ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

CANCER DETECTION METHODOLOGY USING FUZZY BASED CLASSIFICATION TECHNIQUES

A.Sakthivel [*1], Dr.A.Nagarajan [2]
[*1] M.Phil (SSP) Research Scholar, [2] Assistant Professor,
[*1, 2] Department of Computer Applications,
[*1, 2] Alagappa University- Karaikudi.

DOI: 10.5281/zenodo.1207033

ABSTRACT

Extracting knowledge and patterns for the diagnosis and treatment of disease from the medical database becomes more important to promote the development of telemedicine and community medicine. Data mining is a process of extracting hidden knowledge from large volumes of data. It is used intensively in the field of medicine to predict diseases such as heart diseases, lung cancer, breast cancer and more. Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain; such patterns are utilized for medical diagnosis. Medical images play an important role in assisting diagnosis and treatment of healthcare management systems. The advancements and large volumes of medical image data become major challenges. In this paper, a novel fuzzy-based classification method is performed to select the features of the cancer images. This research work mainly focuses on selecting the prominent features to improve the accuracy of the classification algorithms.

KEYWORDS: Cancer Images, Feature selection, Classification, Genetic Algorithm, Support Vector Machine.

I. INTRODUCTION

In generally the data mining techniques are widely used in the detection of the disease occurrence. It is a very popular research tool for medical researchers to predict the outcome of a disease using the datasets. There have been a large number of data mining algorithms rooted in these fields to perform different data analysis tasks. Feature selection is a promising application that is used to increasing the speed of classification technique. The main goal of feature selection is to find a feature subset that produces higher classification accuracy. Feature selection in medical prediction plays a major role as it identifies a prominent feature that influences prediction.

Cancer is a disease of the human cells. In commonly, cells grow and divide in an ordered way. Occasionally, some cells reproduce themselves in an uncontrolled way and these abnormal cells may grow into a lump that is called the tumour. If cancer has spread, treatment becomes more difficult and a person's chance of survival is less. The earlier cancer is found, the greater the chance of survival. With the rapid advancements in information technology, detection of cancer at an early stage is made possible by using classification techniques. In this paper, a new Adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM) based feature selection method is proposed. The Genetic algorithm based Rough Set mechanism is used to select the feature selection.

The rest of the paper is structured as follows; section 1 provides the brief introduction in data mining in the field of cancer medical images. Section 2 discusses the various existing papers which are based on cancer detection and classification methods. It is followed by section 3 includes the proposed model of feature selection and classification. The next section 4 contains the discussions of the proposed method. Finally, section 5 brings conclusion of the accuracy of cancer classification.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

II. MATERIALS AND METHODS

The following papers are motivated to propose the Cancer Detection Methodology using Fuzzy based Classification Techniques.

In 2016, Hamza Turabieh (3) focused on breast cancer recurrence problem, hybridizing two methodologies, Genetic Algorithm (GA) and Adaptive Neuro Fuzzy Inference System (ANFIS), to develop a good diagnosis system. The author examined their proposed hybrid methodology using classification accuracy, sensitivity, and specificity. The literature survey is also classified in this paper to discuss the classification results.

In 2016, Radhanath Patra and Shankha Mitra Sunani (4) has reviewed of diagnosis of breast cancer database, some of following points were much emphasized. Some of the review paper already predicted about the accuracy of different machine learning algorithm .Due to limitation of ANN various modified form of machine learning algorithm as well as hybrid process were adopted to improve accuracy with least time span .some of paper clearly presented that accuracy was above 95% to 99% . Various data mining process also considered predicting the diagnosis of breast cancer with high accuracy and result showed that decision tree, KNN (K-Nearest Neighbor) algorithm and PSO (Particle Swarm Optimization) is better in terms of accuracy calculation. Feature extraction with some good algorithm should be carried out with certain modified form of machine learning algorithm or data mining process to a large volume of a certain data in a small time span with much higher accuracy.

In 2015, Rajamani.R and Rathika.M (5) provides the overview on liver cancer analysis using Adaptive Neuro Fuzzy Inference System (ANFIS) data mining technique. The author's consider the input is 2-D CT (Computed Tomography) images. In data pre-processing step, the noise removal in the CT image, segmentation process, morphological operation and the feature extraction techniques has been discussed. The authors have also discussed the study of Adaptive Neuro Fuzzy Inference System for early detection of Liver Cancer in human. They discussed the implementations of this technique or combination of ANFIS with other data mining techniques can be made to help the medical field at early diagnosis of liver cancer.

III. PROPOSED METHOD

The main aim of this paper is to reduce the features using Rough Set and Genetic Algorithm to improve the accuracy of the classification process. The experimentation of the entire processing was performed with Cancer dataset. This dataset contains the information about benign and malignant tumours. The architecture of the proposed method is figured as follows,

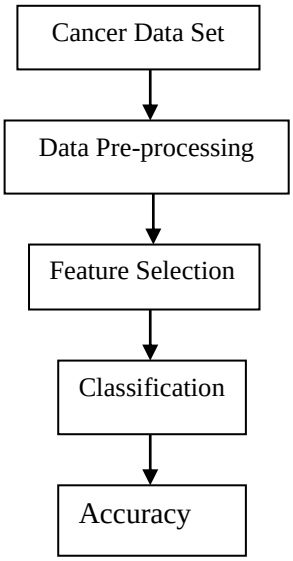

Figure (3.1). Workflow of the Proposed Method

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Data Pre-processing

In the pre-processing method, the raw data is first partitioned into three groups like, i) A finite set of objects, ii) the set of attributes such as features and variables, iii) The domain of attribute. For each group in the dataset, a decision system is constructed. Each decision system is subsequently split into two parts: the training dataset and the testing dataset. Each training dataset uses the corresponding input features and falls into normal (+1) and abnormal (-1) classes.

Feature selection

Two types of feature selection techniques are used in this proposed method.

1. Rough Set Technique

The rough set based feature selection technique as the process of finding a subset of features, from the original set of pattern, optimally according to defined criteria. Rough Sets theory is based on the concept of lower and upper approximation of a set, the approximation space and models of sets. An information system can be represented as,

S = (U,A,V,f)

Where U is the universe, a finite set of N objects ($x_1, x_2, \ldots x_N$) a nonempty set, A is a finite set of attributes, V = $U_{a \in A} V_a$ (where V_a is a domain of the attribute a), f: U * A -> V is the total decision function also called as the information function such that f(x,a) € V_a for every a € A, x € U. B subset of attributes B € Q defines an equivalence relation (called an unnoticeable relation) on U.

(A) = { (x,y) € U for all a € B; F(x,a) = f(y,a)} Denoted also by A.

The straightforward feature selection procedures are based on an evaluation of the predictive power of separate features, followed by a ranking of such evaluated features eventually the choice of the first best m features. A criterion applied to an individual feature could be either of the open-loop or closed-loop type. It can be expected that a single feature alone may have a very low predictive power, whereas when put along with others, it may demonstrate a significant predictive power.

2. ***Feature selection using Genetic Algorithm***

A GA is starting by generating a large set of possible solutions to a given problem which means a solution to a problem corresponds to a genome or chromosome in genetics; a large set of possible solutions to a given problem corresponds to a population (6). It evaluates each of those solutions and decides on a “Fitness level” for each solution set. In generally the genetic algorithm contains the following steps:

- Generate an initial population: An initial population is created from a random selection of solutions.
- Evaluation Fitness: A value for fitness is assigned to each solution (chromosome) depending on how close it actually is to solving the problem, thus arriving at the answer of the desired problem.
- Reproduce, Selection, Mutate and crossover: Those chromosomes with a higher fitness value or more likely to reproduce offspring which can mutate/inverse after reproduction.
- Control Next Generation: If the new generation contains a solution that produces an output that is close enough or equal to the desired answer then the problem has been solved. If this is not the case, then the new generation will go through the same process as their parents did. This will continue until a solution is reached.

Classification

In this proposed methodology the classification mechanism is performed using ANFIS and SVM classifier.

Adaptive Neuro-Fuzzy Inference System Classifier

ANFIS is based conventional mathematical tool. This tool is also known as fuzzy modelling or fuzzy identification. ANFIS is a combination of fuzzy logic and neural network which is used to form a hybrid intelligent system that extends the ability to learn automatically and adapt. It can be trained as hybrid learning algorithm, like this kind of hybrid system are used by the following criteria,

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

- There are no standard methods exist for transforming human knowledge to experience in the rule base and data base of a Fuzzy Inference System (FIS).
- Need for effective methods from tuning the membership functions so as to minimize the output error measure or maximize performance index.

Support Vector Machine (SVM) Classifier

Support vector machines are called as supervised learning methods and it is used for classification and regression a task that is from statistical learning theory (8). They belong to a family of generalized linear classification. Normally, a classification task involves training and test sets which consist of data instances. The original idea of SVM was developed for linearly separable data. SVM tunes the capacity of the classification function by maximizing the margin between the training patterns and decision boundary. In the high dimensional feature space, simpler and linear hyper plane classifiers that have a maximal margin between the classes can be obtained [9].

IV. RESULT AND DISCUSSION

In this methodology, the breast cancer dataset were used to process feature selection and classification. The entire processing of the breast cancer detection methodology is discussed as follows; the data of the cancer images is get from Wisconsin Breast Cancer Dataset (WBCD),

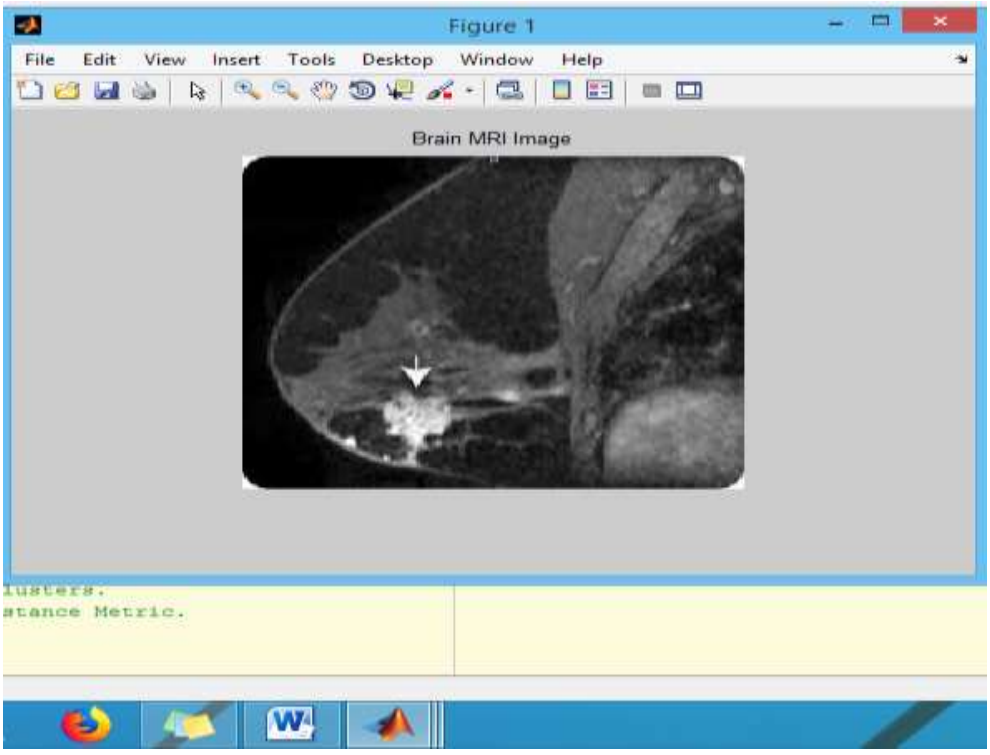

Figure (4.1). Pre-processed Breast cancer Image

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

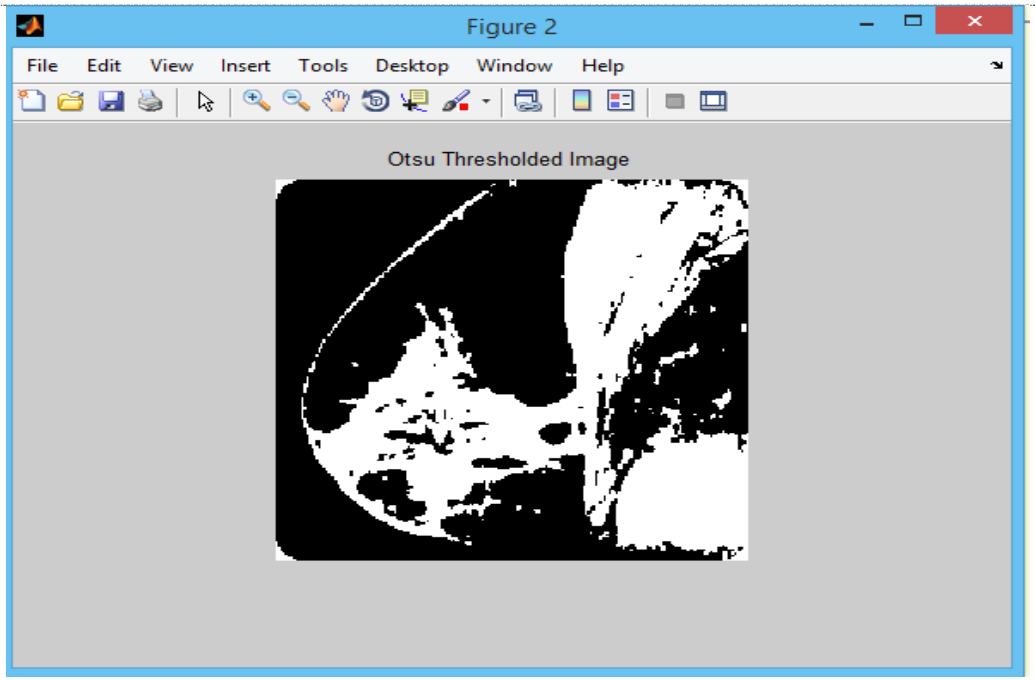

Figure (4.2). Thresholded Image

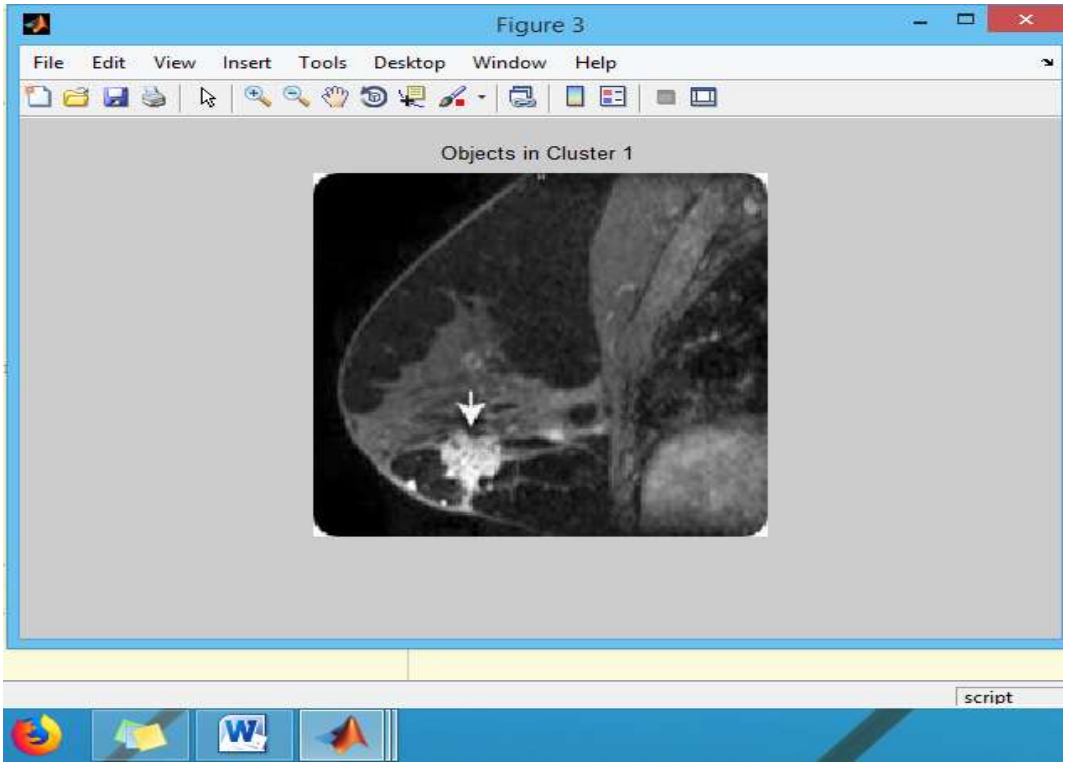

Figure (4.3). Clustered Brest cancer Image

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

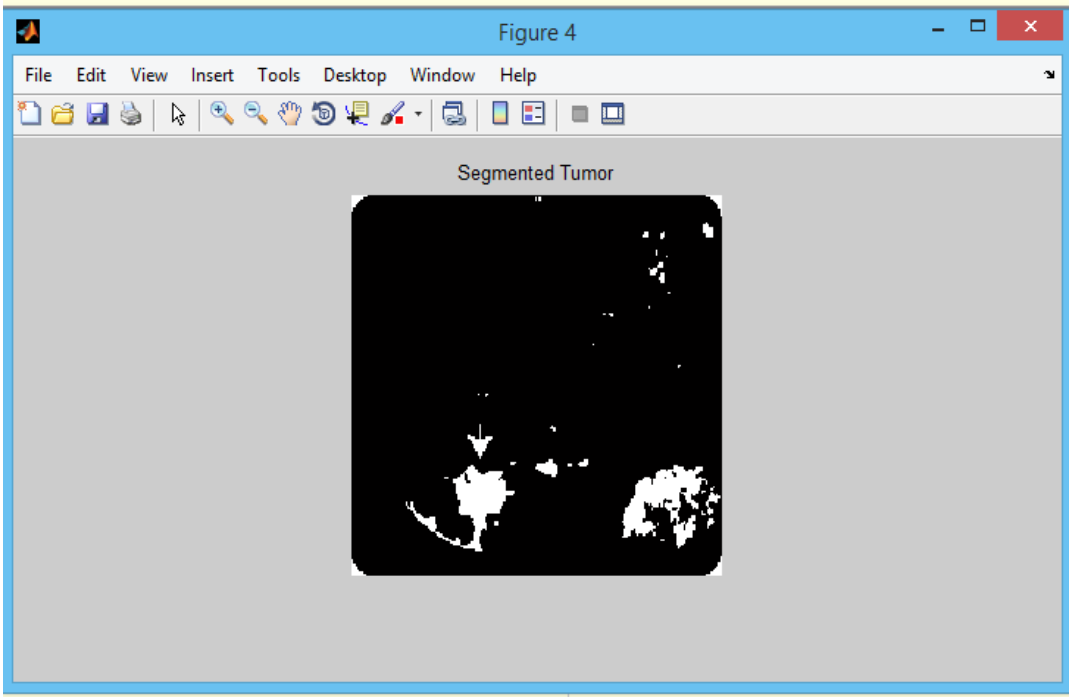

Figure (4.4). Segmented Tumor Image

Table 1. Cancer Dataset Details

Cancer Type	**Number of Dataset**	**Number of Samples**	**Number of Features**
Breast	Wisconsin Breast Cancer Dataset (WBCD)	569	32

From the datasets, the features are reduced by using the Rough set method and Genetic Algorithm. The reduced number of features gives better classification accuracy. The accuracy was calculated using classification techniques SVM and ANFIS.

The performance of the classifiers was evaluated by percentage of accurately. The accuracy indicates the performance of the classifier. The accuracy calculation formula is derived as follows,

Accuracy (%) = (TP+TN) / (TP+FN+TN+FP)*100

The classification accuracy of cancer datasets using SVM and ANFIS,

Table 2. Classification Accuracy for Breast cancer using ANFIS and SVM

Cancer	**Datasets**	**ANFIS (%)**	**SVM (%)**
Breast Cancer	WBCD	98.92	93.02

From the above table, it is evident that the proposed novel method provides better results.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

V. CONCLUSION

The perfect accuracy of cancer classification is important for human's life. Normally, using diagnosis tool, most of the researchers are interested in Artificial Intelligence (AI) classification techniques to classify cancer. The main scope of this paper is conducted in order to compare the performance of two feature selection methods and two AI classification techniques namely, SVM and ANFIS in classifying cancer data. This paper focuses on selecting the main features to improve the accuracy of the classification algorithms. Both the SVM and ANFIS technique are effective in order to classify cancer data.

VI. REFERENCES

[1] American Cancer Society, http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-040951.pdf.
[2] Murat Cınar, Mehmet Engin et al.,"Early prostate cancer diagnosis by using artificial neural networks and support vector machines", Expert Systems with Applications: An International Journal, April 2009.
[3] Hamza Turabieh, "GA-based Feature Selection with ANFIS Approach to BreastCancer Recurrence", IJCSI International Journal of Computer Science Issues, Volume 13, Issue 1, ISSN (Print): 1694-0814. January 2016.
[4] Radhanath Patra and Shankha Mitra Sunani, "A Review On Different Computing Method For Breast Cancer Diagnosis Using Artificial Neural Network And Data mining Techniques", Int. J. Adv. Res. 4(11), pp. 598-610.
[5] Rajamani.R and Rathika.M, "Analysis of Liver Cancer using Adaptive Neuro Fuzzy Inference System (ANFIS)", International Journal of Innovative Research in Computerand Communication Engineering, Vol. 3, Issue 7, July 2015.
[6] Akanksha Sharma et al., 'Optimized Liver Tumor Detection and Segmentation Using Neural Network' International Journal of Recent Technology and Engineering (IJRTE) Volume-2, Issue-5, November 2013.
[7] Anant Bhardwaj et al., 'An Approach to Medical Image Classification Using Neuro Fuzzy Logic and ANFIS Classifier' International Journal of Computer Trends and Technology- volume4Issue3, 2013.
[8] C. Kalaiselvi et al., "A Novel Approach for the Diagnosis of Diabetes and Liver Cancer using ANFIS and Improved KNN", ResearchJournal of Applied Sciences, Engineering and Technology July 10, 2014.
[9] P.Subhasri and Dr.A.Padmapriya., "Enhancing the Security Of Dicom Content Using Modified Vigenere Cipher", International Journal of Applied Engineering Research, Volume: 10(55), January 2015, pp. 1951-1956.