Early Detection of Cancer using Soft Computing

Amjad Khan¹ Zahid Ahmed Ansari²

¹Department of Electronics and Communication Engineering ²Department of Computer Science & Engineering

¹,²P.A College of Engineering Mangaluru-574153, INDIA

Abstract—Cancer is one of the most frequent cause of death for both men and women. The early detection of cancer is very essential to provide the timely treatment so that the probability of curing the disease increases. The pain and toxicity can be minimized if the disease is found in initial stages. Medical image mining refers to the application in data mining techniques on medical image data. The main goal of medical image mining is to extract clinically relevant information from the medical images to assist the physicians in accurate diagnosis and detection of diseases. Soft computing is a consortium of methodologies that provides flexible information processing capability. Soft computing techniques such as fuzzy sets, neural networks, genetic algorithms, and rough sets are most widely applied for Cancer image mining. This paper presents the early detection of Cancer based on image mining using soft computing techniques. Hence the related issues can be resolved suitably using most efficient soft computing based image mining techniques.

Key words: Soft Computing, Cancer, Fuzzy Sets, Neural Networks, Image Mining

I. INTRODUCTION

Cancer is one of the most frequent causes of death for both men and women. The early detection of cancer is very essential to provide the timely treatment so that the probability of curing the disease increases [32]. The pain and toxicity can be minimized if the disease is found in initial stages. Screening is the most popular method to detect the disease in the early stages before the symptoms are shown. Since the disease diagnoses is a very challenging and important task, medical research fraternity has received enormous interest globally in this area [23].

Some of the most popular medical data used are CT scan, PET scan and MRI scan images. Analyzing this large amount of data is a challenging task therefore it has become necessary to use data mining techniques to help in efficient and accurate prediction and diagnosis of cancer [34]. Medical image mining refers to the application in data mining techniques on medical image data. The main goal of medical image mining is to extract clinically relevant information from the medical images to assist the physicians in accurate diagnosis and detection of diseases [10], [11].

In order to analyze the cancer images, two of the most popular data mining techniques are “clustering” and “classification”. Clustering is an unsupervised classification technique used for grouping the images into various disease classes. The selection of appropriate method for clustering is an important task in medical diagnosis and the choice of efficient clustering method should be done carefully to suits the desired task [3], [12]. The fundamental objective of carrying out image classification or clustering in medical image mining is to analyze the image contents and classify them into appropriate disease categories. [4], [13].

Soft Computing played an important role in early detection and diagnosis of cancer diseases analytically with improved effectiveness and suitable accuracy with the help of methods and proper attributes reference value. Soft computing is a consortium of methodologies that provides flexible information processing capability [4]. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solutions [1]. Different clinical values for attributes and biomarkers have been taken as an input and match these with the reference values to predict the diseases accurately. The outcome of this would help doctors, scientists, pharmacists in understanding the characteristic and association of attributes which is responsible for these diseases and provide proper diagnosis method and in discovering new drugs. The early detection of cancer can be helpful in curing the disease carefully. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing [34].

In the following, section II presents a literature review of various cancer based papers and in Table-1 comparison of soft computing methods along with performance. Next in section III, early cancer detection processes along with the diagram were explained. In section IV, soft computing methods and in section V, final conclusions were discussed.

II. LITERATURE REVIEW ON CANCER DISEASES

Here, following are the reviews on various soft computing based cancer disease research papers are described briefly.

Cancer is one of the most important causes of death for both men and women. The early detection of cancer can be helpful in curing the disease absolutely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing [34].

Anam et.al [15] proposed a computerized system for lung nodule detection in CT scan images. The system, lung segmentation and enhancement, feature extraction and classification. Threshold segmentation is applied to remove background and extracts the nodules from an image. Here regions are classified and small nodules are detected using neuro fuzzy classifier which leads to early diagnosis of lung cancer.

Dansheng et.al [16] provides centrosomal numeral and morphological abnormalities. Here Regions of interest were selected based on segmentation, feature abstraction, and optimization using Linear discriminant analysis (LDA) and support vector machines (SVM).
Disha et al. [17] develops a system for early detection of lung cancer by analyzing lung CT images using region growing segmentation algorithm. Then rule based technique is applied to classify the cancer nodules.

Aparna et al. [18] presented an automated process of tumor delineation and volume detection from each frame of PET lung images. K-nearest neighbor and support vector machines (SVM) classifiers were used to analyze performance of features.

Yang et al. [19] proposed a two stage scheme for automatic lung nodules detection in Multi-Slice Computed Tomography (MSCT) scans with multiple SVMs. Three SVMs classifiers for different slice directions are used on preprocessed images to categorize the candidates as nodule or non-nodule.

Samuel et al. [20] presented a nonlinear anomaly detector called kernel RX-algorithm and apply it to CT images for malignant nodule detection. The approach can be an efficient technique for lung cancer detection at early stage.

Yang et al. [21] presented a computer aided lung nodule detection scheme based on analysis of enhanced voxel in three dimensional (3D) CT image SVM classifier is applied to categorize the initial nodule candidates as nodule or non-nodule. Input to SVM classifier is eight features extracted from 3D initial nodule.

Atiyeh et al. [22] aim at presenting a method to improve the efficiency of the lung cancer diagnosis system, through proposing a region growing segmentation method to segment CT scan lung images. Linear-filtering and contrast enhancement used as preprocessing step for noise removal, to prepare the image for segmentation. Afterwards, cancer recognition are presented by Fuzzy Inference System (FIS) for differentiating between malignant, benign and advanced lung nodules.

Anita et al. [23] aim to get the more accurate results by using various enhancement and segmentation techniques. In the segmentation stage the Watershed and Thresholding Segmentation is used and comparison has been made.

Lee et al. [24] presented a method that includes three stages: image acquisition, background removal and nodule detection for identification of lung nodules. The system was tested on few images containing nodules and few containing no nodules which were randomly selected from the database images. Random forest based classifier performs well to detect all the nodules in the images and recorded a low false detection rate.

Vijai et al. [25] proposed a system which efficiently predicts lung tumor from Computed Tomography (CT) images through image processing techniques coupled with neural network classification as either benign or malignant. Optimal thresholding is applied to the denoised image to segment Lung nodules.

Fan et al. [26] presented a feature-based imaging classification method to classify the lung nodules in low dose computed tomography (LDCT) slides into four categories: well circumscribed, vascularized, juxta-pleural and pleural-tail. SVM classifier is used to conduct the classification.

Sivakumar et al. [27] develop an efficient lung nodule detection scheme by performing nodule segmentation through weighted fuzzy possibilistic based clustering is carried out for lung cancer images. Support Vector Machine (SVM), a machine learning technique is used for classification.

Arfan et al. [28] describes a method for lung segmentation based on Genetic Algorithm (GA) and morphological image processing techniques. Susan thinning algorithm [20] is used to reduce the borders to the width of one pixel.

JIA et al. [29] an automatic computer-aided detection (CAD) scheme is presented that can identify the pulmonary nodule at an early stage from CT images. Region-growing algorithm is proposed for Trachea and Main Airway Bronchi Elimination. Then, the filter of nodule candidates, the detection of nodule candidates, the feature extraction and classification, three-dimensional visualization is done.

Hiram et al. [30] presented a very simple but efficient methodology for lung nodule classification without the stage of segmentation. Eight texture features were extracted from the histogram and the gray level co-occurrence matrix (with four different angles) after image acquisition for each CT image. Support vector machine (SVM), used to classify lung tissues into two classes: with lung nodules and without lung nodules.

Fatma et al. [31] presented a Bayesian classification and a Hopfield Neural Network algorithm for extracting and segmenting the sputum cells for the purpose of lung cancer early diagnosis. The HNN segmentation algorithm outperforms the Fuzzy C-Mean clustering.

Kesav et al. [32] proposed an early lung cancer detection methodology using nucleus based features. Seeded region growing segmentation method is used to perform nucleus segmentation. An additional criterion like nucleus size to seeded region growing method is used for better accuracy.

Negar et al. [33] proposed a novel classification method called iterative linear discriminant analysis and use this method in addition to fuzzy c-means clustering for successful false positive reduction. The novel scheme is superior over rule-based FP reduction of detected candidates.

Krishnaiah et al. [34] studied briefly the potential utilization of classification based data mining techniques like as Decision tree, Rule based, Artificial Neural Network and Naive Bayes to huge volume of healthcare data for diagnosis of lung cancer.

Prashant et al. [35] presented an image mining method which diagnosis lung cancer at an early stage by means of CT scan images which are available in Dicom (DMC) format. ANN, SVM and K-NN were the three classifiers used for the prediction of lung cancer and to find the intensity of disease.

Ada et al. [36] developed a hybrid image mining technique, that is based on Principal Component Analysis (PCA) and feature extraction for lung cancer detection in CT scan images. To build the image superior and improve it from corruption, noising or interference the Histogram Equalization was used.
Nancy et al. [37] developed an image mining method to identify lung cancer in its beginning stage. They automated the classification process using classification algorithm, that is Neural Network and for optimization Genetic Algorithm (GA).

Ankit et al. [38] performed investigation on lung cancer data from SEER association rule mining technique in order to recognize hotspots in the cancer data. In hotspots, the patient’s survival times is considerably higher than and lower than the normal survival time across the whole dataset.

Malignant melanoma is the third most frequent type of skin cancer and one of the most malignant tumors. The early diagnosis of malignant melanoma is therefore a crucial issue for dermatologists. The main objective of such systems is to facilitate the dermatologist in different analysis steps, such as the lesion boundary detection, diagnostic features, the classified into different classes of lesions, the visualization, storage, the database management, etc. Skin cancers can be classified into two major groups which are melanoma and non-melanoma skin cancers [39].

Kaur et al. [39] described a system for automatic inspection and estimation of melanoma and pigmented skin lesions. This automated system supported images of skin lesions acquired employing a digital camera. Image Classification and Fuzzy C-means clustering is employed to diagnose the melanoma and region growing for additional improvement to identify lesion area.

Ammara et. al [40] were presented a review on Image-based computer aided diagnosis systems for screening and early detection of malignant melanoma of skin cancer and examine current practices, problems, and prospects of image acquisition, preprocessing, segmentation, feature extraction and selection, and classification of dermoscopic images.

Nishu et al.[41] were described about the early detection of skin cancer using dermoscopic images carried by applying artificial neural networks. The images are subjected to various pre-processing for noise removal, image enhancement and image segmentation using Thresholding. The feature extraction technique - 2D Wavelet Transform method is being used to extract the skin cancer regions. These features are given as the input nodes to the neural network.

Krupal et.al. [42] describes the two predictive models for identification of skill illinesses to help public health management system using multiclassification problem. The models are developed using Artificial Neural Network and Support Vector Machine. These two approaches are applied on the multi class classification dataset and some comparative inferences are generated using F-scores.

Fatima et al. [43] proposed outlines an approach for recognizing breast cancer diagnosis using neuro-fuzzy inference technique namely ANFIS (Adaptive Neuro-Fuzzy Inference System). According to the results obtained it is noted that the used method is very promising approach for recognition.

Ashwini et al. [44] proposed a hybrid approach of feature selection which approximately reduces 75% of the features for the breast mammograms. It uses a new decision tree approach for classification to classify and detect the cancerous tissue. The approach can reduce the computation cost of mammogram image analysis and can be applied to other image analysis applications. The algorithm uses simple statistical techniques in collaboration to develop a novel feature selection technique for medical image analysis.

III. EARLY CANCER DETECTION

Early cancer detection by analyzing the related cancer images is the proposed concept as shown in the figure-1. The Soft Computing techniques are used to detect cancer by comparing the features of sample images taken from the patient with the expert’s documental data.

The cancer image patterns are designed and these patterns are compared with the sample image data to find out the affected cancer patterns by applying the Soft Computing Techniques.

The Soft Computing Techniques and their performance on various Cancer diseases are listed in the Table-1. Then the Clustering techniques are applied to form the clusters of related patterns. By analyzing the discovered cluster, the final detection of cancer is done.

<table>
<thead>
<tr>
<th>Soft computing Methods used</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Image</th>
<th>Citation</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWT, FCM Region Growing Algorithm</td>
<td>97%</td>
<td>99%</td>
<td>Melanoma</td>
<td>[39]</td>
<td>2015</td>
</tr>
<tr>
<td>Threshold segmentation Neurofuzzy classifier</td>
<td>97%</td>
<td>95%</td>
<td>CT</td>
<td>[35]</td>
<td>2014</td>
</tr>
<tr>
<td>Threshold segmentation Neurofuzzy classifier</td>
<td>95%</td>
<td>NA</td>
<td>CT</td>
<td>[15]</td>
<td>2013</td>
</tr>
<tr>
<td>SVM Classifier</td>
<td>84%</td>
<td>NA</td>
<td>CT</td>
<td>[30]</td>
<td>2013</td>
</tr>
<tr>
<td>Seeded region growing Random forest</td>
<td>87%</td>
<td>NA</td>
<td>Sputum</td>
<td>[32]</td>
<td>2013</td>
</tr>
<tr>
<td>Region Growing Artificial Neural n/w</td>
<td>NA</td>
<td>95%</td>
<td>CT</td>
<td>[22]</td>
<td>2013</td>
</tr>
<tr>
<td>SVM Classifier fuzzy possibilistic clustering</td>
<td>80.12%</td>
<td>82.10%</td>
<td>CT</td>
<td>[27]</td>
<td>2013</td>
</tr>
<tr>
<td>Bayesian Neural Network</td>
<td>88.62%</td>
<td>NA</td>
<td>Sputum</td>
<td>[31]</td>
<td>2012</td>
</tr>
<tr>
<td>SVM, Entropy Threshold</td>
<td>85%</td>
<td>NA</td>
<td>CT</td>
<td>[16]</td>
<td>2012</td>
</tr>
<tr>
<td>Sobel Edge detection, Diagnostic Indicators</td>
<td>80%</td>
<td>NA</td>
<td>CT</td>
<td>[17]</td>
<td>2011</td>
</tr>
<tr>
<td>Hotspot algorithm FCM-Assocation</td>
<td>85%</td>
<td>NA</td>
<td>CT</td>
<td>[38]</td>
<td>2011</td>
</tr>
<tr>
<td>Optimal Thresholding BPN classification</td>
<td>86%</td>
<td>NA</td>
<td>CT</td>
<td>[25]</td>
<td>2010</td>
</tr>
<tr>
<td>SVM Threshold Segmentation</td>
<td>87.82%</td>
<td>93.75%</td>
<td>CT</td>
<td>[21]</td>
<td>2009</td>
</tr>
<tr>
<td>k-NN, SVM</td>
<td>96.87%</td>
<td>NA</td>
<td>PET</td>
<td>[18]</td>
<td>2008</td>
</tr>
<tr>
<td>Threshold segmentation Neurofuzzy classifier</td>
<td>NA</td>
<td>94.96%</td>
<td>CT</td>
<td>[29]</td>
<td>2007</td>
</tr>
<tr>
<td>FCM Clustering</td>
<td>NA</td>
<td>81%</td>
<td>CT</td>
<td>[33]</td>
<td>2006</td>
</tr>
</tbody>
</table>

Table 1: Comparison of soft computing methods and performance
Soft computing is a consortium of methodologies that provides flexible information processing capability [4]. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solutions [1]. Soft computing is a collection of methodologies that:

- Exploit the tolerance for imperfection and uncertainty
- Provide capability to handle real life ambiguous situations
- Try to achieve robustness against imprecision

Principal Soft computing techniques includes fuzzy sets, neural networks, genetic algorithms, and rough sets are most widely applied in the medical image mining. Generally fuzzy sets are suitable for handling the issues related to understandability of patterns, incomplete/noisy data, and mixed media information which provide approximate solutions faster.

Neural networks are nonparametric, robust, and exhibit good learning and generalization capabilities in data-rich environments. Genetic algorithms provide efficient search algorithms to select a model from mixed media data based on some objective function. Rough sets are suitable for handling different types of uncertainty in medical image data [5] [8] [9].

Soft computing based classification technique consists of fuzzy classification, neural classification etc. Classification is a supervised learning problem which classifies a data item into one of several predefined categorical classes in which the output is a discrete classification and the possible mutually exclusive classes of the problem[4], [7].

Soft computing based clustering technique consists of fuzzy clustering, artificial neural networks for clustering etc. Clustering is the unsupervised classification of images into groups. Classification of similar objects into different groups, or more precisely, the partitioning of a data set into subsets called clusters.

Clustering is a common technique for statistical data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics [3]. Among the fuzzy clustering methods, fuzzy c-means (FCM) algorithm is the method used in image segmentation because it has robust characteristics for ambiguity and it has a good performance in a large class of images.

The k-Means clustering algorithm is one of the most commonly used methods for partitioning the data. [5]. A fuzzy clustering: This method assigns degrees of membership in several clusters to each input pattern. A fuzzy clustering can be converted to a hard clustering by assigning each pattern to the cluster with the largest measure of membership [14], [3].

Artificial neural networks (ANNs) for clustering: the features of ANNs are inherently parallel and distributed processing architectures. They can act as pattern normalizers and feature selectors by appropriate selection of weights.[14], [3].

REFERENCES

[7] Yushu Xiong, “A Clustering Algorithm Based on Rough Set and Genetic Algorithm” TELKOMNIKA,
Vol. 11, No. 10, October 2013, pp. 5782 ~ 5788, e-ISSN: 2087-278X.


