

Computational Intelligence Optimization Algorithm Based on Meta-heuristic Social-Spider: Case Study on CT Liver Tumor Diagnosis

Mohamed Abu ElSoud
Faculty of Computers and Information,
Mansoura University, Egypt,
Tabouk University, Saudi Arabia.

Ahmed M. Anter
Faculty of Computers and Information,
Mansoura University, Egypt,
Benisuef University, Egypt
Scientific Research Group in Egypt (SRGE)

Abstract—Feature selection is an importance step in classification phase and directly affects the classification performance. Feature selection algorithm explores the data to eliminate noisy, redundant, irrelevant data, and optimize the classification performance. This paper addresses a new subset feature selection performed by a new Social Spider Optimizer algorithm (SSOA) to find optimal regions of the complex search space through the interaction of individuals in the population. SSOA is a new natural meta-heuristic computation algorithm which mimics the behavior of cooperative social-spiders based on the biological laws of the cooperative colony. Different combinatorial set of feature extraction is obtained from different methods in order to keep and achieve optimal accuracy. Normalization function is applied to smooth features between [0,1] and decrease gap between features. SSOA based on feature selection and reduction compared with other methods over CT liver tumor dataset, the proposed approach proves better performance in both feature size reduction and classification accuracy. Improvements are observed consistently among 4 classification methods. A theoretical analysis that models the number of correctly classified data is proposed using Confusion Matrix, Precision, Recall, and Accuracy. The achieved accuracy is 99.27%, precision is 99.37%, and recall is 99.19%. The results show that, the mechanism of SSOA provides very good exploration, exploitation and local minima avoidance.

Keywords—Liver; CT; Social-Spider Optimization; Meta-heuristics; Support Vector Machine; Random Selection Features; Classification; Sequential Forward Floating Search; Optimization.

I. INTRODUCTION

Liver cancer is a serious disease and it is the third commonest cancer followed by stomach and lung cancer [1]. As reported in [2], liver cancer in Egypt is continues to be the second highest cause of cancer incidence and mortality. The most effective way to reduce deaths due to liver cancer is to treat the disease in the early stages. Early treatment requires early diagnosis based on an accurate and reliable diagnostic procedure. One of the most common and robust imaging techniques for the detection of hepatic lesions is CT. The classification of benign and malignant patterns in CT is one of the most significant processes during the diagnosis of liver cancer. Computer aided liver diagnosis (CAD) is a technique that can help radiologists to accurately identify abnormality and help in reducing the risk of liver surgery [3].

The high volume of CT liver tumor data requires some helpful classification approaches to support the analysis of this data. Discriminate analysis is now widely used to distinguishing between normal and abnormal tumor tissues [4]. A critical issue in discriminate analysis is feature selection, instead of using all available features in the dataset, chooses only subset of features to be used in discriminate system [5], [6].

The feature selection stage is one of the important components in any classification system. The performance of a classifier depends directly on the choice of feature selection. The feature selection stage is designed to obtain a compact, relevant, non-redundant and meaningful representation of the data. These selected features are used by the classifier to classify features. It is assumed that, a classifiers that uses smaller and relevant features need less memory, computation speed, and prediction accuracy increased which is desirable for any real-time system [7]. When selecting a small subset of features, their biological relationship with the target tumor is more easily identified. Selecting an effective and more representative feature set is the objective of this paper.

Generally, feature reduction techniques can be divided into two main approaches: filter approach and wrapper approach. Filter approaches are not dependent on machine learning techniques and they are computationally inexpensive and it is more common than wrapper approach. Wrapper approach contains a machine learning techniques as part of the evaluation function, when combine filter approach with wrapper approach usually gives best results [8], [9].

The main challenging problem in feature selection and reduction is the huge search space. The size of the search space increase respect to the available number of features in the dataset. Thus, an exhaustive search is impossible in most cases. Many different search methods have been used to feature selection such as mutual information (MI) [10], document frequency (DF) [11], information gain (IG) [12], Relief [11], [12], Sequential Backward Search (SBS) [12], Sequential Forward Search (SFS) [13], Sequential Forward Floating Search (SFFS) [13], and Random Selection Features (RSFS) [13], almost all these techniques are suffer from the stuck in local minima and computationally expensive. In order to further improve effect of feature selection global search algorithm is needed, many researches try to add intelligent optimization

algorithms into feature selection method, such as improved Gray wolf optimization algorithm [14] and genetic algorithms [15]. The current research work is focused on the determination of an optimal subset feature selection from CT liver tumor dataset using a new intelligence swarm model called social spider optimization algorithm (SSOA) in order to improve the diagnosis accuracy. The choice is a trade-off between computational time and quality of the generated feature subset solutions.

Swarm intelligence is a research field that models the collective behavior in swarms of insects or animals. Several algorithms inspired from the insects and animals behavior to solve a wide range of complex optimization problems [15]. The SSOA algorithm is based on the simulation of cooperative behavior of social-spiders is proposed to optimize our problem. In a social-spider colony, each member depending on its gender and executes a variety of tasks such as ferocity, mating, web design, and social interaction. The communal web is important part of the colony because it is a communication channel among them [16], [17], [18].

SSOA differ to other Evolutionary algorithms (EA). SSOA has a strong capability to search in the problem space and can efficiently find minimal reductions. This algorithm considers two different search agents (spiders): male and female. Depending on gender, each individual is conducted by a set of different evolutionary operators which mimic different cooperative behaviors within the colony. Depending on gender computational mechanisms are applied to avoid the critical flaws such as the premature convergence and the incorrect exploration-exploitation balance. The individuals who have achieved efficient exploration (female spiders) and individuals that verify extensive exploitation (male spiders) [15].

Though the studies mentioned above have contributed extremely to our understanding of the severity of the liver cancer problem, they are lacking to quantitative system to diagnosis these patients. Therefore, the main objective of this study is to develop computerized image analysis system to assist radiologists in interpretation of liver tumor. Multi-classifiers are used in conducting the liver tumor diagnostic problem. A new feature reduction and subset selection approach is used based on natural meta-heuristic model SSOA. Compared with other feature selection methods. SSOA yields more efficient results than any of the other methods tested in this paper.

The reminder of this paper is ordered as follows. Section II discusses the related work for liver tumor characterization. Details of the proposed swarming SSOA model and Texture feature extraction method based on fractal dimension are presented in Section III. In Section IV, the proposed liver tumor diagnosis approach is presented. Section V shows the dataset used and experimental results with discussion. Finally, Conclusion and future work are discussed in the end of this paper.

II. RELATED WORKS

Computer Aided Diagnosis (CAD) plays a key role in the early detection and diagnosis of liver cancer. CAD system is a set of automatic or semi-automatic tools developed to assist radiologists in the diagnosis of liver tumor. Some of the recent

classification results obtained by other studies for liver disease dataset are presented below:

Gletsos et al. [19], proposed first order statistics, SGLDM, gray level difference method, Laws' texture energy features, and fractal dimension measurements methods to extract features from liver tumors. Feature reduction is applied using Genetic Algorithm. Neural Network is applied for classification. Classification performance achieved 91%. Cavouras et al. [20], calculate twenty textural features from the CT density matrix of 20 hemangiomas (benign) and 36 liver metastases (malignant) and were used to train a multilayer perception neural network classifier and four statistical classifiers are used. The performance achieved 83%.

Chen et al. [21], the neural network is included to classify liver tumors. It is implemented by a modified probabilistic neural network (PNN) [MPNN] in conjunction with feature descriptors which are generated by fractal feature information and the gray-level co-occurrence matrix. NFB feature values, spatial gray level dependence matrices give better performance. It is texture based. 30 patients (20 malignant, 10 benign). Classification rate is 83%.

Mougiakakou et al. [22], proposed for each ROI, five distinct sets of texture features to characterize liver based on FOS, SGLDM, GLDM, TEM, and FDM. The genetic algorithm-based feature selection is applied to reduce features. The fused feature set was obtained after feature selection applied. 97 samples is used (38 healthy and 59 abnormal). Weighted voting scheme for 5 classifiers is used. The best performance achieved is 84%.

Kumar et al. [23], proposed Wavelet and Fast Discrete Curvelet Transform (FDHCC) for feature extraction, and to distinguish between benign and malignant tumors the Feed Forward Neural Network classifier is used. The accuracy achieved for Curvelet Transform is 93.3%, and Wavelet is 88.9%.

Duda et al. [24], proposed approach to texture characterization from dynamic CT scans of the liver. The methods applied to recognizing features from hepatic primary tumors are RLM 8 features, COM 11 features, and entropy of image after filtering it with 14 features Laws filters. Experiments with various sets of texture parameters show that the classification accuracy was greater than 90% using Support Vector Machines.

Kumar et al. [25], Improved his work by apply texture features using Gray-Level first-order statistics (GLFOS), Gray level co-occurrence matrix, Contour let coefficient first-order statistics (CCFOS), Contour let coefficient co- occurrence matrices (CCCMs) and for feature selection applied PCA. The classification accuracy based on PNN to classify liver tumor into HCC and Hemangioma. The results obtained from this CAD system for FOS, GLCM, CCFOS, and CCCM are 79%, 86%, 93%, 94% respectively with total accuracy 88%.

The accuracy obtained from above researches are very low and computationally expensive. Therefore, the intelligent optimizations are needed to increase the efficiency and reduce the computation of the methods used. Some of the recently authors are working to optimize hard problems using Social Spider Optimization in different application such as:

James et al. [26], proposed framework based on the foraging strategy of social spiders. SSO can tackle a wide range of different continuous optimization problems and has the potential to be employed to solve real world problems. A set of 20 benchmark functions were used to evaluate the performance of SSO which cover a large variety of different optimization problem types. SSO compared with some widely computational intelligence. Results indicate the effectiveness and robustness of the proposed algorithm to solve optimal hard problem.

Djemame et al. [27], proposed approach to improve segmentation process based on Social Spider optimization. The spiders seem sensitive to the topology of the image, so it is possible to guide spiders movement with a gradient or a laplacian. Indeed, these measures will provide information on the possible presence of contours. It would be then possible to use spiders in two ways: 1) Gradient would be repellent which would partition a colony of spiders in a region. 2) On the contrary, the gradient could have an attractive effect. In this case, spiders would be used to detect the contours of regions.

Pazhaniraja et al. [28], proposed a novel scheme of Discovering new services using SSO. The services that get scattered in the UDDI registry can be discovered by using SSO technique. The SSO method can be used to retrieve more appropriate service from number of services. The proposed approach used to embed the bio-inspired algorithm (SSO) into the web service. The result achieved dynamic web service response to the service requester.

Lenin et al. [29], proposed an improved spider algorithm (ISA) to solve the optimal reactive power dispatch (ORPD) Problem. The structure is based on the foraging social spiders, which make use of the vibrations spread over the communal web to decide the position of preys. The simulation results demonstrate high quality performance of ISA in solving an optimal reactive power dispatch problem. Results indicate the effectiveness and robustness of the proposed algorithm to solve optimal reactive power dispatch problem.

Computational Intelligence based on Bio-inspired SSOA algorithm is used in this paper to decreases time consuming, extract and select relevant, optimal and few features from a huge number of features which are sufficient. Computations also are reduced while prediction accuracy is increased via effective feature selection.

III. METHODOLOGY

1) *Swarming Model based on Bio-inspired Social-Spider Optimization algorithm (SSOA)*: Social-Spider Optimization algorithm can be defined as population-based and algorithmic search meta-heuristic methods that mimic natural evolution process of social spider colony for brief description and more details in [30], [31], [32], [33].

A majority of the spiders are solitary which means that they spend most of their lives without interacting with others. Among the 35 000 spider species observed and described by scientists, some species are social. These spiders live in groups. Based on these social spiders, social spider optimization algorithm (SSOA) is developed to optimize the problems [30], [31]. There are two fundamental components of a social

spider colony, social members and communal web. The social members are divided into males and females. Each spider in the problem represents the solution. Each attribute of features distributed randomly to these spiders. The number of females N_f is randomly selected within the range of 65% - 90% and calculated by the following equation:

$$N_f = \text{floor}[(0.9 - \text{rand}(0, 1) \cdot 0.25) \cdot N] \quad (1)$$

Where S is a population size, and N is number of spider positions (solution). The population contains of females f_i and males m_i . The number of male spiders N_m is calculated as follows:

$$N_m = N - N_f \quad (2)$$

Generate females and males positions randomly on dimension space. The position for female spider calculated as follow:

Generate females and males positions randomly on dimension space. The position for female spider calculated as follow:

$$f_{i,j}^0 = P_j^{\text{low}} + \text{rand}(0, 1) \cdot (P_j^{\text{high}} - P_j^{\text{low}}) \quad (3)$$

Where $i = 1, 2, \dots, N_f; j = 1, 2, \dots, n$

Where f_i is the female spider position, p^{low} lower initial parameter bound and p^{high} upper initial parameter bound.

The position for male spider m_i is generated randomly as follow:

$$m_{i,j}^0 = P_j^{\text{low}} + \text{rand}(0, 1) \cdot (P_j^{\text{high}} - P_j^{\text{low}}) \quad (4)$$

Where $i = 1, 2, \dots, N_m; j = 1, 2, \dots, n$

The evaluations of females and males spiders are defined and weights assigned to each spider. The weighted function for each spider which represents the solution is calculated as follow:

$$w_i = \frac{J(s_i) - \text{worst}_s}{\text{best}_s - \text{worst}_s} \quad (5)$$

Where $J(s_i)$ is the fitness value obtained of the spider position s_i , the values of worst and bests are the maximum and minimum values of the solution in the population respectively. In SSO, the communal web represents the dimension of search space. The search space of the optimization problem seen as a hyper-dimensional spider web. Each solution within the search space represents a spider position. The weight of each spider represents the fitness value of the solution [32]. The information among the colony members is transmitted through the communal web and encoded as a small vibrations. The vibrations depend on the weight and distance of the spider which has generated them [30]. The information transmitted (vibrations) perceived by the individual i from member j are modeled as follow:

$$V_i b_{i,j} = w_j \cdot e^{-d_{i,j}^2} \quad (6)$$

Where the d_{ij} is the Euclidean distance between the spiders i and j .

In each iterations. Female spider presents an attraction or dislike to other spiders according to their vibrations based on the weight and distance of the spiders. Female spiders start looking for any stronger vibration. If there's someone more attractive, the Euclidean distance is calculated. Then the shortest distance between around spiders are calculated and index for the shortest distance. Then female spider do movement and an attraction based on the strong vibration and distance [30], [31]. If r_m is smaller than a threshold PF , an attraction movement is generated; otherwise, a dislike movement is produced as follows.

$$f_i^{t+1} = \begin{cases} f_i^t + \alpha.V_i b_{ci}.(S_c - f_i^t) + \beta.V_i b_{bi}.(S_b - f_i^t) \\ + \gamma.(rand - 0.5) \text{ with probability } PF \\ f_i^t - \alpha.V_i b_{ci}.(S_c - f_i^t) + \beta.V_i b_{bi}.(S_b - f_i^t) \\ + \gamma.(rand - 0.5) \text{ with probability } 1 - PF \end{cases} \quad (7)$$

Where r_m is random number generated between [0 1], α, β, δ and $rand$ are random numbers between [0, 1], PF threshold =.7 and s_c and s_b represent the nearest member to i that holds a higher weight and the best spider of the entire population.

Male spiders are divided into two classes, dominate and non-dominate male spiders. Dominant male spiders have weight value above the median value of the male population. Non-dominate male have weights under the median value [31]. The position of the male spider can be modeled as follows:

$$m_i^{t+1} = \begin{cases} m_i^t + \alpha.V_i b_{fi}.(S_f - m_i^t) + \delta.(rand - 0.5) \\ \text{if } W_{N_{f+i}}^{m_i} > W_{N_{f+m}} \\ m_i^t - \alpha.V_i b_{fi}.(S_f - m_i^t) + \delta.(rand - 0.5) \\ \text{if } W_{N_{f+i}}^{m_i} < W_{N_{f+m}} \end{cases} \quad (8)$$

Where s_f represents the nearest female spider to the male spider i and W is the median weight of male spider population.

The mating in a social spider colony is performed by the dominant males and the female members. Only the Male spiders above median are mating. When a dominant male m_g spider locates a set of female members within a specific range r (range of mating), it mates and forming a new brood [30]. The mating operation calculated as follow:

$$r = \frac{\sum_{j=1}^n (P_j^{high} - P_j^{low})}{2.n} \quad (9)$$

Where n is the dimension of the problem, and l_j^{high} and l_j^{low} are the upper and lower bounds. Once the new spider is formed, it is compared to the worst spider of the colony. If the new spider is better, the worst spider is replaced by the new one. This process is iterated until get the best weighted for each spider and convergence to optimum solution. All weights above 50% will have value '1' that indicates the particular feature indexed by the position of the '1' is selected. If it is '0', the feature is not selected for evaluation process.

2) Texture Feature Extraction based Fractal Dimension:

Geometric primitives that are self-similar and irregular in nature are termed as fractals. Fractal Geometry was introduced to the world of research in 1982 by Mandelbrot [34]. Liver tumor texture is a combination of repeated patterns with regular/irregular frequency. The tumor structure exhibit similar behavior, it has maximum disparity in intensity texture inside and along boundary which serves as a major problem in its segmentation and classification. Fractal dimension reflects the measure of complexity of a surface and the scaling properties of the fractal i.e. how its structure changes when it is magnified. Thus fractal dimension gives a measure of the irregularity of a structure. In fact, the concept of fractal dimension can be used in a large number of applications, such as shape analysis and image segmentation. Segmentation-based Fractal Texture Analysis (SFTA) algorithm consists of decomposing the image into a set of binary images from which the fractal dimensions are computed to describe segmented texture. In order to decompose the image, a new algorithm two-threshold binary decomposition (TTBD) is proposed [35]. Then SFTA feature vector is constructed as the resulting binary images size, mean gray level and bound arias fractal dimension. The fractal measurements are employed to describe the boundary complexity of objects and structures segmented in the input image using box counting algorithm.

IV. PROPOSED CT LIVER TUMOR DIAGNOSIS APPROACH

The proposed approach consists of five main phases to classify CT liver tumors into benign and malignant as shown in Figure 1.

Feature Extraction phase: in this phase feature extraction methods are used such as Gray Level Co-occurrence Matrix (GLCM) [36], [37], First Order Statistics (FOs) [36], Local Binary Pattern (LBP) [37], SAFTA [35], and Feature Fusion to discriminate between benign and malignant tissues.

Normalization phase: the dataset will preprocessed using a normalization technique to decrease the gap between features and smooth data range between [0,1] to increase the classification rate.

Feature Selection phase: The main purpose of feature reduction is to determine a small set of features from a whole features in the problem. The features extracted have irrelevant, redundant, misleading, and noisy features. Remove these data that affects the prediction and classifiers accuracy can be useful. The proposed SSOA is based on K-nearest neighbor (KNN) as fitness function for selecting the optimal feature set as seen in Figure 2. The principles of social spider optimization is used for the optimal feature selection problem. Eventually, they should converge to optimal solution. The solution represents all possible features. Each feature can be seen as a position represented by male and female spider. The optimal position is the subset with high fitness and highest classification accuracy.

The SSOA makes iterations of exploration using female spider for new regions in the feature space and exploitation

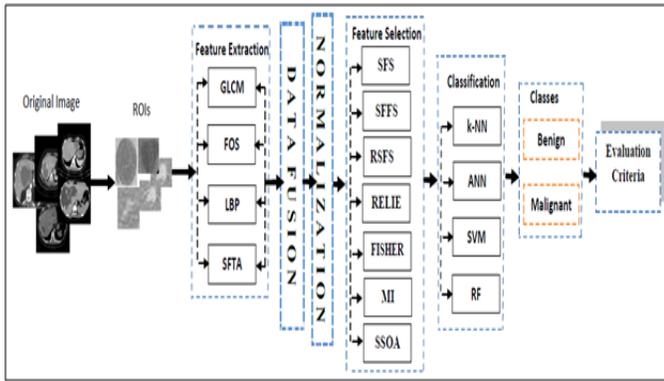


Fig. 1: The proposed liver tumor diagnosis approach.

using male spider until reaching near-optimal solution. To decide if a feature will be selected or not, constant threshold is used. All selected features weights above 50% will have value '1' that indicates the particular feature indexed by the position of the '1' is selected. If it is '0', the feature is not selected for evaluation. The best features with high weight and fitness are selected for evaluation and classification using KNN. The best subset features with high classification accuracy are indexed for classification system. Two fitness functions are used, which are weighted function to measure the weights for each spider in each iteration and changed till reaching the satisfactory solution, and KNN resembling the well-known forward selection. Algorithm 1 shows the steps of the proposed approach for feature subset selection using SSOA and Figure 2, shows the visual representation of the main steps of the proposed system based on bio-inspired SSOA for liver tumor diagnosis.

Classification phase: in this phase the classifiers K-Nearest Neighbor (KNN) [38], Artificial Neural Network (ANN) [39], Support Vector Machine (SVM) [40] and Decision Tree Classifier (DT) [41] are used to classify abnormality into two classes Benign and Malignant tumors. Multi-classifier system are used to obtain high accuracy and to increase the efficiency of our proposed system.

Analysis and evaluation: evaluation criteria for classifiers performance are calculated using confusion matrix, ROC, TP, FP, TN, FN, Precision, Recall, Accuracy, and Over-all accuracy [12], [42].

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the performance of the proposed approach is presented. The simulations were programmed by Matlab 7.9 and performed on Intel (R) Core (TM) i7 CPU 2670QM-2.2 GHz and memory 8GB personal computer and a Microsoft Windows 7.

3) *Data set collection:* CT scanning is a diagnostic imaging procedure that uses X-rays in order to present cross-sectional images ("slices") of the body. The proposed CAD system will be work on difficult dataset. The dataset divided into benign and malignant categories depend on tumor type. The expert

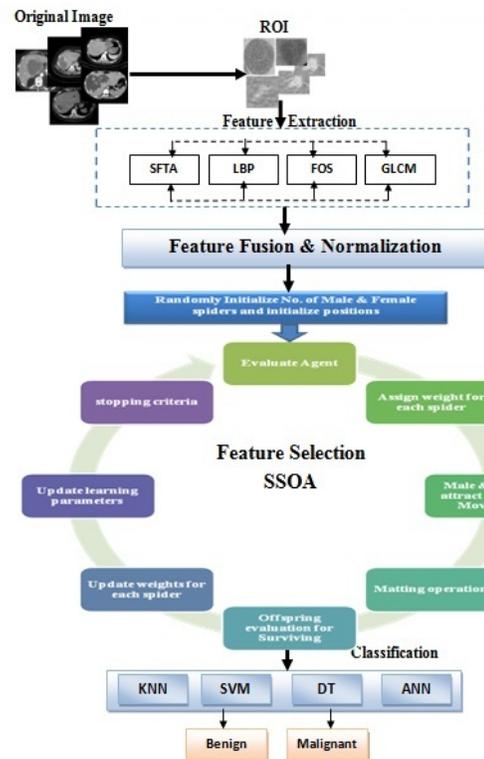


Fig. 2: The steps of the proposed bio-inspired social-spider optimization algorithm for subset feature selection.

physician select from data set 482 region of interests (ROIs) represent benign cases, and 350 ROIs represent malignant cases. Each ROI has dimension size 64×64 pixels [43].

4) *Experimental Results:* In this paper we developed a new approach for liver tumor diagnosis based on meta-heuristic social spider optimizer algorithm to select optimal features with no noise and redundancy.

Firstly, texture features are extracted from each ROIs which represent tumor. The feature extraction methods GLCM, FOs, LBP, and SFTA are proposed. Gray level co-occurrence matrix is constructed to extract feature vector with 68 values which used to represent each ROI. Seventeen features used to represent abnormality Energy, Entropy, homogeneity, Contrast, Dissimilarity, Angular Second Moment (ASM), Correlation, Variance, Maximum Probability (MP), Cluster Tendency, Cluster Shade, Cluster prominence, Sum Average, Sum variance, Sum Entropy, Entropy Difference, and Difference Variance. Four directions (0, 45, 90, and 135) and one distance between pixels equal to 1 are used. LBP is constructed to extract feature vector using mean, variance, skewness and kurtosis, these 4 values used to represent each ROI. FOs is used to extract feature vector with 4 values to represent abnormality (mean, variance, skewness and kurtosis). SFTA is constructed to extract feature vector from liver abnormality with 36 values. The feature vector constructed from SFTA features. The vector corresponds to the number of binary images obtained by TTBD multiplied by three, and from each binary image the following measurements are computed: fractal dimension, mean gray level and size. Finally, the future fusion is applied between

Algorithm 1 Feature subset selection using SSOA.

```
1: Initialize spidn, itern, Dims /* Define spider no., Iter. no.
   & problem dim. */
2: Initialize pflow, pfhigh, Xd, Xu /* Lower & upper Female
   Percent & space dim. */
3: Initialize rm=0, PF=.7 /* Reset Random generator & initial-
   ize proper tuning */
4: Initialize Nf, Nm /* The population of females and males
   */
       Do Equation (2), (3)
5: Initialize fefit, mafit, Fewei, mawei /* Initialize fitness and
   weights Female & males spider */
       /* Population Initialization */
6: Generate female and male positions.
       According to Equations (4), (5)
7: Assign weight for each spider on the colony.
       According to Equation (6)
8: Start iterations for each feature
9: Female and male spiders movement according to the
   vibrations (V) depend on the weight and distance of the
   spiders.
10: Female spider movement and an attraction or dislike occur
   according to stronger vibration
       According to Equations (7), (8)
11: For male spider movement. Select male spiders above
   median
12: Start looking for nearest female with higher vibration and
   shortest distance.
       According to Equations (7), (9)
13: Mating operation: occurs after check whether male spider
   is good or not (above median).
       According to Equation (10)
14: If(Fe is nearest distance) /* Mating occur and produce new
   offspring's*/
15: If(offspring > Worst spider) /*Eval. offspring and calc.
   worst spider*/
16: Delete(Worst spider)
17: Do steps 9-15 again and recalculate the weight for each
   spider and check again the best and worst fitness
18: End For //iteration
19: Calculate the accuracy for each feature from KNN classi-
   fier
```

these features to construct feature vector with 112 values used to represent each ROI. After texture features are extracted. The normalization technique are applied on feature values to enhance, smooth, reduce gap between features and return feature values between [0,1].

Feature subsets are selected with no noise and redundancy and dimensionality reduced using social spider optimization algorithm. A social-spider members maintain a set of complex cooperative behaviors. SSOA can be defined as population-based and algorithmic search meta-heuristic methods that mimic natural evolution process of social spider colony. Each spider in the colony executes a variety of tasks depending on its gender. The fitness's of the spiders are evaluated using a function commonly referred to objective function in Equation

(5). The fitness function reports numerical values which are used in ranking the best spider in the population. The space of the problem solution represents all possible selections of features. Each feature subset values can be seen as a position in this space represented by male and female spider. The numbers of female and male spiders are determined from feature space. In this application number of spiders is equal to number of values in the feature vector which will be 112 spiders. Also the number of iterations are suggested from 5, 10, 50 and 100 iterations to check the SSOA performance and the convergence to optimal solution. Table I shows the initial setting parameters that used in SSOA.

SSOA starts iterations to search for optimal solution, in each one the spiders move to attract or dislike presented in Equation (7) and Equation (8). Each female checked all the spiders and start looking for any stronger vibration presented in Equation (6). These strong vibrations based on the spider characteristics fitness and weight presented in Equation (5). If there's someone more attractive, the Euclidean distance is calculated. Then the shortest distance between the around spiders are calculated and indexed. An attraction is done based on the strong vibration and distance coming from nearest spider. Also repulsion or dislike is done based the gender. The median of spiders is calculated, the male spiders above mean median is start looking for a female with short distance. The spiders below median, go to weighted mean. In mating operation, we check whether male spider is good or not (above median) to generate the offspring's. Only the male spiders above median are mating and the radio (range of mating) is calculated as presented in Equation (10). Then start looking, if there's a good female near. The mating occurs and produces new offspring's. Then the offspring is evaluated and worst spider on the colony calculated. If the fitness of the offspring is better than the worst spider. Then the worst spider removed from the colony. This operation is iterated based on mating and movement occurs till convergence to optimal solution.

After feature values are optimized and the best weights with better fitness are calculated. All weights above threshold 50% will have value '1' that indicates the feature is selected for evaluation and if it is '0' below threshold, the feature is not selected and discard to get the best features for evaluation. The best features with high weights and fitness are selected for evaluation and classification using KNN. The best subset features with high classification accuracy are indexed for classification system. Two fitness functions are used, which are weighted function to measure the weights for each spider in each iteration and changed till reaching the satisfactory solution, and KNN work as forward selection. Forward selection starts with an empty feature set and searches for a feature that achieves the highest classification performance. Then the classification accuracy for the next optimized feature is calculated. Subset Features are selected if achieves the highest improvement in classification accuracy. Then the unused features are removed from features array.

In classification phase the training and testing cases are randomly selected by our algorithm, the number of training cases is 70% and 30% for testing cases. Four classifiers are applied to check the accuracy of the features selection. K-nearest neighbor, Decision Tree, Support Vector Machine, and Artificial neural Network are applied. To prepare KNN we

used $k=1$, in ANN we applied feed-forward back-propagation network with 7 layers in hidden neurons, input neurons depend on the feature sub set selection algorithm, output neurons 2 classes benign and malignant, number of training epochs is 10000. In SVM, we used regularization parameter for weight $\lambda=1$, and we used linear kernel. The DT classifier is much simpler and faster in comparison with the neural network classifier. Each classifier has pros and cons in term of time execution and accuracy for feature vectors as shown in Tables (II-V).

The selection of relevant features and eliminate irrelevant ones is a great problem before classification is applied to train dataset. The redundant and irrelevant features with noise decreases the classification accuracy and makes the computation very complex as shown in Table II.

The normalization approach is applied as shown in Table III, all feature values in the vectors are normalized between [0,1] to decrease gap between values and increase the classifier performance. The visual representations for over-all accuracy, precision, and recall obtained from classification algorithms for non-enhanced dataset compared with normalized dataset are shown in Figure 3.

In this paper an accuracy of classification from different feature reduction methods is applied on abdominal CT liver dataset. In Table IV many different search methods used for feature reduction and selection such as Sequential Forward Search (SFS), Sequential Forward Floating Search (SFFS), Random Feature selection (RFS), Mutual information (MI), Relief, Fisher. Figure 4, shows the precision, recall, and accuracy results of using feature fusion and subset feature selection extracted from classical feature selection methods on the liver tumor dataset.

From the results, all these techniques suffer and stuck in local optima and computationally expensive. For these reasons we improve the effect of feature selection using global intelligent optimization search algorithm SSOA.

SSOA always converge to the optimal or near optimal solution. SSOA model shows a good balance between exploration and exploitation, critical flaws are avoidance such as premature convergence and local minima avoidance. The individuals divided into different search based on gender. The female spiders achieved efficient exploration and male spiders verify extensive exploitation. This assists a meta-heuristic to explore the search space extensively. From the results SSOA can search in the feature space until the optimal solution is converge.

In Table V, shows the results of using SSOA for feature reduction and selection. The number of iterations in SSOA is decreased by normalization phase to 5 iterations, because all features values are smoothed between [0, 1]. Also the time is decreased in all classifiers algorithm, this is because the proposed SSOA algorithm reduced the features and extracted only the optimal ones which are 13 features from 114 features. In Figure 5 we can see that the SSOA gives high accuracy in all classifiers used to diagnosis benign and malignant tumor. SSOA compared with the selection features methods applied in this paper as shown in Figure 6. The results of our proposed approach high, excellent and near to optimum. The achieved accuracy is 99.27%, precision is 99.37%, and recall is 99.19%.

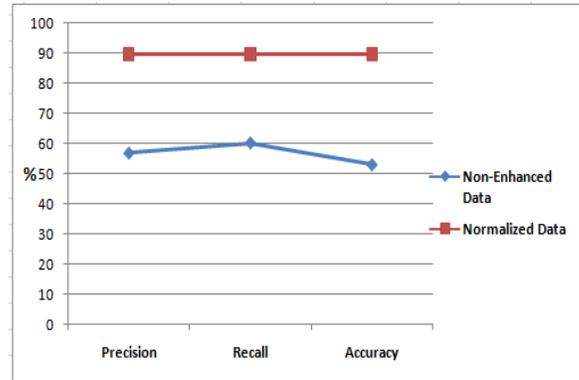


Fig. 3: Results of over-all precision, recall, and accuracy for non-enhanced features (irrelevant, redundant, and noise features) and normalized features.

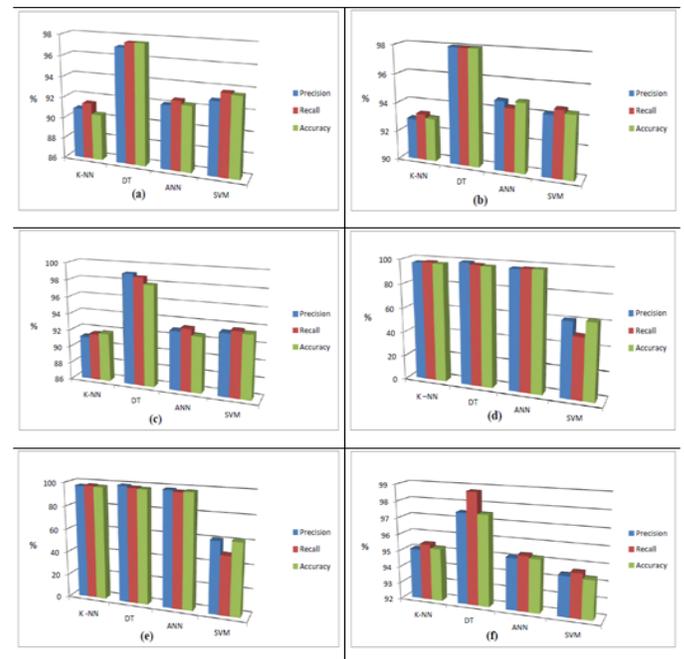


Fig. 4: Accuracy, precision, and recall results using feature fusion and classical feature selection methods on liver tumor dataset ((a) Fisher feature selection method, (b) Relief, (c) Mutual information, (d) SFS, (e) SFFS, (f) RSFS).

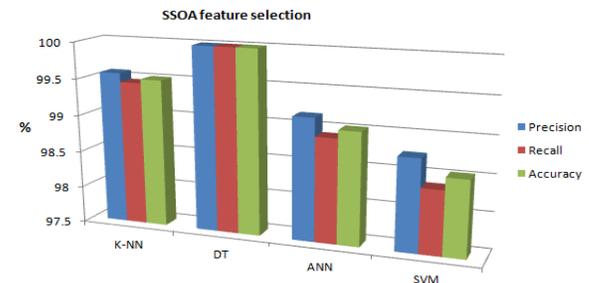


Fig. 5: Results of using the proposed social-spider optimization algorithm to select optimal features.

TABLE I: Initial parameters for Social Spider Optimization Algorithm.

Parameters	Description	Initial value
Spidn	No. of Spider	112
Itern	No. of Iterations	10, 50, 100
X_d	Lower space dim	-800
X_u	Upper space dim	800
p_f^{low}	Lower Female Percent	0.65
p_f^{high}	Upper Female Percent	0.9
Dims	Dimension Space	1
PF	Attraction or dislike	0.7
r_m	Random number	[0,1]
S_d	Search Domain	[0 1]

TABLE II: Classification results using Feature Fusion without normalization.

Classi.	Precision		Recall		Accuracy		Time/s
	B.	M.	B.	M.	B.	M.	
KNN	58.04	47.06	64.84	40	61.25	43.24	0.51
DT	60	52.05	72.66	38	65.72	43.93	9.08
ANN	56.19	50	89.22	50	71.75	60	5.95
SVM	56.70	75	87.22	40	72.16	5.77	0.13

TABLE III: Classification results using Feature Fusion with data normalization.

Classi.	Precision		Recall		Accuracy		Time/s
	B.	M.	B.	M.	B.	M.	
KNN	92.29	86.28	87.84	92	88.87	86.89	2.31
DT	95	91.13	93	91	93	94	1.42
ANN	89.93	84.47	86.18	89	88.02	86.68	9.53
SVM	92.36	86.45	87.75	92	92	89.15	0.23

TABLE IV: Classification results of using fused features with SFS, SFFS, RSFS, Relief, and Mutual Information.

Classi.	Precision		Recall		Accuracy		Time/s
	B.	M.	B.	M.	B.	M.	
*****Fusion*****RSFS*****Subsets(17)*****							
KNN	97	93.15	93.88	97	95.41	95.04	241.31
DT	97	98	97.5	100	97	98	1.14
ANN	96.21	94.06	94.66	96	95.42	95.02	1.95
SVM	97.5	91.34	92.31	97	94.6	94.09	0.29
*****Fusion*****SFS*****Subsets(7)*****							
KNN	97	97	97.5	98	97	97	27.12
DT	98.5	100	98	98	97.5	97	0.943
ANN	97	97	98	97	97	98	2.51
SVM	56.89	65	70	30	72.52	50.83	0.179
*****Fusion*****SFFS*****Subsets(7)*****							
KNN	97	97	97.5	98	97	97	27.23
DT	98.5	100	98	98	98	97	0.946
ANN	96.22	100	98	96	97.61	97.50	2.51
SVM	56.89	65	70	30	72.52	50.83	0.197
*****Fusion*****Relief*****Subsets(38)*****							
KNN	95.37	90.33	91.53	95	93.41	92.61	0.21
DT	98	98	98	98	98	98	1.185
ANN	95.43	94.03	93.66	95	95.04	94.51	4.64
SVM	96.19	92.19	93.09	96	94.62	94.06	0.21
*****Fusion*****Fisher*****Subsets(85)*****							
KNN	95.46	86.18	87.84	95	91.49	89.38	0.534
DT	96.5	97.56	97	98	97	98	1.39
ANN	95.52	88.65	90.19	95	92.78	91.72	6.867
SVM	96.17	89.67	90.97	96.5	93.97	93.19	0.24
*****Fusion*****MI*****Subsets(72)*****							
KNN	93.72	88.57	90.19	93	92.4	91.2	0.48
DT	98	100	100	97.34	97.80	98	1.33
ANN	95.56	90.38	91.75	95	93.62	91.63	4.77
SVM	95.39	91.23	92.31	95	93.83	93.08	0.23

TABLE V: Classification results using fused feature and subset feature selection using SSOA.

Feat. Selec.	Iter.	Fe.No.	Class.	Precision		Recall		Accuracy		Time/s
				B.	M.	B.	M.	B.	M.	
Fusion + SSOA	5	13	KNN	99.15	100	100	98.91	99.57	99.45	0.27
			DT	100	100	100	100	100	100	1.59
			ANN	98.32	100	100	97.83	99.15	98.90	1.49
			SVM	97.5	100	100	96.74	98.73	98.34	0.39



Fig. 6: Results of the proposed SSOA compared with classical feature selection.

A. Discussion

The proposed approach was tested and evaluated on difficult abdominal liver CT data set. The best overall accuracy obtained from Feature Fusion method and KNN classifier is 94%, and the accuracy obtained from other features extraction methods GLCM, LBP, and FD are 85%, 80%, and 87% respectively. To increase the performance of the proposed approach for liver tumor diagnosis. A new subset feature selection algorithm based on Meta-heuristic Bio-inspired Social Spider Optimization algorithm was proposed to select subset of relevant features and eliminate irrelevant ones. The solution space represents all possible selections of features. Each feature subset values can be seen as a position in such a space represented by male and female spider. We used GLCM, FOs, LBP, and SFTA for feature extraction phase. Normalization function is applied to enhance, smooth data and reduce gap between features. SSOA is used to select features with no-noise, no-redundancy and dimensionality reduction. The global intelligent optimization search for subset feature selection SSOA is compared with SFS, SFFS, RFS, MI, Relief, and Fisher. The best optimal features extrated from SSOA with high accuracy and less time consuming are 13 features and 38, 85, 72, 7, 7, 17 from Relief, Fisher, MI, SFS, SFFS, RSFS methods respectively. From results show that, all these techniques suffer from the issues of stuck in local optima and computationally expensive.

SSOA shows a good balance between exploration and exploitation and the results in high local minima avoidance. The female spiders achieved efficient exploration and male spiders verify extensive exploitation. This assists a meta-heuristic to explore the search space extensively. The mechanism of SSOA provides very good exploration, local minima avoidance, and exploitation simultaneously. The proposed approach is high,

TABLE VI: Comparison between the proposed approach and previous works on CT liver tumor diagnosis.

Authors	Year	Dataset	Accuracy
Cavouras et al. [20]	1996	56	83%
Chen et al. [21]	1998	30	83%
Gletsos et al. [19]	2003	147	91%
Bilello et al. [44]	2004	51	80%
Mougiakakou et al. [22]	2007	97	84%
Zhang et al. [45]	2008	44	97.7%
Proposed approach	2015	832	99.27%

excellent and near to optimum solution 99.27%, 99.37%, and 99.19% for accuracy, precision, recall respectively.

Comparing with the other previous works which diagnosis liver abnormality, the works of Cavouras et al. [20] and Chen et al. [21] reached 83% for liver abnormality classification without using feature reduction and selection methods. The results obtained with time computation cost and less accuracy, Gletsos et al. [19] used evolutionary genetic algorithm (GA) for feature reduction and achieved an overall correct classification of 91%. Mougiakakou et al. [22], used genetic algorithm-based feature selection to reduce features and achieved an overall correct classification of 84%. GA gives good results and converge to optimal solution, but has some problems such as premature convergence, crossover, mutation and stuck in local minima. Kumar et al. [25], applied PCA for feature selection and achieved total accuracy 88%. A 99.27% of correct classification and perfect agreement were obtained in our experiments with large dataset as seen in Table VI.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have demonstrated the effectiveness of feature extraction and selection in CT liver tumor classification system. A CAD intelligent system has been built with multi-classifier to achieve high accuracy. The feature fusion extracted from GLCM, FOs, LBP, and SFTA were further subjected to new swarm algorithm called Social Spider Optimization Algorithm (SSOA) to bring the best features with no noise and redundancy for optimal accuracy. SSOA has a strong capability to search in the problem space and can efficiently find minimal reducts. This algorithm considers two different search agents male and female. Depending on gender computational mechanisms are applied to avoid premature convergence and balance between exploration and exploitation. SSOA has selected only 13 features from 112 features to be used in classification phase. The feature selection algorithm SSOA compared with SFS, SFFS, RFS, MI, Relief, and Fisher. From experimental results SSOA prove much better performance, much robustness, and fast convergence speed. With excellent and near optimum solution on abdominal liver CT dataset with accuracy 99.27%, precision 99.37%, and recall 99.19%. In future work, we will apply our new meta-heuristic SSOA algorithm for feature sub set selection on another huge datasets to insure the performance and accuracy.

REFERENCES

- [1] Park, C. J., Cho, E. K., Kwon, Y. H., Park, M. S., Park, J. W. (2005). Automatic separate algorithm of vein and artery for auto-segmentation liver-vessel from abdominal mdct image using morphological filtering. In *Advances in Natural Computation*, Springer Berlin Heidelberg, v.3612, pp.1069-1078.
- [2] Arafa, N., El Hoseiny, M., Rekecewicz, C., Bakr, I., El-Kafrawy, S., El Daly, M., Fontanet, A. (2005). Changing pattern of hepatitis C virus spread in rural areas of Egypt. *Journal of hepatology*, 43(3), 418-424.
- [3] Calle-Alonso, F., Perez, C. J., Arias-Nicolas, J. P., Martin, J. (2013). Computer-aided diagnosis system: A Bayesian hybrid classification method. *Computer methods and programs in biomedicine*, 112(1), 104-113.
- [4] Huang, S. H., Wulsin, L. R., Li, H., Guo, J. (2009). Dimensionality reduction for knowledge discovery in medical claims database: application to antidepressant medication utilization study. *Computer methods and programs in biomedicine*, 93(2), 115-123.
- [5] Costa, A. F., Humpire-Mamani, G., Traina, A. J. M. (2012). An efficient algorithm for fractal analysis of textures. In *Graphics, Patterns and Images (SIBGRAPI)*, 2012 25th SIBGRAPI Conference on IEEE, pp.39-46.
- [6] Ding, C., Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02), 185-205.
- [7] Gunasundari, S., Janakiraman, S. (2013). A study of textural analysis methods for the diagnosis of liver diseases from abdominal computed tomography. *International Journal of Computer Applications*, 74(11), 59-67.
- [8] Kohavi, R., John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1), 273-324.
- [9] Kojadinovic, I., Wotcka, T. (2000). Comparison between a filter and a wrapper approach to variable subset selection in regression problems. In *Proc. European Symposium on Intelligent Techniques (ESIT)*.
- [10] Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8), 1226-1238.
- [11] Lu, Y., Liang, M., Ye, Z., Cao, L. (2015). Improved particle swarm optimization algorithm and its application in text feature selection. *Applied Soft Computing*, 35, 629-636.
- [12] Azar, A. T., Elshazly, H. I., Hassanien, A. E., Elkorany, A. M. (2014). A random forest classifier for lymph diseases. *Computer methods and programs in biomedicine*, 113(2), 465-473.
- [13] Pohjalainen, J., Rsnen, O., Kadioglu, S. (2015). Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Computer Speech and Language*, 29(1), 145-171.
- [14] Emary, E., Zawbaa, H. M., Grosan, C., Hassenian, A. E. (2015). Feature Subset Selection Approach by Gray-Wolf Optimization. In *Afro-European Conference for Industrial Advancement*, Springer International Publishing, pp.1-13.
- [15] James, J. Q., Li, V. O. (2015). A social spider algorithm for global optimization. *Applied Soft Computing*, 30, 614-627.
- [16] Salomon, M., Sponarski, C., Larocque, A., Aviles, L. (2010). Social organization of the colonial spider *Leucauge* sp. in the Neotropics: vertical stratification within colonies. *Journal of Arachnology*, 38(3), 446-451.
- [17] Yip, E. C., Powers, K. S., Aviles, L. (2008). Cooperative capture of large prey solves scaling challenge faced by spider societies. *Proceedings of the National Academy of Sciences*, 105(33), 11818-11822.
- [18] Cuevas, E., Cienfuegos, M., Zaldivar, D., Perez-Cisneros, M. (2013). A swarm optimization algorithm inspired in the behavior of the social-spider. *Expert Systems with Applications*, 40(16), 6374-6384.
- [19] Gletsos, M., Mougiakakou, S. G., Matsopoulos, G. K., Nikita, K. S., Nikita, A. S., Kelekis, D. (2003). A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier. *Information Technology in Biomedicine, IEEE Transactions on*, 7(3), 153-162.
- [20] Cavouras, D., Prassopoulos, P., Karangellis, G., Raissaki, M., Kostariidou, L., Panayiotakis, G. (1996). Application of a neural network and four statistical classifiers in characterizing small focal liver lesions on CT. In *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, vol.3, pp. 1145-1146.
- [21] Chen, E. L., Chung, P. C., Chen, C. L., Tsai, H. M., Chang, C. I.

- (1998). An automatic diagnostic system for CT liver image classification. *Biomedical Engineering*, IEEE Transactions on, 45(6), 783-794.
- [22] Mougiakakou, S. G., Valavanis, I. K., Nikita, A., Nikita, K. S. (2007). Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. *Artificial Intelligence in Medicine*, 41(1), 25-37.
- [23] Kumar, S. S., Moni, R. S. (2010). Diagnosis of liver tumor from CT images using curvelet transform. *International Journal of Computer Science and Engineering*, 2(4), 1173-1178.
- [24] Duda, D. O. R. O. T. A., Kretowski, M., Bezy-Wendling, J. O. H. A. N. N. E. (2006). Texture characterization for hepatic tumor recognition in multiphase CT. *Biocybernetics and Biomedical Engineering*, 26(4), 15.
- [25] Kumar, S. S., Moni, R. S., Rajeeesh, J. (2012). Liver tumor diagnosis by gray level and contourlet coefficients texture analysis. In *Computing, Electronics and Electrical Technologies (ICCEET)*, 2012 International Conference on IEEE, pp.557-562.
- [26] James, J. Q., Li, V. O. (2015). A social spider algorithm for global optimization. *Applied Soft Computing*, 30, 614-627.
- [27] Djemame, S., Nekkache, M., Batouche, M. (2009, December). A Multi-Agent System for Image Segmentation A Bio-Inspired Approach. In *Proceedings of the Third international conference on Innovation and Information and Communication Technology* (pp. 17-17). British Computer Society.
- [28] Pazhaniraja N., Anbusivam P., Raguraman S., Rajeshkanna M. (2014). Service Discovery and Selection using the Bio Inspired Approach. I. *Journal of Engineering And Science (IJES)*, v.3, no.4, pp.42-48.
- [29] Lenin K., Ravindhranath B. (2014). Improved Spider Algorithm for Solving Optimal Reactive Power Dispatch Problem. *International Journal of Recent Research in Interdisciplinary Sciences (IJRRIS)*, vo.1, no.1, pp.35-46.
- [30] Venkatesan, A., Parthiban, L. (2013). Hybridized Algorithms for Medical Image Segmentation. *International Journal of Engineering and Advanced Technology (IJEAT)*, 3(2), pp.305-307.
- [31] Wang, S., Xu, Y., Pang, Y. (2011). A fast underwater optical image segmentation algorithm based on a histogram weighted fuzzy C-means improved by PSO. *Journal of Marine Science and Application*, 10(1), 70-75.
- [32] Deepa, G. (2012). Mammogram Image Segmentation Using Fuzzy Hybrid with Particle Swarm Optimization (PSO). *International Journal of Engineering and Innovative Technology (IJEIT)*, 6(2), 167-171.
- [33] Gopal, N., Karnan, M. (2010). Diagnose Brain Tumor through MRI using Image Processing Clustering Algorithms such as Fuzzy C Means along with Intelligent Optimization Techniques, *International Conference and Computing Research (ICCIC)*, pp.1-4.
- [34] Mandelbrot, B. B. (1983). *The fractal geometry of nature/Revised and enlarged edition*. New York, WH Freeman and Co., pp.1-495.
- [35] Costa, A. F., Humpire-Mamani, G., Traina, A. J. M. (2012). An efficient algorithm for fractal analysis of textures. In *Graphics, Patterns and Images (SIBGRAPI)*, 2012 25th SIBGRAPI Conference on IEEE, pp.39-46.
- [36] Aggarwal, N., Agrawal, R. K. (2012). First and second order statistics features for classification of magnetic resonance brain images, *Journal of Signal and Information Processing*, vol.3, pp.146-153.
- [37] Anter, A. M., El Souod, M. A., Azar, A. T., Hassanien, A. E. (2014). A hybrid approach to diagnosis of hepatic tumors in computed tomography images. *International Journal of Rough Sets and Data Analysis (IJRSDA)*, 1(2), 31-48.
- [38] Kibriya, A. M., Frank, E. (2007). An empirical comparison of exact nearest neighbour algorithms. In *Knowledge Discovery in Databases: PKDD 2007*, Springer Berlin Heidelberg, pp. 140-151.
- [39] Xu, H., Yu, B. (2010). Automatic thesaurus construction for spam filtering using revised back propagation neural network. *Expert Systems with Applications*, 37(1), 18-23.
- [40] Wu, Q., Zhou, D. X. (2006). Analysis of Support Vector Machine Classification. *Journal of Computational Analysis and Applications*, 8(2).
- [41] Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- [42] Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2), 442-451.
- [43] Anter, A. M., Hassanien, A. E., Schaefer, G. (2013). Automatic Segmentation and Classification of Liver Abnormalities Using Fractal Dimension. In *Pattern Recognition (ACPR)*, 2013 2nd IAPR Asian Conference on IEEE, pp.937-941.
- [44] Bilello, M., Gokturk, S. B., Desser, T., Napel, S., Jeffrey Jr, R. B., and Beaulieu, C. F. (2004). Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT. *Medical physics*, 31(9), 2584-2593.
- [45] Zhang, X., Fujita, H., Qin, T., Zhao, J., Kanematsu, M., Hara, T., and Hoshi, H. (2008). CAD on liver using CT and MRI. In *Medical Imaging and Informatics* (pp. 367-376). Springer Berlin Heidelberg.