbacks include difficulty to get accurate results. Then this method is not applicable for multiple images for Breast cancer detection in a short time. Serious inaccuracies classification can be led by Medical Resonance images that contain a noise caused by operator performance.

2.1 Support Vector Machine

The common task in machine learning is data classification. Suppose some given data points belong to one of the two classes, and their goal is to decide in which class, the new data point will be. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, outliers' detection. Support Vector Matrix Representation is shown in Fig 2.1

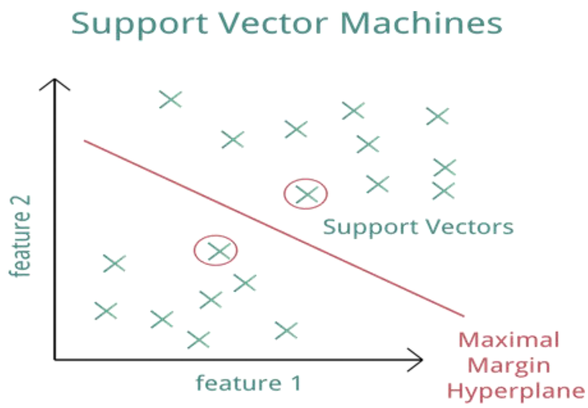


Fig 2.1: Support Vector Matrix Representation

The hyper-plane helps us to achieve a good separation which has the largest distance to the nearest training-data point of any class. Because in general, the larger the margin the lower the generalisation error of the classifier will become .The original problem may be stated in a finite dimensional space and it often happens that the sets to discriminate are not much linearly separable in that space. Due to this reason, the original finite-dimensional space should be mapped into a much higher-dimensional space to make the separation easier in that space was proposed.

To make reasonable computational load, SVM schemes uses mappings that are designed in such a way that ensures that the dot products of pairs input data vectors can be computed easily in terms of the variables in the original space and they are doing so by defining them in terms of a kernel function. We can define the hyper planes in the higher-dimensional space as the set of points whose dot product with a vector in that space is constant.

The vectors defining the hyper planes can be chosen as linear combinations with parameters .The fact that the set of points mapped into any hyper plane can be quite convoluted as a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 3, March 2021 ||
| DOI: 10.15680/IJIRCCE.2021. 0903116|

III. PROPOSED METHOD

Our proposed method uses the Discrete Wavelet Transform (DWT) coefficients as feature vector. The wavelet is a powerful mathematical tool for feature extraction, and has been used to extract the wavelet coefficient from medical images.

3.1 Block Diagram

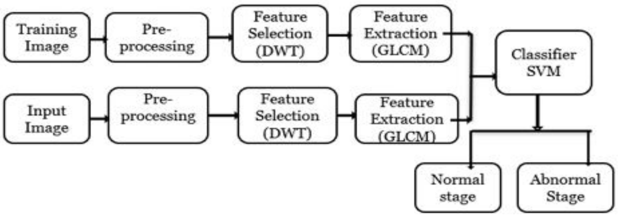


Fig 2.1: Block Diagram of Breast Cancer Detection.

The main advantage of wavelets is that they provide localised frequency information about a function of a signal, is particularly beneficial for classification. The processing stages for breast cancer detection is shown in Fig.2.1

3.2 Pre-Processing:

The aim of pre-processing the images is to selectively remove the redundancy present in the captured images so that it won't affect the details that will play a key role in the overall process. The input is a converter for Gray scale images.

3.3 Working Method:

Initially this method reads the input image that is provided then it is converted to grey scale image that is a combination of grey and white. After converting it to grey and white scale image it is then re-sized to improve the pixel. Then finally the noise present in the image is removed to get accurate results.

3.3.1 Read Images:

In this step, we store the path to our image dataset into a variable then we create a function to load folders containing images into Pixel value. Read Images and Converter for Gray scale Image

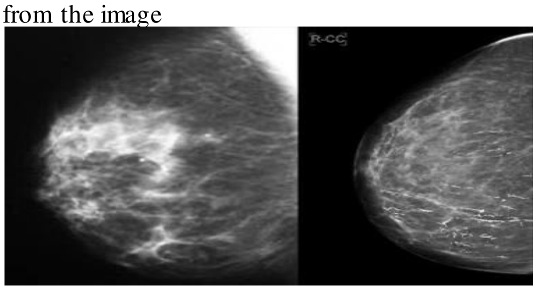
3.3.2 Resizing Images:

Thus, a resized image contains more or less pixels than that of the original image. If the resolution of the image gets increased, the intensity values of additional pixels are obtained through interpolation.

3.3.3 Remove noise(Denoise):

There are some image smoothing techniques like Gaussian Blurring, Median Blurring etc. They are basically used to remove small quantities of noise. In those technics, we took a small neighbourhood around a pixel and did some of its operations like Gaussian weighted average, median of the values etc., to replace the central element. Removing noise at a pixel was local to its neighbourhood.

There is a property of noise. Random variables with zero mean are generally considered as noise. Consider a noisy pixel $p = p_0 + n$ where p0 is the true value of the pixel and n is the noise in the pixel. You can take a large number of small pixels (say N) from different images and compute their average. . Fig 4.2 shows how the image will be changed after the noise is removed from the image



International Journal of Innovative Research in Computer and Communication Engineering

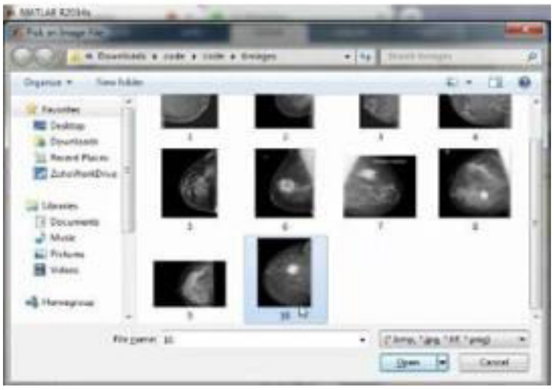
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

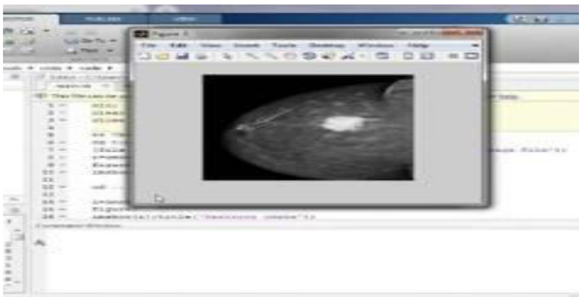
| DOI: 10.15680/IJIRCCE2021. 0903116|

IV. RESULTS & DISCUSSION

The proposed system uses the Discrete Wavelet Transform (DWT) coefficients as feature vectors. The wavelet is a powerful mathematical tool for feature extraction, and has been used to extract the wavelet coefficient from medical images. The test image is compared with the DB image and separated using the SVM mechanism.



Input Image is selected From the List Fig 4.1



Input Image Selected from the List Fig 4.

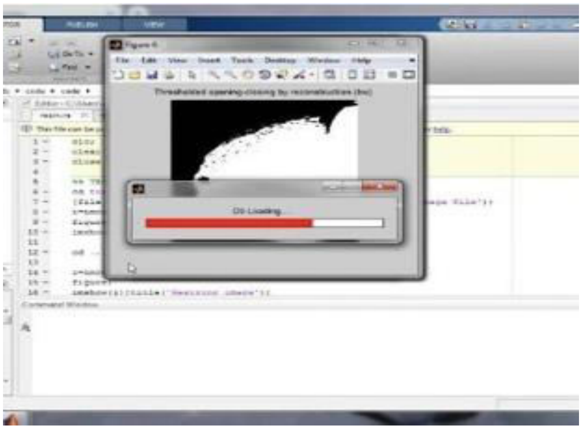


Fig 4.3 The Image is compared with the images in the DB

Fig 4.3.1: Comparison with DB

Fig 4.4 The Result is Displayed after the Comparison Process

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE2021. 0903116|

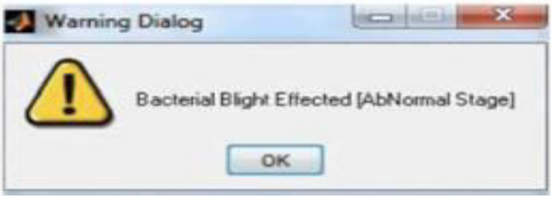


Fig : Result of the in[put image whether Normal/Abnormal

Accuracy, Sensitivity, Specificity of the Input image details

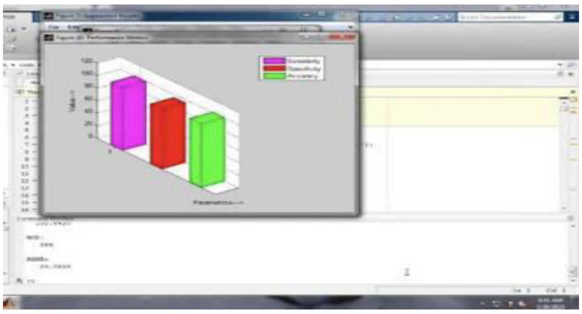


Fig: Input image Properties

1. **Preprocessing –** This process converts to a specified format for better classification. It changes the size, resolution and contrast of all the training and input images without affecting the key details and helps in arranging all the images with the same dimensions.

2. Feature Selection(DWT) - An numerical analysis and functional analysis, a Discrete Wavelet Transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. This method helps in identifying different portions of an image for feature selection.

3. Feature Extraction - This process extracts the pixel for feature analysis using gray scaling. The Image is converted into a gray image whose pixel parameters Range from (0-255). Depending on these values a pixel matrix is generated.

4. Classifier(SVM) - On completion Of the above processes, the image is then passed through the SVM to identify various differences and Segregate into mainly two parts, The first with the ones, affected with breast cancer and the other with the normal ones
Hence, these classifications will help in early diagnosis of breast cancer when the size of the tumour is very small and can still be unidentified during diagnosis and this classification process will act as an extra help for the doctors.

Table I: Performance comparison of ANN and SVM algorithm

Parameters	SVM [12]			ANN [12]		
	WBCD	BUPA	JNCI Data	WBCD	BUPA	JNCI Data
Accuracy	99.51%	63.11%	78.35%	98.54%	57.28%	82.4%
Sensitivity	99.25%	36.67%	100%	99.25%	75%	100%

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 3, March 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0903116|

Specificity	**100%**	**100%**	**0%**	**97.22%**	**32.56%**	**19.05%**
AUC	**99.63%**	**68.34%**	**50%**	**98.24%**	**53.78%**	**59.52%**

Thus the Table I specifies that the proposed algorithm SVM gives more accuracy, sensitivity, specificity and AUC than the ANN algorithm which is compared with four datasets.

V. CONCLUSION

In this paper, Breast cancer is detected at an earlier stage with accuracy using image processing and features extraction techniques that assist radiologists in detecting the tumour. The features extracted from suspicious regions in mammography images helps the doctors to discover the existence of the tumour at real time thus speeding up the treatment process because accurate detection of size and location of Breast cancer plays a vital role in the diagnosis of tumour.

REFERENCES

1. T. C. Cahoon, M. A. Sutton and J. C. Bezdek, "Breast cancer detection using image processing techniques," Ninth IEEE International Conference on Fuzzy Systems. FUZZ- IEEE 2000 (Cat. No.00CH37063), San Antonio, TX, USA, 2000, pp. 973-976 vol.2, doi: 10.1109/FUZZY.2000.839171.
2. C. Varma and O. Sawant, "An Alternative Approach to Detect Breast Cancer Using Digital Image Processing Techniques," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018, pp. 0134-0137, doi: 10.1109/ICCSP.2018.8524576.
3. S. Gupta, N. Sinha, R. Sudha and C. Babu, "Breast Cancer Detection Using Image Processing Techniques," 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 2019, pp. 1-6, doi: 10.1109/i-PACT44901.2019.8960233.
4. P. Kathale and S. Thorat, "Breast Cancer Detection and Classification," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.367.
5. L. Jubairahmed, S. Satheeskumaran, and C. Venkatesan, “Contourlet transform based adaptive nonlinear diffusion filtering for speckle noise removal in ultrasound images,” Cluster Computing, vol. 22, no. S5, pp. 11237–11246, Nov. 2017.
6. L. J. Ahmed, “Discrete Shearlet Transform Based Speckle Noise Removal in Ultrasound Images,” National Academy Science Letters, vol. 41, no. 2, pp. 91–95, Apr. 2018.
7. G. Rajesh Hien Dang, K. Martin Sagayam, S. Dhanasekar,"Image Fusion based on Sparse Sampling Method and Hybrid Discrete Cosine Transformation",International Journal of Scientific and Technology Research,vol.8,no.12,pp.1103-1107,December 2019
8. J. Anitha S. Dhanasekar, P. Malin Bruntha, S.Immanuel Alex Pandian,"Local Ternary Co-occurrence Patterns based Lung Nodules Detection",2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS),pp.489-492
9. Dhanasekar S, Suriavel Rao R S and Victor Du John H, “A Fast and Compact multiplier for Digital Signal Processors in sensor driven smart vehicles”, International Journal of Mechanical Engineering and Technology, vol.9, 2018, pp. 157–167.
10. Dhanasekar S, Malin Bruntha P,Martin Sagayam K, "An Improved Area Efficient 16-QAM Transceiver Design using Vedic Multiplier for Wireless Applications", International Journal of Recent Technology and Engineering, Vol.8, No. 3, (2019), 4419-4425, doi: 10.35940/ijrte.C5535.098319
11. Y. Lu, J. Li, Y. Su and A. Liu, "A Review of Breast Cancer Detection in Medical Images," 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 2018, pp. 1-4, doi: 10.1109/VCIP.2018.8698732.
12. Tobias Christian Cahoon, Melanie A Sutton, James C BezdekNinth IEEE International Conference on Fuzzy Systems. FUZZ-IEEE 2000 (Cat. No. 00CH37063) 2, 973-976, 2000
13. Xueqin Li, Zhiwei Zhao, HD ChengInformation Sciences-Applications 4 (1), 49-56, 1995.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0903116|

14. Sy Ahmad Ubaidillah, S. H., Sallehuddin, R., & Ali, N. A. (2013). *Cancer Detection Using Aritifical Neural Network and Support Vector Machine: A Comparative Study. Jurnal Teknologi, 65(1).* doi:10.11113/jt.v65.1788.
15. Dhanasekar S, Ramesh J, "FPGA Implementation of Variable Bit Rate 16 QAM Transceiver System", International Journal of Applied Engineering Research, vol.10, 2015 pp.26479-26507.
16. Dhanasekar S, Ramesh J, "FPGA Implementation of Variable Bit Rate OFDM Transceiver System for Wireless Applications", Proceedings of IEEE International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology, (2017), 343-346, doi: 10.1109/ICIEEIMT.2017.8116863

4.3 Accuracy Analysis

The accuracy of SVM on the breast cancer dataset is evaluated using the python sklearn metric library which is used to evaluate accuracy of classification algorithms. Table 2 and Figure 3 show the accuracy of the SVM and ANN model.

Table 2 Accuracy of SVM & ANN breast cancer (WDBC) data set

Model	Accuracy %
SVM	98 %
ANN	99%

4.3 Accuracy Analysis

The accuracy of SVM on the breast cancer dataset is evaluated using the python sklearn metric library which is used to evaluate accuracy of classification algorithms. Table 2 and Figure 3 show the accuracy of the SVM and ANN model.

Table 2 Accuracy of SVM & ANN breast cancer (WDBC) data set

Model	Accuracy %
SVM	98 %
ANN	99%