International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

A Survey on Soft Computing Techniques for Early Detection of Breast Cancer

Ashwini K C[1], Jenifer Jacob[2], Sanjana Srinath[3], Vignesh Sharma S[4]

[1,2,3,4] *Students, Dept. of Computer Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India*

Abstract – *Breast cancer is most common among women and is said to be the second major cause of death among women. For every 19 seconds, somewhere around the world a case of breast cancer is diagnosed among women. Report says that for 74 seconds somewhere in the world a women dies from breast cancer. Most effective way to reduce the death rate is to detect at an early stage. By detecting at an early stage proper treatment can be given to save the life of patients. Accurate classification plays an important role in medical diagnosis. Soft computing approaches are gaining importance because of their classification performance in diagnosing the disease. The goal of this survey paper is to identify the current state of research in breast cancer and to summarize the different soft computing techniques that helps in identification and classification.*

Key Words: **KNN, SVM, Fuzzy C-means, ANN, GLCM, ROI**

1. INTRODUCTION

Breast cancer starts when the cells in the breast begin to grow out of control and are said to form a tumour. These tumour can be classified into cancerous or non-cancerous. According to the worldwide survey that was conducted in the year 2010 it is estimated that more than 1.5 million breast cancer cases occurred in women and among the 23% of breast cancer detected 14% of death is reported [1]. There are 12% of chances that a women might develop breast cancer during her life time. Regardless of age and their family history every women is at a risk of developing breast cancer. Early detection and effective treatment is the only way to reduce the death rate due to breast cancer. In this paper various machine learning algorithms and image processing techniques that are available for detecting and classifying the tumour cells at an early stage has been discussed in detail. Different combinations of these will give different outputs and varying accuracy. Thus making it easier to select the most accurate algorithms.

2. LITERATURE SURVEY

A novel methodology proposed in the paper [2] to detect breast asymmetry and calcification cancer cells using combination of different highly efficient technology of digital image processing which are not yet implemented. In this paper it is basically noted that breast asymmetry is one of the major method to identify the suspicious region in the breast and in the segmentation process, the Otsu's thresholding algorithm was being applied using its methodologies, which will segment the micro calcifications from the image with the highest accuracy. They have used the K-Nearest neighbor clustering algorithm which will group the identified micro calcifications into clusters. The next stage is feature extractions which includes brightness, contrast, size, shape and textures which can be obtained from previous micro calcification clusters. The features that were obtained were then extracted using the Sobel algorithm and the extracted features were then classified using the Bayes classifier. Thus the major intention of this paper is to identify the breast asymmetry in the earliest stage as possible with highest possible accuracy.

The major intention of this paper [3] is to classify the medical data with efficient and more accurate processing with simple and faster classification algorithms. The proposed model in this paper involves the steps where in initially the breast cancer dataset was taken as input and if at all few values were missing in the same then those values were handled in the next coming step. Using the K-means algorithm the clustering of the data set was performed as it is well known for its simplicity and hidden pattern and data recognition. Once the clustering was done, the clusters were classified as cluster1 and cluster2 where in each of it is carried out with the process of feature reduction using the FRFS(Fussy Rough Feature Selection) which is efficient in handling noisy, discrete and continuous data with no loss. The clustered data was then merged, reduced and classified using the D-KNN algorithm (Discernibility K-nearest neighbor) classifier which is known for its high accuracy and better classification of the data set. At the end, after all these classifications the overall performance of the proposed system was evaluated.

This paper [4] makes use of K-means and Fuzzy C-means algorithm for detecting the cancer tumour mass and micro calcification. K-means clustering algorithm is used to identify the hidden parts and Fuzzy C-Mean algorithm is mainly used in pattern recognition and it allows one piece of data to be present in two or more cluster. Gray level transformation used in this paper makes use of logarithmic and power law as a contrast stretching methods which is used to highlight the detail in dark or washed out images. Gray level Transformation technique has been used in this paper to obtain negative of an image with gray levels in the range 0 to

L-1 and this type of processing is well suited for enhancing the white or gray detail embedded in the dark region of an image especially when the dark areas are dominant in size. Log transformation is used in the paper to map a narrow range of gray level values in the input image into a wider range of output levels and its transformation. This paper also uses power law, one of the image enhancement technique which is useful for general purpose contrast manipulation and thresholding is used to produce an image with higher contrast than the original image by darkening the levels below and brightening the levels above a threshold value. Robert edge detection, Prewitt edge detection, Sobel edge detection are some of the different edge detection techniques used in this paper for identifying the edges from the original image. Thus, the major intention of this paper was to identify the presence of breast cancer mass and calcification in mammograms using image processing functions, K-means and Fuzzy C-means clustering for clear identification of clusters.

The technique involved in this paper [5] is the concept of fractal dimension measurement along with the ultrasound medicine characteristic and also grouping of the similar objects together in order to obtain knowledge from those data. This is done using the K-mean algorithm which provides the capability to discover new information using the existing data. The Box Counting method is applied to establish fractal dimensions for which a software called as HarFA was built. To calculate fractal dimension on an ultrasound of breast filtering, 4 types of filters were used which are: Max filter- this is known to process the low level image and vision. Min filter- This plays a major part in image processing and vision. Geometric filter- This replaces the gray level by taking the surrounding detail and noise. Median filter- These contain useful algebraic properties which are easier to understand with Fourier transform and other statistical properties. While analysing the tumours, the ROI (Region of Interest) of an ultrasound was considered and then the above filters were used to obtain the quality image. The K-mean algorithm is used to group all these tumours which are of same dimensions using the ultrasound image. The advantage of fractal analysis and clustering is majorly quoted in this paper.

In the referred paper [6], the proposed system for classification of mammogram image is based on GLCM (Gray Level Co-occurrence Matrix). The texture features are extracted using GLCM of ROI. The system is divided into five stages to classify mammogram images. First is Dataset collection where in the mammogram images are obtained from MIAS dataset which contains normal and abnormal images. Second stage is ROI extraction process, the original mammogram images have different types of noise and artifacts in background, so these unwanted elements should be removed from the image. Third stage is Pre-processing, which is further divided into two more steps namely filtering where median filter is used for removing noise and second step is Enhancement where CLAHE (Contrast Limited Adaptive Histogram Equalization) is used to improve the appearance of image. In the fourth step we can see Feature extraction from GLCM, we know that processing of large data is time consuming and less effective in Digital Image Processing. So for reducing time, the input data is transformed into reduced set if feature vector which has relevant information. This transformation process is called Feature extraction. Last and the final stage is classification of mammogram image into normal or abnormal images. Here in this paper classification is done with the help of KNN algorithm and compared with SVM and ANN. KNN is a supervised learning algorithm where different k values gives different results. In this paper they have considered 1NN and 3NN. The accuracy rate for normal and abnormal classification as given by 3NN is 96% when compared with other classifiers.

This paper [7] makes use of KNN algorithm which is a pattern recognition technique, which is used for Classification and Regression. In both the cases, the input will contain K closest training examples in the feature space. The output that is obtained from this method is dependent on whether KNN is used for Classification or Regression. The result of this modelling contains accuracy of 0.99. Support-Vector machines are supervised learning models that are associated with the learning algorithm, where the data is used for Classification and Regression analysis. The clustering algorithm that provides advancement to the support vector machines is called support vector clustering. The result of this modelling contains accuracy of 1. Logistic Regression is a predictive analysis technique. This model is used to describe the data and to explain the relation between one dependent binary variable and one or more nominal. The result of this modelling contains accuracy of 0.98.Decision tree learning is a method commonly used in data mining. The objective of this classification is to develop a model that predicts the value of a target variable based on several input variables. The result of this modelling contains accuracy of 0.98. Artificial Neural Network is a mathematical model that defines a function f: X →Y. For ANN algorithm, working with activation and learning rate parameters are carried out. The result of this model contains an accuracy of 0.98.Greedy Search algorithm uses a heuristic for making optimal choices at each stage to find a global optimum. According to the results of all these modelling the algorithms SVM and KNN are the best for the breast cancer prediction.

In this paper [8] they have discussed about different data mining modelling techniques and algorithms like Neural Network, Decision tree, Support vector machines and Bayesian network. Decision trees are easy to understand and interpret. A decision tree consists of root, internal and leaf nodes. With the help of a decision tree unknown data records can be classified. Decision tree can be applied in medical field for predicting the death rate, feature selection

to improve the classification accuracy etc., decision tree is an effective means of constructing a model to predict the risk of mortality. Neural network has a capability to learn a set of data and construct weight matrixes to represent the learning patterns. Neural network is a computational representation that takes a sequence of numbers input as and outputs another sequence of numbers. These computational nodes are connected in several layers to achieve high accuracy. ANN (Artificial Neural network) can be used as a tool to make decision in diagnosing the various diseases. From this paper we get to know that RBF neural network has resulted in 97% succession rate for classifying the dataset into benign and malignant. Support vector machine is a classification algorithm and a powerful pattern recognizer. The focus is on finding the hyper plane that separates them into cancerous and non-cancerous cells. SVM and ANN for prediction and detection of breast cancer is highly accurate than by the humans. The efficiency of SVM is nearly 97% and the efficiency of manual detection is around 85%. Bayesian belief network is well suited for dealing with the incomplete data. It can be used as both predictive and descriptive models. In prediction we try to infer tasks such as posterior probability, diagnostic reasoning, relevance analysis and classification. As a descriptive tool they represent the dependence/independence relationships among the variables. These are some of the data mining model that can be used for detecting and classifying the tumour cells.

The major intention of this paper [9] is to detect the tumour in breast from the mammogram images using the various image processing techniques. In the proposed system, initially screening and diagnostic techniques are used to get image of the breast by the digital mammogram. The proposed detection algorithm is then initiated which involves image segmentation, image binarization, image thinning, image triangulation and Euclidean distance transformation. In the image segmentation process, the image can be represented in simpler and meaningful way for which the variance method is used to perform thresholding. Binarization does the conversion of image into binary form which is required to know the threshold value. In this paper, threshold equation is being used to calculate foreground and background information of the binary image. Thinning is the next process which is basically used to get a preferred image pattern to focus on the region of interest. This is done using the parallel thinning algorithm which has two iterations to carry out the same. Next, is the triangulation process which is the post processing step that is used to clearly identify the edge of tumour region. A new approach is being used called the algorithm of Delaunay Triangulation which uses both iterative and non-iterative methodologies. EDT is the last step which usually has a lengthy calculation for which this paper suggests a scan recursive algorithm that only requires two scans of images whose coordinates are analyzed. These processes overall will yield the result where in which the features obtained are used to find the cancer cell area. The paper has estimated its overall performance including the accuracy, sensitivity etc.

3. CONCLUSION

In this survey paper we have discussed the various soft computing techniques that are available for detecting and classifying the tumour cell at an early stage. Choosing the best algorithm for detecting the disease plays a major role. Combination of algorithms can be used for disease detection and classification. Different combination of machine learning algorithms will give a different output and the accuracy might be varied. This paper contains various algorithms that can be used in the medical field for breast cancer detection and will be helpful in making the right decision in selecting the algorithms.

REFERENCES

[1] Chandresh Arya, Ritu Tiwari "Expert System for Breast Cancer Diagnosis: A Survey" 2016 International Conference on Computer Communication and Informatics.

[2] Sangeetha R Murthy K and Dr.Srikanth "A novel approach for Detection of Breast Cancer at an early stage by identification of breast Asymmetry and Micro Calcification Cancer cells using Digital Image Processing Technology" 2017 2nd International Conference for Convergence in Technology.

[3] Ibrahim M. EL-Hasnony, Hazem M. EL-Bakry, and Ahmed A. Saleh, "Classification of Breast Cancer using Softcomputing Techniques" International Journal of Electronics and Information Engineering.

[4] Nalini Singh, Ambarish G Mohapatra, Biranchi Narayan Rath, and Guru Kalyan Kanungo "GUI Based Automatic Breast Cancer Mass and Calcification Detection in Mammogram Images Using K- means and Fuzzy C-means Methods" International Journal Of machine Learning and Computing.

[5] Simona Moldovanu, Luminita Moraru," Mass Detection and Classification in Breast Ultrasound Image Using K-means Clustering Algorithm", IEEE Paper.

[6] Ranjit Biswas, Abhijit Nath, Sudipta Roy, "Mammogram Classification using Gray-Level Co-occurrence Matrix for Diagnosis of Breast Cancer", 2016 International Conference on Micro-Electronics and Telecommunication Engineering.

[7] Meraryslan Meraliyev, Meirambek Zhaparov, Kamalkhan Artykbayev "Choosing Best Machine Learning Algorithm for

Breast Cancer Prediction", International Journal of Advances in Science Engineering and Technology.

[8] S. Kharya, D. Dubey, S. Soni," Predictive Machine learning Techniques for breast Cancer Detection", International Journal of Computer science and Information Technologies.

[9] Angayarkanni.N, Kumar.D and Arunachalam.G,"The Application of Image Processing Techniques for Detection and Classification of Cancerous Tissue in Digital Mammograms", Journal of Pharmaceutical Sciences and Research.