

Feature Selection using Artificial Bee Colony Algorithm for Medical Image Classification

Vartika Agrawal¹, Satish Chandra²

Department of Computer Science & Engineering
Jaypee Institute of Information and Technology
Noida, UP, India

¹vartika5290@gmail.com, ²satish.chandra@jiit.ac.in

Abstract— Feature Selection in medical image processing is a process of selection of relevant features, which are useful in model construction, as it will lead to reduced training times and classification model designed will be easier to interrupt. In this paper a meta-heuristic algorithm Artificial Bee Colony (ABC) has been used for feature selection in Computed Tomography (CT Scan) images of cervical cancer with the objective of detecting whether the data given as input is cancerous or not. Starting with segmentation as a first step, performed by implementing Active Contour Segmentation (ACM) algorithm over the images. In this paper a semi-automated the system has been developed so as to obtain the region of interest (ROI). Further, textural features proposed by Haralick are extracted region of interest. Classification is performed using hybridization of Artificial Bee Colony (ABC) and k- Nearest Neighbors (k-NN) algorithm, ABC and Support Vector Machine (SVM). It is observed that combination of ABC with SVM (Gaussian kernel) performs better than combination of ABC with SVM (Linear Kernel) and ABC with K-NN classifier.

Keywords— *Image Segmentation; Computed Tomography; Gray Level Co-occurrence matrix; Artificial Bee Colony; K-Nearest Neighbors; Support Vector Machine.*

I. INTRODUCTION

Medical image processing is becoming an interdisciplinary research field, attracting experts and researchers from various fields like mathematicians, physicians, engineers, doctors etc. As a result of collaborative efforts by these experts, availability of non-invasive imaging modalities has become possible. In recent time much of work is done for information extraction from images. Primary step to start working with medical images is segmentation [1]. Segmentation of image acts as a fundamental step for analysis and information extraction from various kinds of medical images. A user interactive system has been developed giving an idea about region of interest (ROI) and SNAKE algorithm is used for segmentation [2] [3]. SNAKE algorithm uses the concept of internal and external forces termed as balloon operator for segmenting the region of interest from the sample input.

For image classification after segmentation feature extraction and selection are performed in succeeding steps [4] [5]. High dimensionality of data is biggest hurdle in feature extraction [6]. There are many algorithm available for handling high dimensional data providing reduced set of feature vector [6]. From the reduced set of feature vector

selection of features is done. Feature selection is an essential step because many a times there can be redundant or malignant data, which could lead to degrade the overall performance of the model designed. Most commonly feature selection is done via binarization and all possible combinations are tested, but this procedure is time consuming and possesses high computational cost [7].

Optimization of features and classifiers acts as a crucial step in finding optimal solution. But classical methods have limited scope in practical applications so researchers have worked over advanced optimization techniques like Evolutionary algorithms (Genetic Algorithm, Artificial Bee Colony, etc.), Hill Climbing, which have given promising results. Evolutionary Computing is a collective name for the techniques based on biological evaluation. These algorithms can also be termed as nature mimic algorithms.

This paper is divided in the seven sections. Section II discusses about Image Segmentation. Section III presents brief idea about calculation of features and feature selection procedure. Section IV discusses the steps of ABC algorithm. Section V describes the proposed algorithm. Results are shown in Section VI. Finally, section VII ends with conclusion and future scope.

II. IMAGE SEGMENTATION

For applications like image recognition or compression, using whole image is not necessary as usually there is a need for focusing on particular regions only. Thus, image segmentation is used as a fundamental step in data extraction and analyses of an image serving the basic purpose of division of images into non-overlapping regions. This paper endeavors to extract useful features from images and using those features for further processing. A categorization of image segmentation techniques [8] are mentioned below:

A. Otsu Segmentation

Otsu segmentation techniques is one of the simplest segmentation technique which performs segmentation by calculating threshold value (minimizing weights within class variance) [8]. It operates directly on gray level histogram and assumes histogram to be bimodal.

Algorithm works as follows:

Otsu segmentation starts with calculated histogram and probability at each level, assigning ‘0’ to class probability as initial value. Mean is updated for all possible thresholds ranging from 1 to maximum intensity class probability and mean is updated. Variance within class is minimized.

Variance is the sum of variance of two classes. Threshold considered is average of two threshold values calculated.

For this work Otsu Segmentation is implemented, but it was not able to segment the region of interest (fig. 2 shows the segmented result after applying Otsu segmentation algorithm) so no further work was done with this segmentation technique.



Fig. 1- Original Image



Fig. 2 - Image after Otsu segmentation

B. Active Contour Model (ACM):

Active Contour Model (ACM) popularly known as SNAKES emerged as an important framework in image processing used for object tracking, segmentation etc. ACM was introduced in 1988 by Michael Kass *et al.* [2]. Authors have defined “SNAKE is an energy minimization spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges”. For ACM concept of energy minimization is used as a framework. By the addition of appropriate energy terms it becomes possible to reach close to local minima. SNAKE differs from traditional methods as it gets effected by the presence of corners instead of lines and edges. It does not work upon solving whole of the problem instead dependent over other mechanisms so that it can place contour in the region of interest.

In SNAKE internal forces are responsible for smoothness of contour and external forces for placing the SNAKE near to local minima. In this algorithm for minimizing the energy contours over smoothly over the

surface possessing dynamic nature thus the model is active and named as SNAKE by the authors. [3]

$$E_{\text{snake}}^* = \int_0^1 E_{\text{snake}}(v(s))ds = \int_0^1 (E_{\text{internal}} v(s)) + E_{\text{image}}(v(s)) + E_{\text{image}}(v(s)) + E_{\text{con}}(v(s)))ds \quad (1)$$

$$E_{\text{internal}} = \frac{1}{2}(\alpha(s)|v_s(s)|^2) + \frac{1}{2}(\beta(s)|v_{ss}(s)|^2) \quad (2)$$

In equation (1) E_{snake} is defined over a set of points V_i , where $i = 0, 1, 2, \dots, n-1$. E_{internal} is the internal energy term which controls the deformation present in the image and E_{external} controls the contour fitting onto an image, external energy is the combination of forces possessed by an image (E_{image}) and constrained forces (E_{con}) which arises due to user intervention.

Equation (2) presents the internal energy term, where $\alpha(s)$ and $\beta(s)$ are user defined weights, controlling the amount of curvature and constraints over the snake’s shape. There by, leading to image smoothness. [2]

III. TEXTURAL FEATURES EXTRACTION

Classification is also referred as assignment of physical object into one of the predefined categories. For image classification, extraction of feature is an important step. Features extracted are used in classification tool to obtain the final class of an unknown sample. For successful image classification description of image texture is important as it helps in defining material appearance. Texture and tone are always present in an image, many a one of the property dominates the image. In this paper data set considered is medical images whose have underlying textures vary significantly, texture features are invariant to image variations as well as sensitive to spatial structures. It’s believed that texture features play an important role in visualization system.

Image texture classification is done with the help of Gray Level Co-occurrence matrix (GLCM) introduced by Haralick popularly came to be known as Haralick Features [9] [10], GLCM is the combination of pixel brightness values occurring in an image [11]. A co-occurrence matrix is a symmetric matrix of dimensionality N , N is total number of possible gray levels of a particular image. Co-occurrence matrix is a 2-D histogram estimating probability that a pixel has a specific gray-level while a displaced pixel exhibits another gray-level. It encodes structural information which is useful for derivation of informative data representation in texture classification problems.

$$P_x(i) = \sum_{j=0}^{G-1} P(i, j) \quad (3)$$

Equation (3) is obtained by doing the summation of all the rows GLCM matrix.

$$P_y(i) = \sum_{j=0}^{G-1} P(i, j) \quad (4)$$

Equation (4) is obtained by doing the summation of all the columns GLCM matrix.

$$\mu_x = \sum_{i=0}^{G-1} i P_x(i) \quad (5)$$

Equation (5) represents the mean of P_x obtained from equation (3)

$$\mu_y = \sum_{i=0}^{G-1} j P_y(j) \quad (6)$$

Equation (6) represents the mean of P_y obtained from equation (4)

$$\sigma^2 x = \sum_{i=0}^{G-1} (P_x(i) - \mu_x)^2 \quad (7)$$

Equation (7) represents the variance of P_x obtained from equation (3)

$$\sigma^2 y = \sum_{j=0}^{G-1} (P_y(j) - \mu_y)^2 \quad (8)$$

Equation (8) represents the mean of P_y obtained from equation (4)

$$P_{x+y}(K) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j), \quad |i-j|=k \quad (9)$$

Contrast:

Contrast is a measure of intensity or grey-level variations between the reference pixel and its neighbor. If contrast will increase picture will become more clear that is dark portion will become more dark and light shade will become more light but if it will increase above certain value the colors present in image will get mingled and picture will be unclear.

$$Contrast = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i, j)\} \right\}, \quad |i-j|=n \quad (10)$$

Correlation:

Correlation is a measures of linear dependency of grey levels on neighboring pixels. More is the correlation value of an image it will imply the presence of considerable amount of linear structures.

$$Correlation = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i * j\} * p(i, j) - \{\mu_x * \mu_y\}}{\sigma_x * \sigma_y} \quad (11)$$

Homogeneity:

Homogeneity is a value that measures the closeness of the distribution of elements in the GLCM to the GLCM

diagonal. There will be cluster formation around main diagonal if there is small range of gray levels. If there is good range of homogeneity present in an image then similar pixels will represent an object else each pixel will alone act as an individual. Angular Second Moment (ASM) is the measure of homogeneity [30]. High similarity in neighboring pixel leads to large ASM value.

$$ASM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(i, j))^2 \quad (12)$$

Energy:

Energy helps in detecting various types of textures present in an image. Energy is an approach of using local kernels to detect various types of textures. If a region possess high energy then it means that region is distinguishable.

$$Energy = (ASM)^{0.5} \quad (13)$$

Inverse Difference Moment (IDM):

IDM is also said to be as Local Homogeneity. IDM possess lesser value for inhomogeneous images, and a relatively higher value for homogeneous images.

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} p(i, j) \quad (14)$$

Entropy:

Entropy measures the randomness of intensity distribution in an image or it can also be said as the measure of information content. High entropy value represents that an image do not contain any distinguishable object in it. If entropy measured is low then darker portion is darker and lighter portion will become lighter.

$$Entropy = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i, j) * \log(p(i, j)) \quad (15)$$

Dissimilarity:

Measure of dissimilarity is a metric, which if produces a higher value implies that dependency between the corresponding values in a sequence decreases.

Sum of squares: Variance:

Variance is a measure of heterogeneity, referring to variation in gray-level of pixel pairs. Elements showing the difference from average value of $p(i,j)$ shows high variance value. If an image have more number of edges it will show high variance value in comparison with images having smoother texture.

$$Variance = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - mean)^2 * P(i, j) \quad (16)$$

Sum Entropy:

Entropy sum is useful as it weights the contribution of each pixel, not only statistically but also by taking pixel

entropy into account. Thus entropy sum is helpful in rejection of contribution from empty image regions.

$$SumEntropy = - \sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i)) \quad (17)$$

IV. ARTIFICIAL BEE COLONY ALGORITHM

Artificial Bee Colony algorithm popularly referred as ABC algorithm, defined by Dervis Karaboga in 2005 [12]. ABC is inspired by the intelligent foraging behavior of honey bees. The model of forage selection which head towards the collective intelligence of bees comprises of three components namely food sources, employed bees and unemployed bees.

Food Sources: Food source value gets effected by many factors like reachability to the nest, concentration of energy and how much effort is required to extract this energy.

Employed Bees: Employed bees or employed foragers are associated with any particular food source, which is currently exploited by them. These bees carry information about the food source they are exploiting. Information comprises of distance between the two and the direction of the nest from other bees. All the information is shared with some probability.

Unemployed Bees: These are the bees which are continuously at look-out for the food source which is exploitable. Unemployed bees are classified in two types scout bees and onlooker bees. Scout bees are the bees which search for the new food sources. Onlooker bees are the bees which wait for the employed bees to share the information about the food source.

It is a powerful metaheuristic population based algorithm used by researchers to solve numerical optimization problems. ABC is popularly used as an optimization tool by many researchers in wide range of applications. However there have been few work related to feature selection also. Many comparative works with PSO, GA and other evolutionary computing algorithms over differential algorithm have shown that ABC's performance is better for solving engineering problems in terms of local and global minima [13]. In this work ABC is used for feature selection. In ABC individuals called food positions are modified by the artificial bees. A notable aspect which makes ABC unique is that it takes care of both local and global search. Local search is taken care by employed bees and onlooker bees. Global search is taken care by scout bees, balancing both exploration and exploitation task. Thus ABC over comes the drawbacks of other evolutionary algorithms which usually get stuck in their local minima, giving inappropriate result.

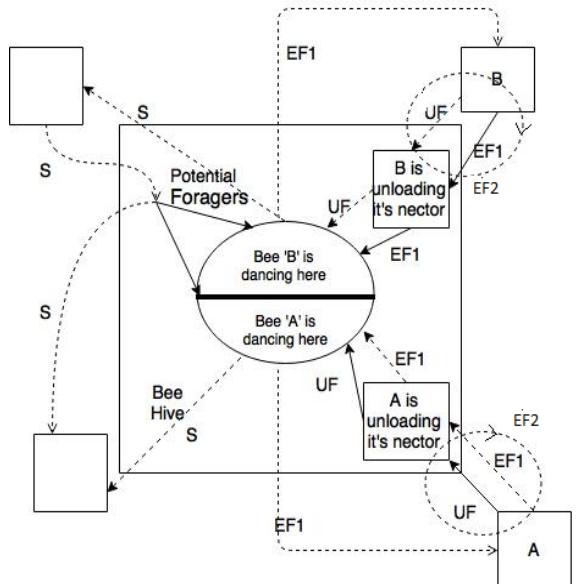


Fig. 3- Representation of behavior of bees [12]

The Fig. 3 above represents the behavior of honey bees foraging for nectar. Considering two bees 'A' and 'B' dancing in their respective areas, bees do not have any knowledge initially it can be scout (S is when it becomes scout) or get itself employed using the information shared by waggle dance (R is food source). After allocation of food source bee will memorize its location and starts exploiting it, hence becoming an 'employed forager'. Employed forager takes nectar from the source and return to its initial place or hive and unloads it. After unloading it can become an uncommitted follower after abandoning the food source (UF), performs waggle dance and recruits other bees (EF1), or continue to forage at the food source (EF2) [12].

Pseudo code for ABC algorithm:

Assumptions:

- There is one artificial bee for one food source.
- Number of employed bees = Number if food sources available.
- Total colony size = Twice the number of employed bees.

Algorithm:

1. Initialization phase- all employed bees are provided with food source.
2. Employed bees move to the food sources allocate and search for the source in neighborhood and the nectar amount is evaluated which is termed as fitness and start doing waggle dance for giving directions to the onlooker bees.
3. Onlooker bees follows the dance and choose food source for themselves and evaluates the nectar or fitness.

4. Scout bees look for the abandoned food sources and replace the existing.
5. Step 2 is repeated until the termination condition is not fulfilled

V. PROPOSED METHOD

For classification of a new image into existing classes, feature value is calculated and then classification is done on the basis of training data. This process is very time consuming as many a times there are some irrelevant features present, which don't play much role in classification. We are using ABC for optimization of features.

Step 1: Representation of the individual:

The individual is composed of: bee or particle position, X , represented by D -dimensional vector namely,
 $X = (x_1, x_2, x_3, \dots, x_D)$
where, D is the number of features extracted from a medical image. Here, $(x_1, x_2, x_3, \dots, x_D)$ represents feature value.

Step 2: Initialize the initial population in the searching space randomly:

The initial population is composed of N individuals; each individual X is D dimensional.

Step 3: Evaluation of fitness values:

Fitness value is evaluated on the basis of classification done k-NN classifier and accuracy is calculated using confusion matrix.

Step 4: If the termination condition is met, then stop and choose the optimum solution. Otherwise, continue.

Termination condition can be desired accuracy to be obtained or maximum number of cycles.

Step 5: The population is updated according to ABC:

Assign new positions to first half bees randomly and chose among the last and new position according to least fitness. Next, assign probabilistically new positions to second half bees in the neighborhood of employee bees. Choose the position with least fitness.

Step 6: Step 5 is executed repeatedly until the termination condition is met. Termination Condition can be the desired accuracy or maximum number of cycles

Step 7: When the algorithm stops, we will get the minimum number of features which can give us maximum possible accuracy.

For better results and hybridization of classification algorithm with meta-heuristic algorithm is done. Feature selection is performed randomly with the help of ABC algorithm and the selected features are given to SVM classifier with performs classification using RBF kernel (Gaussian) [14]. Gaussian kernel is represented as:

$$\exp \frac{(-\|x-x'\|^2)}{2\sigma^2} \quad (18)$$

Where x and x' are feature vector in same input space and σ is a free variable.

Likewise, features selected are given to k-NN classifier and on the basis of those features for $k=3$, k-NN classifier predicts the class of the input sample [15].

VI. EXPERIMENTAL SET-UP AND RESULT

For this work implementation is done in PYTHON language. Data set (CT-Scan images of cervical cancer) are 271 in number with 108 instances as negative and 163 as positive. Images were collected from various diagnostic centres of Delhi and Noida. Starting with image segmentation as a first step user have to select a point on region of interest, then the SNAKE algorithm will start working and region of interest will be highlighted, image obtained will be saved.

Figure below represents the image input and consecutive results:

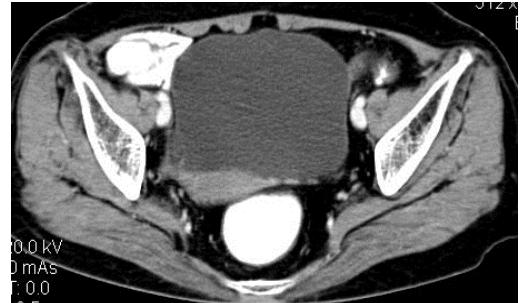


Fig 4: Input image

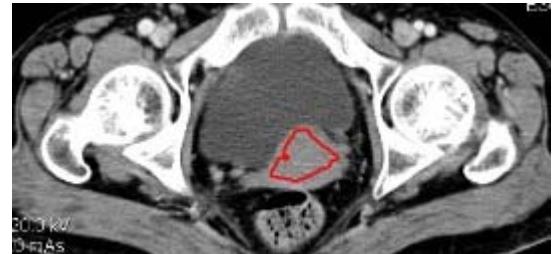


Fig 5: Segmented Image of cervix

Image is cropped and saved again, as finally from the saved cropped image features are calculated.



Fig 6: Cropped Image of region of interest

Features are computed from the cropped image, classification performed and accuracy is calculated using the concept of confusion matrix [16].

A. Result

TABLE I. Result

Algorithm	ABC+k-NN	ABC+SVM (Linear Kernel)	ABC+SVM (Gaussian Kernel)
Accuracy	97%	93%	99%

VII. CONCLUSION

In this paper comparative study of image segmentation and classification over CT images of cervical cancer is presented.

For image segmentation results obtained from Otsu segmentation is not good enough for further processing but ACM has provided satisfying result. So it can be concluded that ACM is appropriate segmentation technique for CT images of cervical cancer.

Furthermore, image classification is performed by selecting features with the help of Artificial Bee Colony algorithm, classifier applied is k-NN and SVM. K-NN is showing 97% accuracy with biased dataset and 100% accuracy with unbiased dataset. While SVM with linear kernel is showing 93% accuracy with biased data set and 99% accuracy when SVM with radial basis as a kernel is applied.

Thus it can be concluded that SVM with radial basis function as kernel is best for classification, but biasness in data set will effect the result.

VIII. REFERENCES

- [1]. Pham, D. L., Xu, C., & Prince, J. L. (2001). A Survey of Current Methods in Medical Image Segmentation, Technical Report JHU/ECE 99-01. Baltimore: Department of Electrical and Computer Engineering, The Johns Hopkins University.
- [2]. Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, 1(4), 321-331.
- [3]. Álvarez, L., Baumela, L., Henríquez, P., & Márquez-Neila, P. (2010, June). Morphological snakes. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 2197-2202). IEEE.
- [4]. Ashburner, J., & Friston, K. J. (2003). Image segmentation. *Human Brain Function*
- [5]. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The 74 Journal of Machine Learning Research*, 3, 1157-1182.
- [6]. Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.
- [7]. Sharma, N., & Aggarwal, L. M. (2010). Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1), 3.
- [8]. Zhao-dong, A. C. J. N., & Zeng-ping, L. Z. J. C. (2010). Otsu Threshold Comparison and SAR Water Segmentation Result Analysis [J]. *Journal of Electronics & Information Technology*, 9, 033.
- [9]. Albertson, F. (2008). Statistical texture measures computed from gray level co-occurrence matrices. *Image processing laboratory, department of informatics, university of oslo*, 1-14.
- [10]. Albrechtsen, F. (2008). Statistical texture measures computed from gray level coocurrence matrices. *Image processing laboratory, department of informatics, university of oslo*, 1-14.
- [11]. Thakare, V. S. (2013). Survey on image texture classification techniques. *International Journal of Advancements in Technology*, 4(1), 97-104.
- [12]. Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization* (Vol. 200). Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.
- [13]. Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. *Applied soft computing*, 8(1), 687-697.
- [14]. Shi, Y. (2012). Comparing K-Nearest Neighbors and Potential Energy Method in classification problem. A case study using KNN applet by EM Mirkes and real life benchmark data sets. *arXiv preprint arXiv:1211.0879*.
- [15]. Suganya, R., & Rajaram, S. (2013, July). Feature extraction and classification of ultrasound liver images using haralick texture-primitive features: Application of SVM classifier. In *Recent Trends in Information Technology (ICRTIT), 2013 International Conference on* (pp. 596-602). IEEE.
- [16]. Powers, D. M. (2013, April). Characteristics and heuristics of human intelligence. In *Computational Intelligence for Human-like Intelligence (CIHLI), 2013 IEEE Symposium on* (pp. 100-107). IEEE.