A Particle Swarm based Approach for Classification of Cancer based on CT Scan

Ahmed Sultan Al-Hegami Professor of Intelligent Information Systems, University of Sana'a, Yemen

ABSTRACT

In this paper, the technique is propose that made use of Particle Swarm Optimization (PSO) algorithm for tumor detection by using the techniques of Image Processing. This proposed algorithm is based on three steps. First, it identities the affected area, Second, it makes enhancement to the image, Finally, it performs segmentation and extraction of characteristics of the affected area. The propose approach takes any medical image of Computed Tomography CT scan, and provides indicators for physicians decision-making to build treatment plans with minimal diagnostic errors and more accurate description of the treatment plan at minimal cost. The proposed approach has been implemented and tested using data from Oncology Center (the National Center for Oncology Therapy, Hadramout El Wadi, Yemen) and shown very promising .

General Terms

CT scan, Particle Swarm Optimization (PSO) algorithm, Image Processing, Image Classification, cancer, Artificial Intelligence, Feature Extraction, Contrast Improvement Index(CII).

Keywords

CT scan, Particle Swarm Optimization (PSO) algorithm, Image Processing, Image Classification, cancer, Artificial Intelligence, Feature Extraction, Contrast Improvement Index(CII).

1. INTRODUCTION

Owing to no obvious early symptoms of cancer most of patients are diagnosed at an advanced stage [1], which usually results in high costs with a poorer prognosis.

Medical images have become essential in medical diagnosis and treatment. These images playa substantial role in medical applications because doctors exhibit interest in exploring the internal anatomy [2]. Many techniques have been developed based on X-ray and cross-sectional Many techniques have been developed based on X-ray and cross-sectional images like Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), or other topographic modalities [3,4,5]. Medical image processing application has occupied its importance in both technical and clinical aspects for its help towards the detection and analysis of abnormalities, by making the work easier for medical professionals in treating the problem with more scientific and sophisticated methods [6].

The word tomography is derived from two Greek words; tomes, which means slice or section, and graphia, which means description. CT Scan is an imaging modality which uses X-rays to obtain structural and functional information about the human body. The CT Image is the reconstructed Abeer Saleh Hamdi Bin-Ghodel Department of Information Technology, Yemen Academy for Graduate Studies, Sana'a, Yemen

image and is reconstructed on the basis of X-ray absorption profile. X-rays are electromagnetic waves that are used in diagnosis based on its property that all matters and tissues differ in their ability to absorb X-rays [7,8]. A computer compiles the multi images into complete, cross-sectional pictures (slices) of soft tissue, bone, and blood vessels. A CT Scan obtains images of an organ that cannot be seen on a standard x-ray that results in earlier diagnosis [9].Dense tissues such as the bones appear white on a CT film while soft tissues such as the brain or liver appear gray. The cavities filled with air such as lungs appear black. CT performs better in cases of trauma and emergent situations. It provides better bone detail and has high sensitivity for acute hemorrhage. CT has become an important tool in medical imaging to supplement X-rays, Medical Ultrasonography (USG) and MRI. Although it is still quite expensive, it is the gold standard in the diagnosis of a large number of different disease entities. It is, more recently, Intelligent Optimization Algorithms are developed by simulating or revealing some natural phenomena, and are widely used in many research fields because of their versatility[1]. The PSO Algorithm originated from the study of Bird Predation Behavior, first proposed by Kennedy and Eberhart in 1995. The algorithm is easy to implement and the rules are simple.

The position of each particle represents a potential solution, each particle has a fitness value determined by fitness function, and each particle has three characteristics: position, velocity, and fitness value; a population-based optimization algorithm modeled after the simulation of social behavior of birds in a flock. This algorithm looks for optima by updating generations after initializing with a group of random particles. Every particle's position is adjusted based on its distance from its own personal best position and also the distance from the best particle of swarm throughout the search space [1,10,11].

The PSO (Particle Swarm Optimization) Algorithm has been successfully applied to cancer classification because of its simplicity and generality. However, PSO is easily falls into the local optimal solution [1].

This thesis aims to propose a medical application model that can be expanded to include CT scans of all members of the body. The classification through PSO algorithm is used to doing the data as abnormal or normal. Abnormal refers to cancerous nodule. By the process used, the complexity is reduced in the system and diagnosis confidence is enriched.

2. PROBLEM STATEMENT

Medical Images have become essential in medical diagnosis and treatment. These images play a substantial role in medical applications because the doctors exhibit interest in exploring the internal anatomy. Over the years, medical image processing has contributed a lot in medical applications; for example, the use of image segmentation, image registration, and image guided surgery is so common in medical surgery.

One of the most important problems of cancer is the discovery of late disease, which is not detected early and the diagnosis of the disease, which leads to a low proportion of treatment and the high cost of treatment and Makes death is certain Therefore, early diagnosis and rapid classification of the disease Sure death-saving factor.

3. RELATED WORK

The scope of the use of CT scans in a number of stages and years of research in order to distinguish the edges and the detection of cancer cells or structural calcifications.

In [1], The researchers, used the FCBF (Fast Correlation-Based Feature Selection) method to filter irrelevant and redundant features in order to improve the quality of cancer classification. Then, the classification is perform based on SVM (Support Vector Machine) optimized by PSO (Particle Swarm Optimization) combined with ABC (Artificial Bee Colony) approaches, which is represented as PA-SVM. The proposed PA-SVM method is applied to nine cancer datasets, including five datasets of outcome prediction and a protein dataset of ovarian cancer. By comparison with other classification methods, the findings demonstrate the effectiveness and the robustness of the proposed PA-SVM method in handling various types of data for cancer classification.

In[12], The paper published by discusses the formation of Lung Cancer Detection System by using the techniques of Image Processing. The system formed can take any type of medical image within the three choices consisting of CT, MRI and Ultrasound Images. Here the proposed model is developed using PSO, Genetic Optimization and Support Vector Machine(SVM) algorithm used for feature selection and classification. This paper is an extension of image processing using lung cancer detection and shows the findings of feature extraction and feature selection after segmentation. The system formed accepts any one of medical images within the three choices consisting of MRI, CT and Ultrasound Image as input. After preprocessing of image, canny filter is used for Edge Detection. The present thesis proposes a method to detect the cancerous cells effectively from the CT Scan, MRI and Ultrasound Images. Super Pixel Segmentation has been used for segmentation and Gabor Filter is used for de-noising the medical images. Comparison is done between the three medical images.

In[13], study the researcher discussed Major issues in the implementation of screening for lung cancer by means of lowdose computed tomography (CT) are the definition of a positive result and the management of lung nodules detected on the scans. a population-based is conducted prospective study to determine factors predicting the probability that lung nodules detected on the first screening low-dose CT scans are malignant or will be found to be malignant on follow-up. Predictive tools based on patient and nodule characteristics can be used to accurately estimate the probability that lung nodules detected on baseline screening low-dose CT scans are malignant. he analyzed data from two cohorts of participants undergoing low-dose CT screening. The development data set included participants in the Pan-Canadian Early Detection of Lung Cancer Study (Pan Can). The final outcomes of all nodules of any size that were detected on baseline low-dose CT scans were tracked. Parsimonious and fuller multivariable logistic-regression models were prepared to estimate the probability of lung cancer.

In[14], The algorithm proposed by Nishikawa consisted of four steps. He used in the first step spatial filters on the first two steps in order to improve the advantages of the tumor. The second is to suppress the noise, after which the two images produced from each other to get most of the locations of micro-calcifications after canceling the background of the original image almost. This research did not use a standard database to evaluate algorithm findings.

4. THE PROPOSED MODEL

In this paper, a new Model is proposes called (PCDM) Particle Swarm Optimization Cancer Detection Model, PCDM is specifically aimed cancer detection through image processing used on CT, The present study does not focus on determining the degree or level of cancer. PCDM focuses on method for segmentation of CT images. Correct identification of cancer cell is done by studying the necessary features extracted for the images. Moreover, the filter is have used for noise reduction for the image. Filter is used for edge detection and finally we go for super pixel segmentation. Feature selection and classification. Then the classification through PSO algorithm is used giving the result as abnormal or normal nodule. Abnormal nodule refers to cancerous nodule. By the process used the complexity is reduced in the system and diagnosis confidence is enriched. The proposed model involves the major image modalities that have been studied in this survey of cancer detection through image processing used on CT. Correct identification of cancer cell is done by studying the necessary features extracted for the images, in addition, to detect the validity of this system. We used feature selection as well by the use of PSO. The Proposed Model consists of six main steps. These steps are not isolated and every one of them is linked to some other steps. This is an essential assumption to ensure that these steps are making up the Proposed Model.

Figure (1) shows the PCDM .



Fig 1: The Architecture of PCDM



and figure(2) shows Feature Selection for Model :

Fig 2: Feature Selection in PCDM 5. IMPLEMENTATION AND EVALUATION

To evaluate the performance of the proposed method in improving CT images, the criterion is use for improving Contrast Improvement Index (CII), and use the Peak Signal To Noise Ratio (PSNR) as a quantitative measure to evaluate the performance of images. two images have contain blocks and natural images. The findings show that the ratio of (CII) is high for images that contain a tumor compared with natural images and show that the natural images have better values (PSNR). In this work, the findings of the proposed algorithm were compared to improve CT images, with a study by researchers Rajkumar K.K. and G. Raju [15] who compared the optimization techniques on the Baseline wavelet, the top hatting filtering method and bit plane wavelet in decomposition of the median method CII.

The general methodology of the proposed algorithm for detection of blocks and infected areas includes several steps including:

- 1. Identification of The Affected Area
- 2. Enhancement of the Image.
- 3. Segmentation and Extraction of Characteristics of the Affected Area.

5.1 Identification of the Affected Area

CT images contain areas as diverse as the background of the image, the tissue area, and labels that contain image information. They also show areas around the affected area that do not need and take too long to process unnecessarily. It is assumed that the elements of the image that form the area of interest need to be a subset shared with a group of adjacent neighboring image elements of adequate intensity and consider that the lowest threshold of intensity and the maximum threshold of density are the two thresholds used to determine the appropriate densities.

In order to achieve this, algorithm assumes that the image that will be working threshold has two categories of image elements is a histogram with a binary style, for example, the background and original image that will help to get rid of any mark that existed in the picture and identify the area of the largest area as the affected area. The injury area is then detected.

This happens after converting the image to gray in order to make it easier to handle the pixels of the image and was applied to the application of Matlab.



Fig 3: Histogram of the Original Image



Fig 4: Histograms of the Image after Conversion to Gray

5.2 Enhancement of the Image

After obtaining the affected area the improve the image do not affect the value of the pixels of the image in actual fact, but work on the removal of the asterisk, softening the picture, the collection of pixels similar to the work and filtering it by using filters to determine the edges of the area of interest. These filters gave results and a good speed for the first treatment of the image:



Fig 5: Adjustment



Fig 6: Weiner Filter of Image

5.3 Segmentation and Extraction Characteristics

The use of the histograms to enhance the high frequency components in the resulting image and remove the low frequency in the background structure. A general mechanical threshold value has been applied to the image that has been rebuilt for each CT image and a binary image has been formed that includes all possible points of being (calcification or tumor).

The threshold value is set according to wavelet coefficients, where the maximum values of many wavelet coefficients are calculated for each line forming the image. The image element locations corresponding to wavelet coefficients values that are higher than the specified value are indicated as points of calcification or tumor.

It is necessary to determine the nature of the tissue of the treated area in CT images in order to derive tumor and classification characteristics. The properties depend on histograms of gray pixels from selected regions of the affected area. Therefore, at this stage, the histogram was calculated as the image expressing the distribution of the probabilistic density, and then the histogram was adjusted to extend the low pixels of CT images with narrow histograms. According to theoretical information, regular distribution achieves containing most information. Thus, the redistribution of gray pixels in CT images will achieve a more uniform histogram. By applying the PSO threshold, a general threshold of the image was obtained, the threshold value was automatically deduced from the histogram of the gray pixels in a peak detection method.

The steps of the algorithm were applied in the work environment of Matlab through the following code which was added to the algorithm after calculating the best site or more site clustered around the pixels after the process of improvements to the image, the work of the regular distribution of pixels, work filter for the image, and then the separation and identification of tumor area. It is clarified through Histogram and compared the histogram of the original image. This is the image after applying the algorithm to it:

Input: ProblemSize, Population _{size}
Output: P_{g_best}
Population $\leftarrow \emptyset$
$P_{g_best} \leftarrow \emptyset$
$\mathbf{For}(i = 1 \mathbf{To} Population_{size})$
$P_{velocity \leftarrow \texttt{RandomVelocity}}$
$P_{position \leftarrow RandomPosition}(Population_{size})$
$P_{p.best} \leftarrow P_{position}$
$if(Cost(P_{p_best}) \leq Cost(P_{g_best}))$
$P_{g_best} \leftarrow P_{p_best}$
End
End
While (-StopCondition())
For ($P \in$ Population)
$P_{velocity} \leftarrow UpdateVelocity(P_{velocity}, P_{g_{best}}, P_{p_{best}})$
$P_{position} \leftarrow UpdatePosition(P_{position} P_{velocity})$
$T_{f}(C_{ost}(P_{position}) < C_{ost}(P_{p,best}))$
$P_{n hest} = P_{nosition}$
$P_{-} \to P_{-} \to P_{-$
$If (Cost(* p_{-best}) \ge Cost(* g_{-best}))$
$I_{g_best} \leftarrow I_{p_best}$
End
End
End
End
Return (P_{g_best})

Fig 7:General Formatting PSO Algorithm[16]







Fig 9: (a) The actual location of the tumor in the original image (b) tumor location through the proposed algorithm.

5.4 Criteria for Evaluating the Performance of the Proposed Algorithm

To evaluate the performance of the proposed method in improving CT images, for improving contrast improvement index (CII) use the criterion given by relationship:

$$CII = \frac{C_{enhanced}}{C_{original}} \dots \dots \dots \dots \dots (1)[15]$$

Where (C original and C enhanced) the improved image and the original image were compared, and the image contrast was calculated according to the following relationship:

 $C = (Imax - Imin)/(Imax + Imin) \dots (3)[15]$

(b, r) defines the mean value of gray and forward pixels respectively, and uses the peak signal to noise ratio (PSNR) as a quantitative measure to evaluate the performance of fractional images and defines the following relationship:

PSNR= 10
$$log_{10} \frac{(L-1)^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[F'(i,j)-F(i,j)]^2} (4)[15]$$

Where f(i, j) the original image is the size M*N, f'(i, j) is the image enhanced and the maximum value in the image f(i, j) the small value of PSNR indicates that the image is of poor quality [15].

two criteria on the images are have calculated that contain blocks and natural images. The table below shows the category of selected images, their number and values (CII, PSNR) respectively and the calculation of the rate of each of them:

Table (1) Comparison of Performance Measurements for Different Styles of Images

Table (1) Compare performance measurements for different styles of images					
Pe mea	rformance asurements				
PSNR	СП	Number of pictures	Category Image		
39.6	0.1	1	Natural images		
34.5	0.56	1	Images contain lumps or a tumor		

The findings show that the ratio of (CII) is high for images that contain a tumor compared with natural images, and showed that the natural images show better values (PSNR). The table shows the comparison of the findings of each of the values (PSNR, CII)

In this thesis, the findings of the Proposed Algorithm were compared to improve CT images, with a study by researchers Rajkumar K.K. and G. Raju [15] who compared the optimization techniques on the Baseline wavelet, the top hatting filtering method and bit plane wavelet decomposition the median method CII.

The comparison as shown in Table (2) indicates

 Table (2) Measure the performance of different methods of image optimization with the method proposed using the variance standard CII

Method	Average CII Index
Top Hat	1.02762
Wavelet decomposition (Sure Shrink)	1.16525
Top Hat +Sure Shrink	1.24294
Top Hat + level dependent Wavelet Shrink	1.20331
Top Hat +Visual Shrink	1.18645
Top Hat +level dependent Visual Shrink	1.18835
Top Hat +modified level independent Visual Shrink	1.18517
Top Hat +Bit Plane decomposition	1.06999
Proposed method	1.28649

Table (3) Measurement of the Performance of Different				
Methods of Image Enhancement using	Contrast Criterion			
CII [15]				

Method	Average CII Index
Top Hat	1.02762
Wavelet decomposition (Sure Shrink)	1.16525
Top Hat +Sure Shrink	1.24294
Top Hat + level dependent Wavelet Shrink	1.20331
Top Hat +Visual Shrink	1.18645
Top Hat +level dependent Visual Shrink	1.18835
Top Hat +modified level independent Visual Shrink	1.18517
Top Hat +Bit Plane decomposition(up to 4 th MSB)	1.06999



Fig 10: Measure the performance of different methods of image optimization with the method proposed using the variance standard CII

Since there are problems with processing digital images related to the quality of the digital image itself, in addition to the number of pests in the data structure and the degree of vagueness of the lesions in the data structure, it is difficult to achieve 100% accuracy. However, when you execute the algorithm, in returns errors in optimization in some low-quality images.

6. CONCLUSIONS AND RECOMMENDATIONS

Recent studies have shown that the first option for early detection of cancer is radiography. However, the explanation of the images of CT images largely depends on the opinion of doctors. In this thesis, the PSO algorithm was improved for automated detection of clusters or tumors in digital CT images. The methodology is based on the threshold technique of the algorithm and its application in Matlab environment and on distinction between its characteristics.

In this proposed work, the method of automated detection of CT images obtained from Oncology Center was evaluated. The performance of the algorithm was evaluated by calculating special indicators such as CII to evaluate the

improvement of image contrast and PSNR to evaluate the image's retail performance. Quantitative of the proposed algorithm with findings for another research study. The experimental and quantitative findings demonstrate that the algorithm is capable of helping doctors interpret CT images as well as improving their performance in diagnosis.

7. REFERENCES

- L. Gao, M. Ye, and C. Wu, "Cancer Classification Based on Support Vector Machine Optimized by Particle Swarm Optimization and Artificial Bee Colony," Molecules, vol. 22, p. 2086, 2017.
- [2] T. Saba, S. Al-Zahrani, and A. Rehman, "Expert system for offline clinical guidelines and treatment," Life Sci J, vol. 9, pp. 2639-2658, 2012.
- [3] A. Norouzi, M. S. M. Rahim, A. Altameem, T. Saba, A. E. Rad, A. Rehman, et al., "Medical image segmentation methods, algorithms, and applications," IETE Technical Review, vol. 31, pp. 199-213, 2014.
- [4] R. S. Kapse, S. Salankar, and M. Babar, "Literature survey on detection of brain tumor from MRI images," IOSr Journal of electronics and communication engineering, vol. 10, pp. 80-86, 2015.
- [5] M. Majid, A. Abidin, N. Anuar, K. Kadiran, M. Karis, Z. Yusoff, et al., "A comparative study on the application of binary particle swarm optimization and binary gravitational search algorithm in feature selection for automatic classification of brain tumor MRI," Journal of Fundamental and Applied Sciences, vol. 10, pp. 486-498, 2018.
- [6] A. Kashyap, V. K. Gunjan, A. Kumar, F. Shaik, and A. A. Rao, "Computational and Clinical Approach in Lung Cancer Detection and Analysis," Procedia Computer Science, vol. 89, pp. 528-533, 2016.
- [7] J. L. Prince and J. M. Links, Medical imaging signals and systems: Pearson Prentice Hall Upper Saddle River, NJ, 2006.
- [8] E. E. Nithila and S. Kumar, "Segmentation of lung nodule in CT data using active contour model and Fuzzy

C-mean clustering," Alexandria Engineering Journal, vol. 55, pp. 2583-2588, 2016.

- [9] A. Krishan and D. Mittal, "Detection and classification of liver cancer using CT images," International Journal on Recent Technologies in Mechanical and Electrical Engineering, vol. 2, pp. 93-98, 2015.
- [10] A. Asuntha, N. Singh, and A. Srinivasan, "PSO, Genetic Optimization and SVM Algorithm used for Lung Cancer Detection," Journal of Chemical and Pharmaceutical Research, vol. 8, pp. 351-359, 2016.
- [11] C. Qin, Z. Xue, Q. Feng, and X. Huang, "Selecting Parameters of an Improved Doubly Regularized Support Vector Machine based on Chaotic Particle Swarm Optimization Algorithm," Journal of Universal Computer Science, vol. 23, pp. 603-618, 2017.
- [12] A. Asuntha, N. Singh, and A. Srinivasan, "PSO, Genetic Optimization and SVM Algorithm used for Lung Cancer Detection," Journal of Chemical and Pharmaceutical Research, vol. 8, pp. 351-359, 2016.
- [13] A. McWilliams, M. C. Tammemagi, J. R. Mayo, H. Roberts, G. Liu, K. Soghrati, et al., "Probability of cancer in pulmonary nodules detected on first screening CT," New England Journal of Medicine, vol. 369, pp. 910-919, 2013.
- [14] Nishikawa RM, Ema T, Doi K, Jiang Y, Papaioannou J. 1995 Image feature analysis and computer-aided diagnosis in mammography: reduction of false-positive clustered microcalcifications using local edge-gradient analysis". Med Phys. Feb;22(2):161-9.PMID:7565347.
- [15] R. KK and G. Raju, "Enhancement of mammograms using tophat filtering and wavelet decomposition," Journal of Computer and Mathematical Sciences Vol, vol. 2, pp. 780-898, 2011.
- [16] A. Asuntha, N. Singh, and A. Srinivasan, "PSO, Genetic Optimization and SVM Algorithm used for Lung Cancer Detection," Journal of Chemical and Pharmaceutical Research, vol. 8, pp. 351-359, 2016.