Lung Cancer Detection in its Early Stages Using a C-means based Fuzzy Hopfield Neural Network

Farhad Kazemi MUN #:201576196 July 4, 2016

Abstract

Lung cancer is the second most common cancer in both men and women.
The early detection of lung cancer is a challenging problem due to the structure of the cancer cells, where most of the cells are overlapped with each other. Our algorithm aims to detect cancer before a person has any symptoms, which can help finding cancer at an early stage when it might be easier for early detection and treatment of lung cancer. In particular, we propose a C-means based training of a Fuzzy Hopfield Neural Network and apply it to detect and segment cancer cells in images. We present a new objective function, and its minimization by a type of Lyapunov energy function which is based on a two dimensional Fuzzy Hopfield Neural Network. This objective function is the same energy function of the Hopfield Neural Network which is improved and includes average distance between image pixels and cluster centers. Our research will be done on lung cells images. There are currently no universally applicable criteria for validating image segmentation methods. Nevertheless, an impact criterion for performance evaluation is whether the method can indicate interesting or important regions in the image.Therefore, we will propose new criteria for solving this problem. An exploratory analysis will also be done to understand how this algorithm operates in comparison to

the Hopfield Neural Network, Fuzzy C-Means, Gustafson-Kessel and C-Means algorithms.

1 Motivation

Lung cancer is the second most common cancer in both men and women (not counting in skin cancer), and also, is the leading cause of cancer death among both men and women. It is difficult to detect in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancer. Each year, more people die of lung cancer than of colon, breast, and prostate cancers combined. There is significant evidence indicating that the early detection of lung cancer will decrease the mortality rate. The most recent estimates according to the latest statistics provided by world health organization indicates that around 7.6 million deaths worldwide each year because of this type of cancer. Furthermore, mortality from cancer are expected to continue rising, to become around 17 million worldwide in 2030 [?].

There are many techniques to diagnosis lung cancer, such as Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology [?]. However, most of these techniques are expensive and time consuming. In other words, most of these techniques are detecting the lung cancer in its advanced stages, where the patient's chance of survival is very low. Therefore, there is a great need for a new technology to recognize the lung cancer in its early stages. Image processing techniques provide a good quality tool for improving the manual analysis. A number of medical researchers utilized the analysis of lung cells for early detection of lung cancer [?], most recent research relay on quantitative information, such as the size, shape and the ratio of the affected cells [?].

For all these reasons we attempt to produce automatic diagnostic system for detecting lung cancer in its early stages based on the analysis of the lung cells images. Moreover, we have developed a Fuzzy Hopfield technique for segmentation of the lung cells image to divide the images into several meaningful sub regions. Image segmentation has been used as the first step in image classification and clustering. There are many algorithms which have been proposed in other articles for medical image segmentation. A review of image segmentation techniques can be found in the next section.

2 Previous and Related Work

Image segmentation, a process to divide a given image into meaningful regions with homogeneous properties, is an important step in image analysis and recognition. A large number of algorithms have been proposed in previous years. Those conventional image segmentation algorithms can be categorized generally into three classes:1) histogram-based schemes, where the pixels are segmented into classes based on overall gray levels; 2) clustering by which homogeneous properties around a given pixel are enlarged; and 3) edge-based schemes, which detect the pixels with abrupt changes in gray levels, and then connects selected pixels to form completely enclosed boundaries[**?**].

Clustering is useful in several exploratory pattern analysis, grouping, and machine-learning situations, decision-making, data mining, document retrieval, image segmentation, and pattern classification. In image segmentation techniques, image is segmented to different regions separated with contours. Region growing, k-means, and split and merge methods are used generally for image segmentation. Besides these crisp classical segmentation methods, the fuzzy logic methods were also seen to be very effective for segmentation [**?**, **?**, **?**]. The clustering techniques that Naz et al.[**?**] reviewed are segmentation for noisy medical images with spatial probability, Novel Fuzzy C-Means Clustering (NFCM), Fuzzy Local Information C-Means (FLICM) and Improved Spatial Fuzzy C-Means Clustering (ISFCM) algorithms. Recently, a novel unsupervised image-based SPM (Subpixel mapping) model based on the fuzzy c-means (FCM) clustering approach was proposed by Zhang et al [**?**].

Jiang et al. [**?**] provided a focused literature survey on recent neural network developments in computer-aided diagnosis, medical image segmentation and edge detection towards visual content analysis, and medical image registration for its pre-processing and post-processing, with the aims of increasing awareness of how Neural Networks can be applied to these areas and to provide a foundation for further research and practical development. In 1982, Hopfield proposed the so-called Hopfield Neural Network, which possesses auto-associative properties. It is a recurrent (fully interconnected) network in which all neurons are connected to each other, with the exception that no neuron has any connection to itself[**?**]. The Hopfield neural network is a well-known technique used for solving optimization problems based on Lyapunov energy function. Amatur, Piriano and Takefuji used the two dimensional

Hopfield Neural Network for segmentation of multi-spectral MR images[?]. An approach to tomographic image reconstruction from projections based on Hopfield Neural Network is developed and investigated by Cierniak [?]. Robust segmentation of medical images using competitive Hopfield Neural Network as a clustering tool was proposed by Roozbahani, Ghassemian and Sharafat[?].

Combination of Fuzzy and Hopfield is a good technique for some problems. For example, Lin, Cheng and Mao proposed the segmentation of single and multi-spectral medical images using a Fuzzy Hopfield Neural Network [?, ?].Fuzzy Hopfield Neural Network with fixed weight for medical image segmentation was proposed by Chang and Ching [?].
Suresh and others [?] proposed the task of segmenting skin lesions in Dermoscopy images based on clustering techniques such as Fuzzy C Means Algorithm (FCM), Possibilistic C Means Algorithm (PCM), Hierarchical C Means Algorithm (HCM); C-mean based Fuzzy Hopfield Neural Network, Adaline Neural Network and Regression Neural Network for the early diagnosis of Malignant Melanoma.

In this study, we examine the ability of the new algorithm to detect and segment cancer cells based on Fuzzy Hopfield Neural Network. Our algorithm will be done on lung cells images. We present a new objective function, and its minimization by a type of Lyapunov energy function which is based on two dimensional Fuzzy Hopfield Neural Network.

3 Research Question(s)

The early detection of lung cancer is a challenging problem due to the structure of the cancer cells, where most of the cells are overlapped with each other. In this study, we propose a new Fuzzy Hopfield Neural Network (FHNN) for lung cells image segmentation. An analysis will also be done to understand how this algorithm operates in comparison to Hopfield Neural Network (HNN), Fuzzy C-Mean (FCM), Gustafson-Kessel and C-Means algorithms. In particular, we propose a C-means based training of a Fuzzy Hopfield Neural Network and a new objective function, and its minimization by a type of Lyapunov energy function which is based on a two dimensional Fuzzy Hopfield Neural Network. This objective function is the same energy function of the Hopfield Neural Network which is improved and includes average distance between image pixels and cluster centers.

However, the lung cells images are characterized by a noisy and cluttered background patterns that make the segmentation and automatic detection of the cancerous cells very problematic. In addition to that there are many debris cells in the background of the images. We aim to produce a new method that detects lung cancer cells promising than other methods. In other words, we provide a pre-processing technique which can mask all these debris cells and keep the nuclei and cyto-plasm. In the literature we found, there have already been attempts to solve this problem using Hopfield Neural Network and Fuzzy C-Means[**?**], thus we propose a new method (Fuzzy Hopfield Neural Network) to improve the Hopfield Neural Network technique which was used in [**?**]. Moreover, Taher et al. presented two segmentation methods, Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm for segmenting sputum color images to detect the lung cancer in its early stages. Our goal is to well extract the lung cells, the nuclei and cytoplasm to be ready for the segmentation process where we want to partition these cells into regions, then these regions will be diagnosed to see if its a normal cells or a cancerous cells.

There are currently no universally applicable criteria for validating image segmentation methods. Nevertheless, an impact criterion for performance evaluation is whether the method can indicate interesting or important regions in the image. Therefore, we will propose new criteria for solving this problem.

4 Proposed Method: Fuzzy Hopfield Neural Network

In this section, we present the proposed Fuzzy Hopfield Neural Network to simulate the membership matrix for cancer cells image segmentation. Each pixel in the image is a point in the plane. If there are n pixels to be divided into c clusters, then each pixel has c neurons associated with it. There are n*c neurons in this Neural Network.

The number of neurons depends on the image size; the larger the image size is, the more neurons are required. If the size of image is R*C pixel, so the total number of data points is R*C=N. So, the Fuzzy Hopfield Neural Network consists of N *c neurons that can be conceived as a 2-D array for the image segmentation problem. Where N is the number of pixels and c is the number of clusters.

4.1 The proposed segmentation algorithm is summarized as following:

1) Given the image X in square size, choose the number of clusters $1 < c < cmax$**, the termination tolerance** $\epsilon > 0$**.**
2) Convert RGB image to gray scale image and Normalization (gray levels between 0 and 1) The normalization operation guarantees that each image pixel will be absorbed on several classes with certain probability degrees so there will be N data points assigned among c clusters.
3) convert 2-dimensional image to 1-dimensional array. Now, we can consider membership of each pixel to each cluster (depends on the number of clusters). Also, we need to convert address of pixel (i,j) in 2-dimensional image to a new address in 1-dimensional array. So we applied the following formula:

Converting address of pixel(i,j) in two-dimensional image to new address in one-dimensional array:

$$= [(i-1)(U_1) + (j-1)] \tag{1}$$

Where U_1 = c (maximum number of pixels in one row or column in the original image)
4)Determine the neighborhood of neurons through the original image.
Two neurons are neighbors to each other if their corresponding pixels in the image are neighbors to each other. Actually each neuron receives contributions from the neighboring neurons and itself as its input.
5) Calculate of primary centroids (v_0). This method requires a set of initial cluster centers. The initial cluster centers do not need to be exact, but they should not be far away from the true centers. The initial cluster centers can be either given by user assists or obtained from the global information about the gray-scale distribution of the image. In most of the cases for medical image segmentation, the intensities of the regions of interest are known. Users can provide the cluster centers from such available knowledge. Otherwise, the cluster centers can be estimated using a C-Means method or Fuzzy C-Means. Furthermore, we used Fuzzy C-Means for estimating of

initial cluster centers.

6) Compute the initial partition as:

$$\mathrm{O}_{\mathrm{i,k}}^{(0)} = \frac{((I_i - v_k)^{-2})^{1/q-1}}{\sum_{j=1}^{c}((I_i - v_k)^{-2})^{1/q-1}} \tag{2}$$

where I(i) and v_k are, respectively, the intensity values of pixel i and the k 'th class center ($1 \leq i \leq n, 1 \leq k \leq c$). q is the fuzzification parameter.

7) Compute the weights between two neurons i and j

$$\mathrm{w}_{\mathrm{i,j}} = \frac{1}{\Delta I(i,j)^2 + D(i,j)^2} \tag{3}$$

In this method, the weights are fixed and I determined weights first. The $w_{i,j}$ is weight between two neurons i and j. D(i,j) is the Euclidean distance between two neighbor neurons and $\Delta I(i,j)$ is the difference of their intensities in the image.

8) Calculate the input to each neuron (i,k) [?] (The net value of the neuron (i,k))

$$\mathrm{Net}_{\mathrm{i,k}}^{(\mathrm{t+1})} = \sum_j w_{i,j} O_{j,k}^{(t)} \tag{4}$$

9) Compute new output values [?]

$$\mathrm{O}_{\mathrm{i,k}}^{(\mathrm{t+1})} = \frac{((Net_{i,k}^{(t+1)})^{1/q-1}}{\sum_{j=1}^{c}(Net_{i,j}^{(t+1)})^{1/q-1}} \tag{5}$$

10) If $max_{i,k}|O_{i,k}^{(t+1)} - O_{i,k}^{(t)}| < \epsilon$ **then continue** ; otherwise t=t+1 and go to Step 8.

11) Output the final result using the defuzzification process as $S_i = k$, If $o_{i,k} = max_{1 \leq j \leq c}\{O_{i,k}\}$ where S_i is the segmentation label of pixel i.

5 Timeline

Note: Bi-monthly meetings with supervisor.

Table 1: **Timeline of the research.**

Objective	**Schedule**
Literature Review on theories on image segmentation	1st April 2016 - 15 April 2016
Literature Review on theories on cells image segmentation	16 April 2016 - 30 April 2016
Literature Review on datasets (Lung cells images)	1st May 2016 - 10 May 2016
Theoretical analysis of FHNN	11 May 2016 - 31 May 2016
Implementation	1st June 2016 - 30 June 2016
Understanding how this algorithm operates in comparison to another methods	1st July 2016 - 31 July 2016
Writing report	1st Aug 2016 - 31 Aug 2016
Publish my experimental results in Journal or Conference	1st September 2016