

Genetic Algorithm for Feature Selection of MR Brain Images Using Wavelet Co-occurrence

Ahmed Kharrat

University of Sfax National Engineering School
Computer & Embedded Systems Laboratory (CES)
B.P 1173, Sfax 3038, Tunisia
Ahmed.Kharrat@ieee.org

Mohamed Ben Messaoud

University of Sfax National Engineering School
Electronics and Information Technologies
B.P 1173, Sfax 3038, Tunisia
M.BenMessaoud@enis.rnu.tn

Nacéra Benamrane

Department of Computer Science Faculty of Science
Vision and Medical Imagery Laboratory U.S.T.O.
B.P 1505, EL-Mnaouer, Oran, 31000, Algeria
nabenamrane@yahoo.com

Mohamed Abid

University of Sfax National Engineering School
Computer & Embedded Systems Laboratory (CES)
B.P 1173, Sfax 3038, Tunisia
Mohamed.Abid@enis.rnu.tn

Abstract— The selection of features has a considerable impact on the success or failure of classification process. Feature selection refers to the procedure of selecting a subset of informative attributes to build models describing data. The main purpose of feature selection is to reduce the number of features used in classification while maintaining high classification accuracy. A large number of algorithms have been proposed for feature subset selection. Here we compare classical sequential methods with the genetic approach in terms of the number of features, classification accuracy and reduction rate. Genetic Algorithm (GA) achieves an acceptable classification accuracy with only five of the available 44 features. The optimal feature such as mean of contrast, mean of homogeneity, mean of sum average, mean of sum variance and range of autocorrelation provide best classification performance. Similar classification performance is obtained with SFFS and SFBS but with larger feature set.

Keywords— Feature selection; classification accuracy; reduction rate; Genetic Algorithm.

I. INTRODUCTION

Texture is a fundamental property in many natural images. It plays an important role in pattern recognition and computer vision. Also it helps our visual perception system to understand a scene better and to extract more information from it. In the last few decades texture analysis is investigated by research. Texture analysis proves to be important in many fields such as industrial inspection, document segmentation, remote sensing of each resources and medical imaging [1, 2, 3].

Texture classification and texture segmentation arise as two major texture analysis problems. If a textured image is

segmented it doesn't achieve satisfactory results. The failure to get the desired segmentation results for the textured images is explained by the fact that the distance from which we observe and the visual attention that we give to a texture gives us different interpretations of the texture. Many researchers have tried to solve the problem of texture segmentation by using different features selection techniques such as directional grey-level energy and different segmentation methods like region growing.

In texture classification, image textures are categorized into different classes and observed image will be determined to belong to one of the given set of texture classes.

Researchers employed various ways to solve the problem of texture classification such as co-occurrence matrix [4], an analysis of the image spectrum using Wigner distribution and Gabor filter, etc. [5], fractals [6], rotational invariant operators [6], the extraction of the structural and stochastic components of an image [7], model-based method and so on.

Due to the large numbers of features, feature selection procedure seems to be a key to classification problem. This procedure reduces the number of features that need to be collected and provides better classification accuracy. Feature selection refers to algorithms that output a subset of the input feature set since it is unfeasible to measure all possible features and it is unknown in advance which features will provide the best discrimination between classes. More discriminatory features are kept, retaining sufficient information and maintaining an acceptable classification accuracy. By reducing the number of features, feature selection improves classification speed and efficiency. Genetic algorithm appears as a seductive approach to find optimal solution to optimization problem and to offer an alternative to heuristic tree search methods.

Feature selection methods use various criteria such as T-statistics, F-statistics as statistics criteria, and information-theoretic criteria such as mutual information and entropy-

based measure. The diversity of criteria induces substantially different outcomes when applied to same data set.

The goal of this paper is to highlight the value of feature selection in having a considerable effect on the effectiveness of the resulting classification algorithm. Several algorithms are implemented. A taxonomy of feature selection algorithms is presented to show rank of genetic algorithm.

The outline of this paper is organized as follows. In section 2 we will introduce our method based on feature extraction, feature selection and classification. Different feature selection algorithms are presented in the following section. Section 4 we focus our interest on genetic algorithm. Section 5 compares our method using genetic algorithm with other feature selection algorithms with experimental results. Conclusions are presented in the final section.

II. THE PROPOSED APPROACH

We have developed an automated brain tumor classification method that is made up of three phases: features extraction, dimensionality reduction and finally SVM training and testing. The whole process is shown in the diagram of Fig.1. Our whole approach aims at classifying MR brain images into normal, benign tumor and malignant tumor tissue. To achieve this goal, texture features need to be extracted by using Wavelet Transform-Spatial Gray Level Dependence Method (SGLDM) approach [4]. Having as input extracted features the dimensionality reduction is applied. This step is ensured by GA [8]. The output of this phase is a selected set of features used as entries to the classification step which will be done via training and testing techniques. These tools enable us to distinguish between the brain tissue, benign tumor and malignant tumor tissue. Among all classification methods we adopt the Support Vector Machines (SVM) as a classifier. Our implementation of SVM is explained by its robustness, its capacity to handle very large feature spaces and to have good generalization properties compared to conventional classifiers.

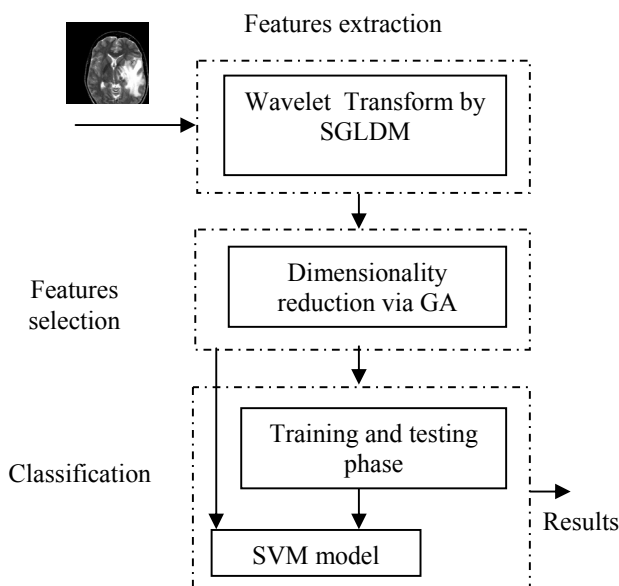


Figure 1. Overview of our approach

III. FEATURE SELECTION ALGORITHMS

A. Selection Problem

Feature selection refers to algorithms that output a subset of the input feature set. Y represents the original set of features and X represents the selected subset that is $X \subseteq Y$. Feature selection criterion are of crucial importance. They divide feature selection methods into two categories: the filter method and the wrapper one [9]. Whereas the wrapper method uses classification accuracy as feature selection criteria; the filter method employs various measurements as shown in Fig.2.

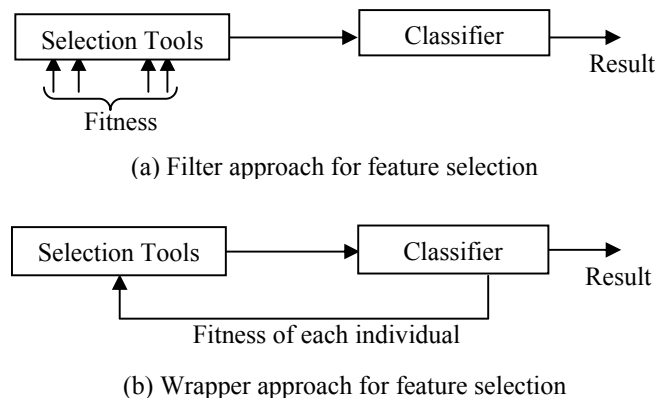


Figure 2. Feature selection methods (a) Filter approach for feature selection (b) Wrapper approach for feature selection

Despite the rapidity of the filter approach, it does not improve the performance of the classification stage. In our paper we use the wrapper method and especially the probability function as a criterion. One possible criterion function is $(1-p_e)$, where p_e denotes the probability of error. The use of probability of error as a criterion function makes feature selection dependent on the specific used classifier and the size of the training and test data sets.

B. Overview of Feature Selection Algorithms

Feature selection algorithms are divided into two main categories: artificial neural networks (ANN), and statistical pattern recognition (SPR) techniques giving the optimal solution or suboptimal feature set. In the suboptimal methods, one can maintain a population of subsets or store a single "current" feature subset and make modifications to it. Also algorithms may be deterministic, producing the same subset on a given problem, or stochastic having a random element to produce different subsets on every run. The Fig.3 shows the tree of some representative feature selection algorithms.

1) *Suboptimal Method*: These methods are not guaranteed to produce the optimal result as they don't examine all possible subsets. They include deterministic,

Single-Solution Methods and deterministic, stochastic Multiple-Solution Methods.

a) *Deterministic, Single-Solution Methods*: They are the most commonly used methods for performing selection. Being referred to as sequential method, deterministic single solution method start with a single solution and iteratively add or remove features until some termination criterion is met. They are split into those that start with the full set and delete features. Kittler [10] compares these algorithms with the optimal branch-and-bound algorithm by applying a synthetic two-class Gaussian data set. Pudil et al [11] modify Kittler's comparative study by introducing sequential floating forward selection (SFFS) and sequential floating backward selection (SFBS).

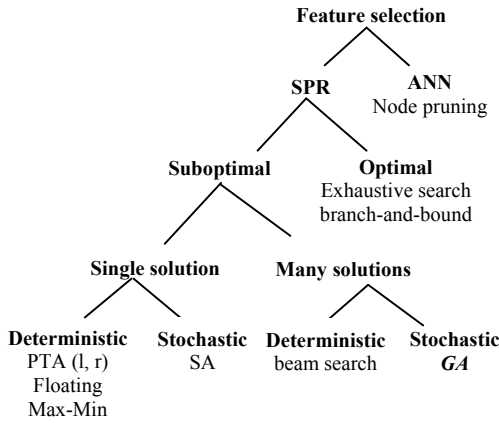


Figure 3. Categories of feature selection algorithms

b) *Deterministic, Multiple-Solution Methods*: They are referred to as “feature selection lattice” since they treat the space of subsets as a graph. Siedlecki and Sklansky [12] have discussed the performance of “beam search” and a best-first search in the space of feature subsets and induced that both methods maintain a queue of possible solutions.

c) *Stochastic, Multiple-Solution Methods*: Among Stochastic, Multiple-Solution Methods, we can use the genetic algorithm. GA is an evolutionary method inspired by the natural process of evolution. It allows a randomised search guided by a certain fitness measure. A feature subset is identified by a particular binary string of length n , with a zero or one in position i denoting the absence or presence of feature i in the set. In each iteration of the algorithm (generation), a fixed number (population) of possible solutions (chromosomes) is generated by means of applying certain “genetic” operators in a stochastic process guided by a fitness measure. Each chromosome is evaluated to determine its “fitness”. New chromosomes are created from old chromosomes by the processes of recombination, crossover and mutation which represent the most important genetic operators. Siedlecki and Sklansky [12] introduced the use of genetic algorithms (GA) for feature selection.

2) *Optimal Methods*: Among optimal method brand-and-bound (BB) feature selection algorithm. Narendra and Fukunaga [13] introduced this algorithm to find the optimal subset of features much more quickly than exhaustive search. Yu and Yuan [14] modified Narendra and Fukunaga's branch and bound algorithm and introduced BAB+. They showed that BAB+ outperforms the original algorithm both analytically and experimentally. Their modification essentially recognizes all “string-structure subtrees”.

IV. GENETIC ALGORITHM

Genetic algorithms are stochastic global adaptive search techniques based on the mechanisms of natural selection. GAs comprise a subset of Darwinian evolution-based optimisation techniques. They are introduced by Siedlecki and Sklansky [12] and later expanded to allow linear feature extraction by Punch et al. [15] and independently by Kelly and Davis [16].

Let d be a set of dimensional input patterns, m a dimensional space and ($m < d$). GA seeks to find a transformed set of patterns in m that maximizes a set of optimization criteria.

A population of competing feature transformation matrices is maintained by GA. The input patterns are multiplied by the matrix in order to evaluate each matrix in this population. The evaluation of matrices produces a set of transformed patterns which are then sent to a classifier. The available samples are divided into a training set, used to train the classifier, and a testing set, used to evaluate classification accuracy. Then GA, being a measure of the quality of the transformation matrix, measures the accuracy obtained. The dimension of the transformed patterns is then reduced increasing the classification accuracy. GA is identified by a particular method of coding the features into string of D bits, $\alpha_1, \dots, \alpha_D$, where $\alpha_i = 1$ if the feature i is in the subset and included in the classifier; or $\alpha_i = 0$ otherwise. Each resulting subset of features is evaluated according to its classification accuracy on a set of test data [17]. Fitness value is allocated based on a performance measure in classification, e.g. the classification accuracy. The fitness function is represented in (1).

$$fitness = W_A \times Accuracy + W_{nb} \times \frac{1}{N} \quad (1)$$

Where W_A is the weight of accuracy and W_{nb} is the weight of N feature participated in classification where $N \neq 0$.

Feature reduction aims at improving classification by searching the best features subset. A reduction uses the fixed set of the original features according to a given processing goal and a feature evaluation criterion (i.e.

classification accuracy). For particular algorithm, accuracy classification is more importance than its execution time.

The effectiveness of the GA in producing desired feature sets was evaluated and compared with several feature selection methods for each dataset. The performance of GA, SFS and SFBS is compared by Ferri et al. [18]. They indicated the superiority of GA performance in very high dimensional problems. Also they showed that the performance of GA decreases as the dimensionality grows.

The performance of SFBS, SFBS, branch and bound algorithm and GA is compared by Siedlecki and Sklansky [12]. They showed the superiority of GA in terms of classification performance and computational effort on a synthetic 24-dimensional data set. Different settings of the feasibility threshold and the tolerance margin are tested by Jain et al [19]. The best recognition rate of 78.9% is obtained by using threshold $t=0.1$ and margin $m=0.05$.

V. EXPERIMENTAL RESULTS

A. Data

In order to test the efficiency of our classification method, we employ the real data base relative to 83 images of seven patients (four men and three women) collected from the Harvard Medical School website [20]. The images consist of 29 images normal, 22 malignant tumors suffering from a low grade Glioma, Meningioma and 32 benign tumors suffering from a Bronchogenic Carcinoma, Glioblastoma multiform, Sarcoma and Grade IV tumors. These normal and pathological benchmark images used for classification, are axial, three weighted images (enhanced T1, proton density (PD) and T2) and acquired at several positions of the transaxial planes. Patients' ages are ranged from 22 to 81 years.

B. Experiments

The data base is divided in two sets. The training set is composed by 12 normal brain images and 20 abnormal images and the remaining slices are used for testing.

To distinguish between benign and malign tumors, 9 benign and 16 malign tumor images are applied in the training phase and the lasting images are employed for testing. The SVM algorithm is implemented for classification in the test phase. Each feature vector, corresponding to a test image, is individually input to the SVM classifier, which produces the corresponding class.

In our approach we have applied Daubechies-2 level 1 wavelet co-occurrence in feature extraction phase [21] which enables us to extract a total of 44 features including the mean and the range. Then we employ SVM classifier especially RBF kernel as an effective option for kernel function. The choice of RBF is explained by the fact that it can classify multi-dimensional data, unlike a linear kernel function and it has fewer parameters to set than a polynomial kernel. Two major RBF parameters applied in SVM, C and γ , must be set appropriately. Parameter C represents the cost of the penalty and parameter γ is the width of the kernel function. The choice of C value influences the classification outcome

and γ value affects the partitioning outcome in the feature space [22]. Hence, the values of $C=8$ and $\gamma=2$, as the best optimised parameters to apply in our implementation. To guarantee valid results for making predictions regarding new data, the dataset is further randomly partitioned into training sets and independent test sets via a k -fold cross validation. Each of the k subsets acted as an independent holdout test set for the model trained with the remaining $k-1$ subsets. The advantages of cross validation are that all of the test sets were independent and the reliability of the results could be improved. The data set is divided into k subsets for cross validation. A typical experiment uses $k=5$. Other values may be used according to the data set size. For a small data set, it may be better to set larger k , because this leaves more examples in the training set. This study used $k=5$, meaning that all of the data will be divided into 5 parts, each of which will take turns at being the testing data set. The other four data parts serve as the training data set for adjusting the model prediction parameters.

C. Results

The efficiency of the classifier is tested for 44 extracted features and is found to be 100% for normal and pathological brain [23]. The classification of benign and malignant brain tumors yields an accuracy rate of 98%.

TABLE I. RESULTS OF FEATURE SELECTION BY SEQUENTIAL SEARCH ALGORITHMS (FORWARD AND BACKWARD) AND GENETIC ALGORITHM

Feature selection	Number of features	Classifier accuracy by normal and pathological brain	Classifier accuracy by benign and malignant brain tumors	Percentage reduction
SFBS	11	100%	98%	75%
SFFS	7	100%	98%	84.09%
GA	7	100%	98%	84.09%
	6	100%	98%	86.36%
	5	100%	98%	88.63%

We perform various feature selection algorithms to evaluate their performances. Table 1 summarized the classification results. In the case of classification of normal and pathological brain, the Sequential Floating Backward Selection (SFBS) method achieves a classification result of 100% with 11 of the available 44 features. This accuracy is similar to that obtained by only 7 features for the Sequential Floating Forward Selection (SFFS) method. In GA-based feature selection method, a population of 30 chromosomes is randomly generated. Each chromosome contains 44 genes (One gene for each feature). The genetic operators, one point crossover and mutation are used. The crossover rate is 90% and mutation rate is 10%. Tournament selection method is used to select the mating pool. Using GA, the selected feature set contains only 5 features to achieve the same classification accuracy of 100%. The feature size is reduced by 88.63%. The classifier accuracies for different feature set sizes for the feature selection methods are illustrated in table 1. This

theoretical result is due to the perfect separation between data in the selected base.

The classification accuracy of 100% is achieved by using only five features: *mean of contrast, mean of homogeneity, mean of sum average, mean of sum variance and range of autocorrelation* [23]. Actually, the features selected by the genetic algorithm, are very related with the appearance of images of the tumors database. By examining the images of abnormal brain we can see that the area of the tumor is characterized by a high degree of brightness. Furthermore, its color distribution is regular. This explains well the selection of the contrast and the auto-correlation features as descriptive characteristics of the tumor. In fact, a contrast is a distinctive characteristic of light distribution of an image or between two points of an image. The auto-correlation can detect regularity and repeated profiles in a signal. In addition, a tumor is an area where the distribution of colors is regular. So that the values are fairly close. Therefore, these aspects also explain the choice of the variance, the homogeneity and the average features as descriptive characteristics of the tumor. In particular, the homogeneity has an opposite behavior of the contrast. The more the texture has homogeneous regions the higher this characteristic is. The variance characterizes the distribution of gray levels around the mean value.

For the identification of benign and malignant brain tumors, the three selection methods achieve the same classification accuracy of 98% although they differ in the size of selected features. GA is the best selection method due to the least number of selected features. In fact GA reduces the feature size by 88.63% without compromising the accuracy of the classifier.

VI. CONCLUSION

In this study we have classified normal, benign and malignant brain tumors using texture analysis for classification. The various feature selection algorithms as SFFS, SFBS and GA are tested. Although, SFFS and SFBS achieve high classification accuracy but they give non-optimal selected features for classification.

The experimental results indicate that SFFS is the best among the Sequential Floating Algorithms. The optimal subset of features established using GA reveal that the classification accuracy can be obtained with only five wavelet co-occurrence features. In this case feature size is reduced by 88.63%. The optimal features are *mean of contrast, mean of homogeneity, mean of sum average, mean of sum variance and range of autocorrelation*. The experiments show that GA is the best feature selection method followed by SFFS and SFBS for the classification of brain tumors in MR images.

ACKNOWLEDGMENT

We would like to thank Mrs Ines Kallel for her helpful review of the manuscript and also be grateful Ridha Mkaouer radiology doctor in the clinic of El Bassetine cité jardin II (Sfax) for their help in interpreting data.

REFERENCES

- [1] R. Haralick, Statistical and structural approaches to texture, *Proc. of IEEE*, 1979, 786-804.
- [2] H. Derin, W.S. Cole, Segmentation of textured images using Gibbs random fields, *Comput. Vision Graphics Image Process*, 1986, 72-98.
- [3] J. Mao, A. K. Jain, Texture classification and segmentation using multiresolution simultaneous autoregressive models, *Proc. of Pattern Recognition*, 1992, 173-188.
- [4] R. Haralick, K. Shanmugam, I. Dinstein, Textural Features for Image Classification, *Proc. of IEEE Trans On Systems Man and Cybernetics*, 1973, 610-621.
- [5] A. BODNAROVA, M. BENNAMOUM, S. LATHAM, Optimal Gabor filters for textile flaw detection, *Proc. of Pattern Recognition*, 2002, 2973-2991.
- [6] D.A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, *Canadian Journal of Remote Sensing*, 28(1), 2002, 45-62.
- [7] T. Kaspas, D. Charalampidis, M. Georgiopoulos, J.P. Rolland, Segmentation of textured images based on fractals and image filtering, *Proc. of Pattern Recognition*, 2001, 1963-1973.
- [8] M. Analoui, M. Fadavi Amiri, Feature Reduction of Nearest Neighbor Classifiers Using Genetic Algorithm, *In Proceedings of the Third International Conference on Computer Vision, Image and Signal Processing*, 2006, 36-39.
- [9] R. Kohavi, H.J. George, Wrappers for feature subset selection, *Artificial Intelligence*, 97(1-2), 1997, 273-324.
- [10] J. Kittler, Feature Set Search Algorithms, *Pattern Recognition and Signal Processing*, 1978, 41-60.
- [11] P. Pudil, J. Novovicova, J. Kittler, Floating Search Methods in Feature Selection, *Pattern Recognition Letters*, 1994, 1119-1125.
- [12] W. Siedlecki, J. Sklank, A note on genetic algorithms for large-scale feature selection, *Pattern Recognition letters*, 1989, 335-347.
- [13] P.M. Narendra, K. Fukunaga, A Branch and Bound Algorithm for Feature Subset Selection, *IEEE Trans. Computers*, 1977, 917-922.
- [14] B. Yu, B. Yuan, A More Efficient Branch and Bound Algorithm for Feature Selection, *Pattern Recognition*, 1993, 883-889.
- [15] W. F. Punch, E. D. Goodman, M. Pei, L. Chiahun, P. Hovland, R. Enbody, Further research on feature selection and classification using genetic algorithms, *in Proc. Int. Conf. Genetic Algorithms*, 1993, 557-564.
- [16] J. D. Kelly, L. Davis, Hybridizing the genetic algorithm and the k nearest neighbors classification algorithm, *in Proc. 4th Int. Conf. Genetic Algorithms Appl.*, 1991, 377-383.
- [17] D.P. Muni, N.R. Pal, J. Das, Genetic programming for simultaneous feature selection and classifier design, *IEEE Transactions on Systems, Man and Cybernetics*, 2006, 106-117.
- [18] F. Ferri, P. Pudil, M. Hatef, J. Kittler, Comparative Study of Techniques for Large Scale Feature Selection, *Pattern Recognition in Practice IV*, E. Gelsema and L. Kanal, eds., 1994, 403-413.
- [19] A. Jain, D. Zongker, Feature Selection: Evaluation, Application, and Small Sample Performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 153-158.
- [20] Harvard Medical School, <http://med.harvard.edu/AANLIB/>
- [21] A. Kharrat, M.B. Messaoud, N. Benamrane, M. Abid, Detection of Brain Tumor in Medical Images. *In: 3rd IEEE International Conference on Signals, Circuits & Systems*, doi: 10.1109/ICSCS.2009.5412577, 2009, 1-6.
- [22] B. Scholkopf, A. J. Smola, Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond, Cambridge, MIT Press, 2001.
- [23] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane, M. Abid, Automated Classification of Magnetic Resonance Brain Images Using Wavelet Genetic Algorithm and Support Vector Machine, *Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10)*, doi: 10.1109/COGINF.2010.5599712, 2010, 369 - 374.