Biomedical Research 2017; Special Issue: S152-S157 *ISSN 0970-938X*
www.biomedres.info

Breeding firefly association rules for effective medical image retrieval.

Veeramuthu A[1*], Meenakshi S[2]

[1]Department of Information Technology, Sathyabama University, Chennai, India

[2]Department of Information Technology, SRR Engineering College, Chennai, India

Abstract

Multimedia files are increased in size and include affordable memory storage along with the popularity of World Wide Web (www) so that the requirement for effective tools for retrieving images from huge database is critical. For diagnosis, reference, therapy, surgery and training medical images are significant one. A meta-heuristic protocol is built for solving a vast variety of hard optimization issues with no need of deep adaptation to every issue. In this work, an automated Computed Tomography (CT) classification system with novel features selection technique on the basis of FireFly (FF) optimization protocol is suggested for medical images. Wavelets extract the features from medical images and selected CT image features are classified through usage of Naïve Bayes as well as k Nearest Neighbour. The suggested FF protocol attains improved accuracy in comparison to mutual information (MI) as well as the generic FF. Outcomes of simulations reveal that the suggested CFS-FF technique with NB enhanced classification accuracy by 2.83% than MI with NB and with kNN enhanced classification accuracy 2.86% than MI with kNN.

Keywords: Digital medical images, medical image retrieval.

Accepted on January 18, 2017

Introduction

Digital medical images ensure visual information for diagnosis and progress in medical treatment. They X-Rays, MRI, CT-consume much space in medical databases and their retrieval from archives is a challenge. Historically, image retrieval models were text-based with addition from database management as images had to undergo annotation as well as indexing. Increase of image size and image database results in user-based annotation, which are cumbersome, subjective, and incomplete as text does not convey an image rich structure. To offset this, Content Based Image Retrieval (CBIR) research started in early 1990s with retrieval being based on automatic query image feature matching with database image through image-image similarity evaluation [1]. So, images based on visual content like color, texture, shape or other feature or combination of visual features set, are indexed. In an ideal situation, the attributes ought to be incorporated for providing improved discrimination in the comparison. Colour is the most typical visual attribute utilized in CBIR, essentially due to the ease of extraction of colour data from images. For extracting data about shapes as well as textural features are more complicated as well as expensive jobs, typically carried out after the first filter given by the colour attributes [2]. Advances in research are aided by the computer vision community. Multimedia databases are popular for applications like digital libraries, medical images, and news photos. CBIR systems ensure effective and efficient retrieval to access multimedia databases for users querying relevant images through perceptual (or low-level) attributes such as texture, colour, shape or spatial object layouts [3].

A CBIR model has two tasks. The first refers to extracting features wherein features set, known as feature vector, is created and represents an image's database content. The second one is similitude measurement wherein distance between query image as well as database images uses feature vectors for retrieval of "closest" images. CBIR features extraction's two methods are features extraction in spatial as well as transform domains. Features vectors have high dimension. Feature selection is a data mining problem in CBIR. This issue locates adequate as well as discriminatory features subset from a set, for a specific application domain ensuring accuracy including locating a minimum subset representing the entire set, or significance based ranking features from overall set. The advantages of getting features subsets saves unnecessary computing cost and unnecessary features and excludes noisy features and keeps information with "clean" features. Pattern recognition attributes are the observed image's assessable heuristic characteristics. Choosing independent and discriminating features are a key in pattern recognition algorithms efficiency for classification. Bad feature sets lower classifier performance.

Features selection refers to an active as well as encouraging research field in machine learning, patterns recognition, as well as data mining. Common features selection methods include filters, wrappers, as well as embedded techniques. It discards the noise-filled attributes and is an optimum method for

distinguishing between classes. Conventional optimization algorithms offer attractive approaches to find near-optimal solutions to optimization issues [4]. Optimization techniques like Ant Colony Optimizations (ACOs), Particle Swarm Optimizations (PSOs), Bacteria Foraging Algorithm (BFOA) and FF algorithm select subset features. FF is a recently developed SI method that stochastic, bio-inspired, metaheuristic and may be employed for resolving NP-hard issues. Integration of many features leads to dimensionality curse and time in retrieval [5]. The new model includes: 1) extracting features from image database with colour coherence vectors as well as Gabor filter protocol for extracting colour as well as textural features 2) discriminating features through maximal entropy substituting numerical features with nominal ones representing numerical domain intervals and discrete values through usage of class attribute interdependence maximization protocol 3) selecting features through PSO to extract relevant attributes from original set. CBIR based applications are utilized in Internet as well as Law Enforcement to identify and censor images.

A frame work of decision, feature fusion and SVM classifiers for medical images classification was proposed by [6]. Discussions included weighted combination, classifier selection and class-indifferent methods in addition to feature fusion, features selection methods, ranking, as well as features combination. Experiments were done on three benchmark datasets. Results revealed that SVM with polynomial kernel of degree 2 (SVM-P) as well as features fusion achieved optimal classification precision of 90.65%. A new ACO based relevance feedback method with chaos for image retrieval proposed by Pan et al. [7] dynamically reflected users' subjectivity in retrieval results through feature selection. ACO algorithm with chaos assigned weight values to feature vectors images. To avoid algorithm search being trapped in local optimum, chaotic approach found a solution after ants completed some operations. Experiments revealed the proposed image retrieval method efficiency.

A GA-SVM feature selection technique to optimize SVM classification parameters ensured accuracy and saved computation time. Spam assassin dataset validated the new system's performance. Hybrid GA-SVM was an improvement over SVM regarding classification accuracy and computation time. This study developed a new, hybrid genetic Algorithm - SVR (HGA-SVR), for kernel functions as well as kernel parameter value optimization in SVRs, for forecasting maximal electrical daily loads. A new HGA searched for optimum kind of kernel function as well as kernel parameter SVR values for increasing precision. The system was evaluated at electricity load forecasting competition on European Network on Intelligent Technologies (EUNITE) network [8]. Results reveal that HGA-SVR outperformed earlier models and identified optimal kernel function as well as optimum values of SVR variables with low estimation error in electricity load forecasting.

CBMIR using a hybrid method on the basis of optimization of features vectors, feature extraction, and classification of features/similitude metrics proposed by Jaganathan and Vennila [9] was called as Features Optimized Classification Similarity (FOCS) model. Chosen features were textures through usage of Gray level Co-occurrence Matrix (GLCM) as well as Tamura Features (TF), where extracted attributes are turned into a features vector dataset. Fuzzy based PSO (FPSO) reduced features vectors' dimensionality while classification was through Fuzzy based Relevance Vector Machine (FRVM) forming image feature groups classifying dimensionally reduced image feature vectors naturally. Euclidean Distance (ED) was similarity measure to assess similitude between queried as well as target image. FOCS method receives query from user and retrieves images from database. Retrieval protocol performances based on precision as well as recall were estimated. FOCS model has many advantages over current CBMIR. GLCM/TF extracted texture features forming a features vector dataset. Fuzzy-PSO reduced features vector dimensionality problems and chose significant attributes in features vector dataset reducing operational complexities. FRVM utilized for features classification improves response rates as well as retrieval speeds. The new FOCS framework helps physicians get confident in diagnosis decisions while medical research students get images for research.

A novel FF algorithm based method for enhancing the retrieval efficacy of query expansions and concurrently ensuring lower computation complexity was presented by Khennak & Drias [10]. An FF algorithm discovers the optimal expanded query amongst a set of expanded queries (candidate set) and it permits the definition of length of expanded queries in an empirical manner. Outcomes of the simulation on MEDLINE, an online medical database, prove that the suggested method was more efficient in comparison to current best methods. The author increased the performance by optimizing image feature by adopting the FF algorithm [11]. Further, to improve the retrieval accuracy, random walk concepts based on Gaussian distribution was used to move all the fireflies to global best at the end of each iteration. Experimental results show that the proposed method achieved more accuracy and better performance than particle swarm optimization and genetic algorithm, and the use of Gaussian distribution further improved the retrieval accuracy. An automated CT medical image classification method is suggested wherein attributes are extricated via bi-orthogonal spline wavelets. A novel FF algorithm based features selection method is utilized. The chosen features are classified through NB as well as kNN.

Materials and Methods

The patient's condition through digital medical image recordings and clinical care are part of the database for diagnostic, research and educational purposes. With huge amount of digital medical images generated, medical databases are huge and multi-varied. Retrieving required medical image from databases is a challenge. Most of the time, focus is within images relevant to clinical or research query. CBIR retrieving images from databases similar to query images, are now used for medical image retrieval. The major disadvantage is that

every retrieval system handles a specific medical image type like mammography, brain tumour or specific disease. This study addresses retrieval of medical images from a multi-varied database. The study uses Biorthogonal spline wavelets for feature extraction with feature reduction. Features are selected using a new FF based feature selector. Naïve Bayes as well as kNN classify CT images after features selection. Figure 1 shows the flowchart for proposed methodology.

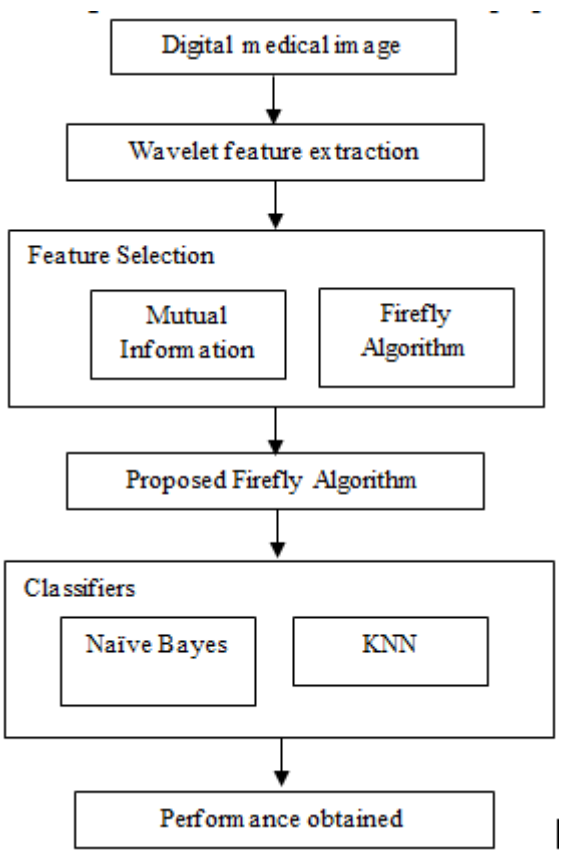

Figure 1. *Flowchart for proposed methodology.*

Feature extraction

Wavelet: Wavelet transforms provide an adequate basis for images handling due to its advantageous characteristics such as its capacity to compact almost all of the signal energy into a minute number of transformation coefficients, which is called energy compaction. The capacity to obtain as well as denote efficiently lesser frequency parts like image background and higher frequency transients like image edges [12].

For image matrix [g(n,m)]N,M, the high-pass filters as well as complementary low-pass filters is employed on image columns as well as to rows followed by downsampling after every processing unit at first. Resultant coefficients may be utilized as attributes.

Feature selection

Mutual information (MI): MI refers to a fundamental information theory notion [13] and a metric of typical interdependence between arbitrary parameters. In two arbitrary parameters X as well as Y, mutual information I(X; Y) is given by equation (1):

$$I(X; Y)=H(X)+H(Y)-H(X,Y) \rightarrow (1)$$

H( ) is a random variable entropy measuring associated uncertainty. H(X) For a continuous arbitrary parameter X is given by equation (2):

$$H(x) = -\int p(x)\log 2p(x)dx \rightarrow (2)$$

If X represents a discrete arbitrary parameter H(X) may be given by equation (3):

$$H(x) = \sum p(X)\log 2p(X) \rightarrow (3)$$

In both, p(X) denotes marginal probability distribution of arbitrary parameter X.

FireFly (FF) feature selection: Xin-She Yang developed FF algorithm at Cambridge University, on the basis of flashing patterns as well as fireflies behaviour. There are nearly 2000 FF species, and almost all emit short as well as rhythmic flashes. The patterns of the flashes are singular almost always for particular species. Flashes rhythms, rate of flashing and time when flashes are seen together form a pattern which attract males to females. Females respond to individual pattern of males of same species. FF algorithm is based on the FF's swarm intelligence behavior. It is a metaheuristic protocol that owes its inspiration to the FF's social behavior. FF algorithm finds global optimal solution. The FF algorithm focuses on task completion within a minimum make span and flow time to use grid resource efficiently. FF algorithm uses three rules [14]:

1. Fireflies are attracted to another disregarding the sex of the other, because they are unisex.
2. Attractiveness is always in proportion to brightness, and both reduce when the distance rises. For two flashing fireflies, the one less bright moves to brighter FF. If one is not more bright than the other, it moves arbitrarily.
3. FF brightness is determined by objective function landscape.

FF optimization is described as

- FF attracts and is attracted by other Fireflies
- Brighter one attracts the less bright
- Brightness decreases with distance
- Brightest FF moves randomly
- FF particles move randomly

Proposed firefly (FF) feature selection: The feature selection algorithm [15] incorporates FF algorithms behavior to improve features selection. The protocol begins with n fireflies x_i, i=1,2, ….,n relating to all features in features set C. Intensity Ii of every FF xi is initialized. A FF *i* moves to its best mating partner j possessing minimal distance with i. If a FF is unable to find best mating partner, the FF i intensity is absorbed by system and is invisible to other fireflies in space. All subsets are tested for a terminating condition. The same procedure is followed by the algorithm for new FF groups, generated in

earlier iterations determining intensity I_{ij} of every group x_{ij} till a stopping criterion is satisfied.

Three rules are combined with the features of levy flight and are formulated a novel L´evy-flight FF algorithm (LFF). The real format of attractiveness function β(r) may be a monotonically decreasing function like the one given by equation (5):

$$\beta(r) = \beta_0 e^{-\gamma r^m}, \quad (m \geq 1) \rightarrow (5)$$

For a particular γ, the characteristic length is Γ=γ-1/m → 1 as m → ∞. With the terms of the relation reversed, for a particular length scale Γ in an optimization issue, variable γ may be utilized as a general initial value. Which is $\gamma = \frac{1}{\Gamma^m}$.

Distance between two fireflies i - j at x_i and x_j, correspondingly, refers to the Cartesian distance by equation (6):

$$r_{ij} = \left|\left| x_i - x_j \right|\right| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \rightarrow (6)$$

Wherein $x_{i,k}$ refers to k^{th} component of spatial coordinate x_i of i^{th} firefly. For other applications like scheduling, distance may be time delay or other adequate form.

Classifiers

The primary notion of classifiers is its usage for segmenting training data and later choosing optimal classifiers for classifying every segment.

Naïve Bayes classifier

Naïve Bayes refers to a probability model based supervised learning classifier. It is a flexible way to handle many attributes and classes, and based on probability theory. Bayes rule for supervised learning is given for unknown target functions by *f: X→Y, or P(Y|X)* If Y is assumed as a boolean-valued arbitrary parameter, then X is a vector comprising 'n' boolean features. Otherwise put as in equation (7):

$X=(X_1,X_2,\ldots,X_n) \rightarrow (7)$

wherein X_i refers to boolean arbitrary parameter representing i^{th} feature of X. Employing Bayes rule, P(Y = yi|X) is in equation (8):

$$P(Y = y_i \left| X = x_k) = \frac{P(X = x_k \right| Y = y_i) P(Y = y_i)}{\sum_i P(X = x_k \left| Y = y_j) P(Y = y_j)} \rightarrow (8)$$

wherein y_m represents m^{th} potential value for Y, x_k represents k^{th} potential vector value for X, and summation in denominator is over legal values of arbitrary parameter Y. An approach for learning $P(Y_jX)$ uses training data for estimating P(X|Y) as well as P(Y). These are later utilized with Bayes rule for determining $P(Y|X=X_k)$ for fresh sample x_k. Naïve Bayes is an asymptotically fast learning algorithm examining all training input. It performed surprisingly well in various problems despite the model's simplistic nature. Further, limited bad data amounts or "noise," do not affect results much.

k Nearest neighbor (kNN)

kNN classifiers have their basis in analogy learning; through comparison of similar test as well as training tuples. Training tuples are described in attributes and every tuple denotes points in an n-dimensional space so that all training tuples are stashed in a n-dimensional pattern space. kNN classifiers search pattern spaces for k training tuples nearest to the unknown tuple when provided with an unknown. A k training tuple is the k nearest neighbor of unknown tuple. Closeness refers to a distance measure like Euclidean distance, which is the distance between two points, for instance, $X_1=(x_{11},x_{12},\ldots,x_{1n})$ as well as $X_2=(x_{21},x_{22},\ldots,x_{2n})$, and may be given by equation (9):

$$Dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \rightarrow (9)$$

Experimental Setup and Results

Simulations were performed with 150 brain, chest and colon CT scan images. The efficacy of the suggested features selection is compared with CFS and MI. Figure 2 to 6 give the classification accuracy, precision, recall, f measure, as well as specificity correspondingly.

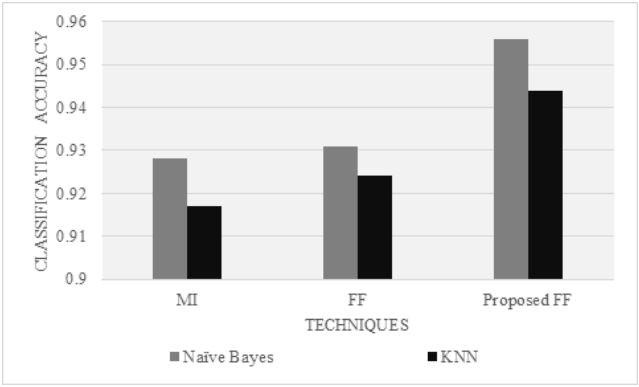

Figure 2. *Classification accuracy.*

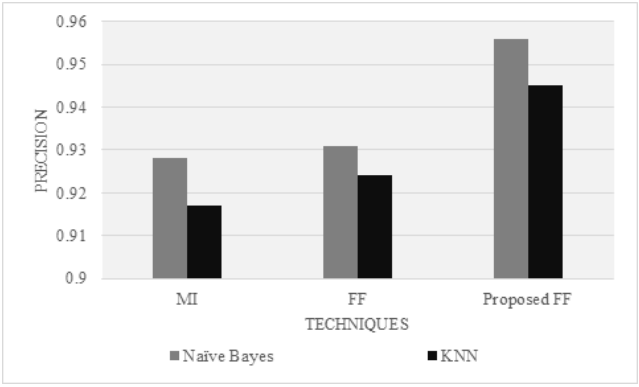

Figure 3. *Precision.*

Figure 2 reveals that the classification accuracy of the suggested CFS-FF technique with NB enhanced classification

accuracy by 2.83% in comparison to MI with NB. The classification accuracy of the suggested CFS-FF technique with kNN enhanced classification accuracy 2.86% in comparison to MI with kNN. Figure 3 gives that the precision of the suggested CFS-FF technique with NB enhanced precision by 2.85% in comparison to MI with NB. The precision of suggested CFS-FF technique with kNN enhanced precision by 2.9% in comparison to MI with kNN.

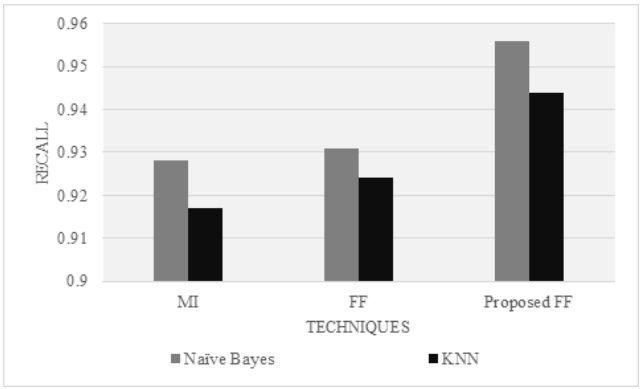

Figure 4. *Recall.*

Figure 4 reveals that the recall of the suggested CFS-FF technique with NB enhanced recall by 2.83% in comparison to MI with NB. The recall of the suggested CFS-FF technique with kNN enhanced recall by 2.86% n comparison to MI with kNN.

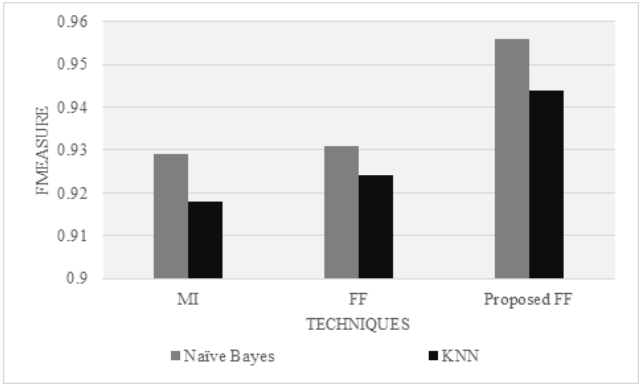

Figure 5. *F Measure.*

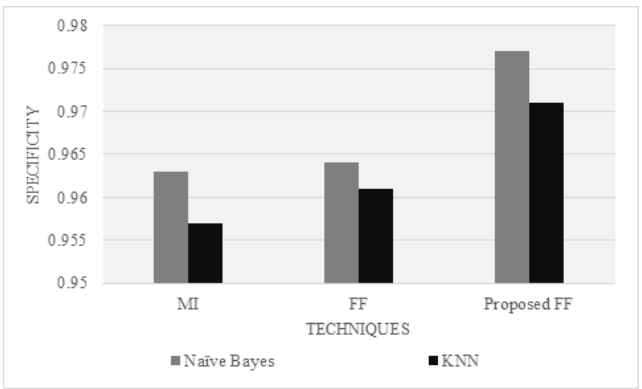

Figure 6. *Specificity.*

Figure 5 reveals that the F measure of the suggested CFS-FF technique with NB enhanced f measure by 2.83% in comparison to MI with NB. F measure of the suggested CFS-FF technique with kNN enhanced f measure by 2.86% in comparison to MI with kNN. Figure 6 reveals that the specificity of the suggested CFS-FF technique with NB improved specificity by 1.46% n comparison to MI with NB. The specificity of the suggested CFS-FF technique with kNN enhanced specificity by 1.48% in comparison to MI with kNN.

Discussion

In the domain of CBIR systems, technology driven techniques perform better than others. A lot of research is being taken up in the domain of CBIR systems according to the users' needs. Modelling complicated human behaviour is a troublesome task; but, the current state of knowledge on the domain is very little. The advantages of CBIR is an active domain of research for 2 major research public, data base organization, as well as computer vision. Features extraction techniques are simple, efficient as well as cheap. Time needs are lesser for finding every associated image. Several associated results happen solely through one search. CBIR disadvantages include semantic gaps, i.e., the lack of coincidences between data through which visual info as well as comprehension of data in a particular situation can be excerpted. Visual contents of pictures, i.e. visual subjects, are colours, size, texture, metadata and so on of images. It is hard to comprehend the information need of users through the query images as there are several objects to search from [16].

The benefits of FF include: FF automatedly splits the population into sub-groups, as local attractions are more strong than farther ones [17]. FF does not utilize historical individual best or global bests. This decreases the possible disadvantages of premature convergences. Furthermore, velocity is not utilized; therefore, issues related to velocity in PSOA is automatedly dropped. FF has the capacity of mobility to control variables like . Therefore, it is seen that FF is more effective in its control of variables, local search capacity, resilience, as well as removal of premature convergences. FF however, has some drawbacks like forced local optima, inability to get rid of local searches, its variables are set and unable to change over time. Moreover, FF algorithm does not remember the past situations of all fireflies and makes them move randomly despite previous better positions [18].

Conclusion

CBIR systems biggest challenge is incorporating adaptable techniques to process images of various characteristics and classes. Factors like illumination variation, image resolution, and occluded objects affect system performance. Low-level features do not coincide with high-level concepts like events or emotions conveyed by an image. Memory and disk space needed to store images and processing is a CBIR systems issue. Feature vectors high dimensionality, results in high computational cost which in turn affect system usability and efficiency. This study presents an automatic classification

system for CT medical images with a new FF optimization algorithm based feature selection method. Experiments proved the new feature selection method's efficiency in classifying multi-varied databases.

References

1. Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J. Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014.
2. Chaudhari R, Patil AM. Content based image retrieval using color and shape features. Int J Adv Res Elect Electron Instrument Eng 2012.
3. Yadav PK, Rizvi S. An exhaustive study on data mining techniques in mining of Multimedia database. In Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on IEEE.
4. Kharrat A, Gasmi K, Messaoud MB, Benamrane N, Abid M. A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine. Leonardo J Sci 2010; 17: 71-82.
5. Bhargavi PK. A novel content based image retrieval model based on the most relevant features using particle swarm optimization. J Global Res Compute Sci 2013; 4: 25-30.
6. Mangai UG, Samanta S, Das S, Chowdhury PR. Survey of decision fusion and feature fusion strategies for pattern classification. IETE Technical Review 2010; 27: 293-307.
7. Pan Z, Chen L, Zhang G. A relevance feedback method based on ant colony algorithm with chaos for image retrieval dependencies. J Comput Informa Syst 2009; 5: 1767-1774.
8. Wu CH, Tzeng GH, Lin RH. A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. Expert Syst Appl 2009; 36: 4725-4735.
9. Jaganathan Y, Vennila I. A hybrid approach based medical image retrieval system using feature optimized classification similarity framework. Am J Appl Sci 2013; 10: 549.
10. Khennak I, Drias H. A Firefly Algorithm-based Approach for Pseudo-Relevance Feedback: Application to Medical Database. J Med Syst 2016; 40: 240.
11. Kanimozhi T, Latha K. An Adaptive Approach for Content Based Image Retrieval Using Gaussian Firefly Algorithm. In International Conference on Intelligent Computing (pp. 213-218). Springer Berlin Heidelberg, 2013.
12. Daubechies CI, Feauveau JC. Bi-orthogonal bases of compactly supported wavelets. Comm Pure Appl Math 1992; 45: 485-560.
13. Battiti R. Using mutual information for selecting features in supervised neural net learning. IEEE Transact Neural Networks 1994; 5: 537-550.
14. Yang XS, He X. Firefly algorithm: recent advances and applications. Int J Swarm Intell 2013; 1: 36-50.
15. Yang XS. Firefly algorithm, Levy flights and global optimization. In Research and development in intelligent systems XXVI. Springer London, 2010.
16. Sukhman Kaur RK. Survey of Content Based Image Retrieval Architecture, Advantages and Disadvantages. Int J Res Elect Comput Eng (IJRECE) 2016; 4: 108-110.
17. Saxena N. Economic Load Dispatch Using Firefly Algorithm. for Doctoral dissertation, Thapar University, Patiala, 2014.
18. Pal SK, Rai CS, Singh AP. Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems. Int J Intell Syst Appl 2012; 4: 50.

*Correspondence to

Veeramuthu A

Department of Information Technology

Sathyabama University

India