

# A Novel Prostate Cancer Classification Technique Using Intermediate Memory Tabu Search

## Muhammad Atif Tahir

School of Computer Science, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, UK  
Email: a.tahir@qub.ac.uk

## Ahmed Bouridane

School of Computer Science, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, UK  
Email: a.bouridane@qub.ac.uk

## Fatih Kurugollu

School of Computer Science, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, UK  
Email: f.kurugollu@qub.ac.uk

## Abbes Amira

School of Computer Science, Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, UK  
Email: a.amira@qub.ac.uk

Received 31 December 2003; Revised 2 November 2004

The introduction of multispectral imaging in pathology problems such as the identification of prostatic cancer is recent. Unlike conventional RGB color space, it allows the acquisition of a large number of spectral bands within the visible spectrum. This results in a feature vector of size greater than 100. For such a high dimensionality, pattern recognition techniques suffer from the well-known curse of dimensionality problem. The two well-known techniques to solve this problem are feature extraction and feature selection. In this paper, a novel feature selection technique using *tabu search* with an intermediate-term memory is proposed. The cost of a feature subset is measured by leave-one-out correct-classification rate of a nearest-neighbor (1-NN) classifier. The experiments have been carried out on the prostate cancer textured multispectral images and the results have been compared with a reported classical feature extraction technique. The results have indicated a significant boost in the performance both in terms of minimizing features and maximizing classification accuracy.

**Keywords and phrases:** feature selection, dimensionality reduction, tabu search, 1-NN classifier, prostate cancer classification.

## 1. INTRODUCTION

Prostate cancer has become the second most commonly diagnosed cancer in the male population after lung cancer, with approximately 22 800 new cases diagnosed every year in the UK alone. Currently, prostate needle biopsy remains the only conclusive way to make an accurate diagnosis of prostate cancer [1]. Recently Roula et al. have described a novel approach in which additional spectral data is used for the classification of prostate needle biopsies [2, 3]. The aim of this novel approach is to help pathologists reduce the diagnosis error rate. Instead of analyzing conventional grey scale or RGB color images, spectral bands have been used in the analysis. Results have shown that the multispectral image classification outperforms both RGB and grey-level-

based classification. The following four classes have been discriminated.

- (i) Stroma: STR (muscular normal tissue).
- (ii) Benign prostatic hyperplasia: BPH (a benign condition).
- (iii) Prostatic intraepithelial neoplasia: PIN (a precursor state for cancer).
- (iv) Prostatic carcinoma: PCa (abnormal tissue development corresponding to cancer).

Figure 1 shows samples of the four classes.

In their research [2, 3], the total number of features used is greater than 100. For such a high-dimensionality problem, pattern recognition techniques suffer from the well-known curse-of-dimensionality problem [4] as the number

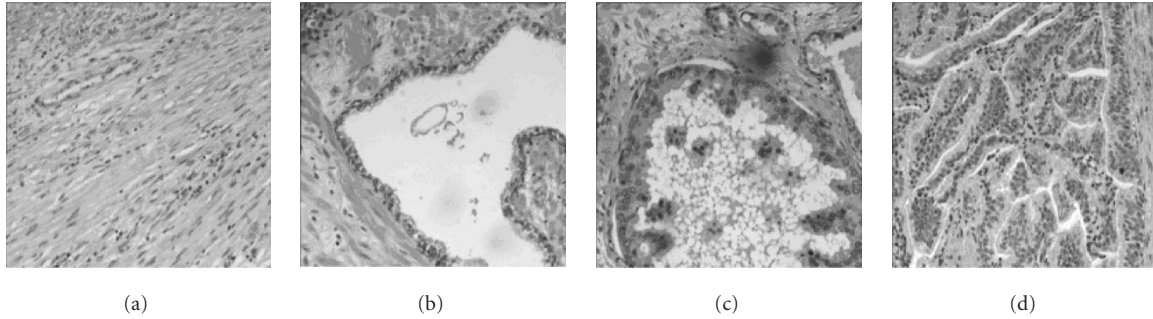


FIGURE 1: Images showing representative samples of the four classes: (a) stroma, (b) BPH, (c) PIN, and (d) cancer.

of training samples that are used to design the classifier is small relative to the number of features. One way to overcome the curse-of-dimensionality problem is to reduce the dimensionality of the feature space. Principal component analysis (a well-known feature extraction method) has been used by Roula et al. on the large resulting feature vectors to reduce its dimensionality to a manageable size. The classification tests have been carried out using supervised classical discrimination method [5]. Although Roula et al. have achieved a classification accuracy of over 94% in their experiments, this is still unacceptable in medical applications. This work attempts to further improve this classification accuracy.

The other way to reduce the dimensionality of the feature space is by using feature selection methods. The term feature selection refers to algorithms that select the best subset of the input feature set. These algorithms used in the design of pattern classifiers have three goals: (1) to reduce the cost of extracting features, (2) to improve the classification accuracy, and (3) to improve the reliability of the estimation of performance [4, 6]. Feature selection leads to savings in the measuring features (since some of the features are discarded) and the selected features retain their original physical interpretation [4]. This feature selection problem can be viewed as a multi-objective optimization problem since it involves minimizing feature subset and maximizing classification accuracy.

Mathematically, the feature selection problem can be formulated as follows. Suppose  $X$  is an original feature vector with the cardinality of  $n$  and  $\bar{X}$  is the new feature vector with the cardinality of  $\bar{n}$ ,  $\bar{X} \subseteq X$ ,  $J(\bar{X})$  is the selection criterion function for the new feature vector  $\bar{X}$ . The goal is to optimize  $J()$ .

This feature selection problem is NP-hard problem. Therefore, the optimal solution cannot be guaranteed to be acquired except when performing an exhaustive search in the solution space [7]. However, exhaustive search is feasible only for small  $n$ . Different algorithms have been proposed for feature selection to obtain near-optimal solutions [4, 6, 8, 9, 10, 11]. The choice of an algorithm for selecting the features from an initial set depends on  $n$ . The feature selection problem is of small scale, medium scale, or large scale if  $n$  belongs to  $[0, 19]$ ,  $[20, 49]$ , or  $[50, \infty]$ , respectively [6, 10]. Sequential algorithms such as sequential forward floating search (SFFS) and sequential backward floating

search (SBFS) are efficient and usually find fairly good solutions for small- and medium-scale problems [9]. But they suffer from the problem of trapping into local optimal solutions for large-scale problems when  $n > 100$  [6, 10]. Modern iterative heuristics such as tabu search and genetic algorithms have been found effective in tackling this category of problems which have an exponential and noisy search space with numerous local optima [10, 11, 16].

Recently Zhang et al. have used tabu search with short-term memory to solve the optimal feature selection problem. The experimental results on synthetic data have shown that the tabu search not only has a high possibility to obtain the optimal or near-optimal solution, but also requires less computational time than the other suboptimal and genetic algorithm methods [10].

The aim of this paper is to improve the classification error rate of prostate needle biopsies for the four groups mentioned above while removing as many features as possible from the original feature vector. Tabu search method is used to find a feature subset. A novel approach based on intermediate-term memory has been developed to improve the search through tabu search. Due to multiobjective nature of the problem and the various cost parameters, that is, minimizing the number of features, maximizing the classification accuracy, and so forth, it is not clear what constitutes the best solution. Therefore, the objective function is measured by using the fuzzy formula, since fuzzy logic provides a suitable mathematical framework to address such problem. Classification accuracy is measured by leave-one-out correct-classification rate of a nearest neighbor (1-NN). Neighbors are calculated using the Euclidean distance. The 1-NN classifier is simple and it always provides a reasonable classification performance in most applications. The most straightforward 1-NN rule can be conveniently used as a benchmark for all other classifiers. Further, as the 1-NN classifier does not require any user-specified parameters, its classification results are implementation independent [4].

The rest of the paper is organized as follows. Section 2 gives an overview of tabu search and fuzzy logic followed by proposed tabu search design using intermediate-term memory for feature selection problem. The experiments and the comparison with other reported classifiers are presented in Section 4. Section 5 concludes the paper.

```

Begin
Start with an initial feasible solution  $S \in \Omega$ .
Initialize tabu list and aspiration level.
For fixed number of iterations, do
    Generate neighbor solutions  $V^* \subset N(S)$ .
    Find best  $S^* \in V^*$ .
    If move  $S$  to  $S^*$  is not in  $T$ , then
        Accept move and update best solution.
        Update tabu list and aspiration level.
        Increment iteration number.
    Else
        If  $\text{Cost}(S^*) < \text{AL}$ , then
            Accept move and update best solution.
            Update tabu list and aspiration level.
            Increment iteration number.
        End if
    End if
End for
End

```

ALGORITHM 1: Algorithmic description of a short-term tabu search (TS) [16].

## 2. OVERVIEW OF TABU SEARCH AND FUZZY LOGIC

### 2.1. Tabu search

Tabu search (TS) was introduced by Glover [17, 18] as a general iterative metaheuristic for solving combinatorial optimization problems. Tabu search is conceptually simple and elegant. It is a form of local neighborhood search. Each solution  $S \in \Omega$  has an associated set of neighbors  $N(S) \subseteq \Omega$ , where  $\Omega$  is the set of feasible solutions. A solution  $S' \in N(S)$  can be reached from  $S$  by an operation called a *move to  $S'$* . TS moves from a solution to its best admissible neighbor, even if this causes the objective function to deteriorate. To avoid cycling, solutions that have been recently explored are declared forbidden or tabu for a number of iterations. The tabu status of a solution is overridden when certain criteria (aspiration criteria) are satisfied. Sometimes, intensification and diversification strategies are used to improve the search. In the first case, the search is accentuated in the promising regions of the feasible domain. In the second case, an attempt is made to consider solutions in a broad area of the search space. The tabu search algorithm is given in Algorithm 1, where

- (i)  $\Omega$  is the set of feasible solutions,
- (ii)  $S$  is the current solution,
- (iii)  $S^*$  is the best admissible solution,
- (iv) Cost is the objective function,
- (v)  $N(S)$  is the neighborhood of solution  $S$ ,
- (vi)  $V^*$  is the sample of neighborhood solutions,
- (vii)  $T$  is the tabu list,
- (viii)  $AL$  is the aspiration level.

### 2.2. Fuzzy logic

Fuzzy set theory has been recently applied in many areas of science and engineering [20]. In the most practical situations, one is faced with several concurrent objectives. Classic

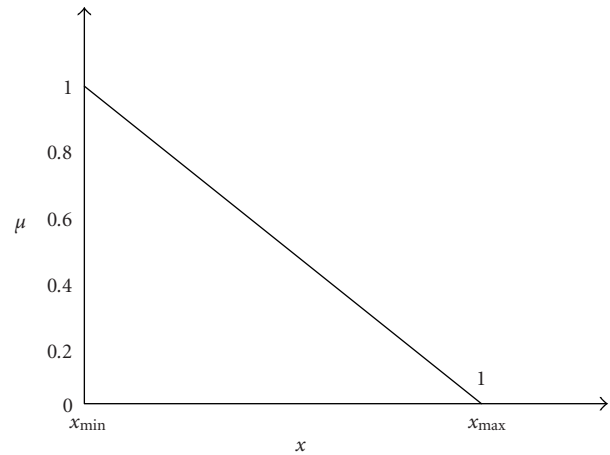


FIGURE 2: Membership function for fuzzy subset  $X$ , where  $X$  is the number of features  $F$ , the number of incorrect predictions  $P$ , or the classification error rate  $E$  in this application.

approaches usually deal with such difficulty by computing a single utility function as a weighted sum of the individual objectives, where more important objectives are assigned higher weights. Balancing different objectives by weight functions is at best controversial. Fuzzy logic is a convenient vehicle for trading off different objectives. It allows to map values of different criteria into linguistic values, which characterize the level of satisfaction of the designer with the numerical value of objectives, and to operate over the interval  $[0, 1]$  defined by the membership functions for each objective.

## 3. PROPOSED FEATURE SELECTION METHOD USING TABU SEARCH WITH INTENSIFICATION

### 3.1. Fuzzy objective function

In this paper, we present a tabu search algorithm, where the quality of a solution is characterized by a fuzzy logic rule expressed in linguistic variables of the problem domain. Three linguistic variables are defined to correspond to the three component objective functions: number of features  $f_1$ , number of incorrect predictions  $f_2$ , and average classification error rate  $f_3$ . One linguistic value is defined for each component of the objective function. These linguistic values characterize the degree of satisfaction of the designer with the values of objectives  $f_i(x)$ ,  $i = \{1, 2, 3\}$ . These degrees of satisfaction are described by the membership functions  $\mu_i(x)$  on fuzzy sets of the linguistic values. The membership functions for the minimum number of features, the minimum number of incorrect predictions, and the low classification error rate are easy to build. They are assumed to be nonincreasing functions because the smaller the number of features  $f_1(x)$ , the number of incorrect predictions  $f_2(x)$ , and the classification error rate  $f_3(x)$ , the higher the degree of satisfaction  $\mu_1(x)$ ,  $\mu_2(x)$ , and  $\mu_3(x)$  of the expert system (see Figure 2). The fuzzy subset of a good solution is defined by the following fuzzy logic rule:

“if a solution has *small number of features* and *small number of incorrect predictions* and *low classification error rate*, then it is a good solution.”

According to the and-like/or-like ordered weighted averaging logic [21, 22], the above rule evaluates to the following:

$$\mu(x) = \gamma \times \min(\mu_i(x)) + (1 - \gamma) \times \frac{1}{3} \sum_{i=1}^3 \mu_i(x), \quad (1)$$

where  $\gamma$  is a constant in the range  $[0, 1]$ . The shape of the membership function  $\mu(x)$  is shown in Figure 2. Membership of data in a fuzzy set is defined using values in the range  $[0, 1]$ . The membership values for the number of features  $F$ , the number of incorrect predictions  $P$ , and the classification error rate  $E$  are computed by using (2), (3), and (4), respectively:

$$\mu_1(x) = \begin{cases} 1 & \text{if } F \leq F \text{ Min,} \\ \frac{F \text{ Max} - F}{F \text{ Max} - F \text{ Min}} & \text{if } F \text{ Min} \leq F \leq F \text{ Max,} \\ 0 & \text{if } F \text{ Max} \leq F, \end{cases} \quad (2)$$

$$\mu_2(x) = \begin{cases} 1 & \text{if } P \leq P \text{ Min,} \\ \frac{P \text{ Max} - P}{P \text{ Max} - P \text{ Min}} & \text{if } P \text{ Min} \leq P \leq P \text{ Max,} \\ 0 & \text{if } P \text{ Max} \leq P, \end{cases} \quad (3)$$

$$\mu_3(x) = \begin{cases} 1 & \text{if } E \leq E \text{ Min,} \\ \frac{E \text{ Max} - E}{E \text{ Max} - E \text{ Min}} & \text{if } E \text{ Min} \leq E \leq E \text{ Max,} \\ 0 & \text{if } E \text{ Max} \leq E. \end{cases} \quad (4)$$

The maximum number of features ( $F \text{ Max}$ ) is the size of the feature vector and the minimum number of features ( $F \text{ Min}$ ) is 1. The maximum number of incorrect predictions ( $P \text{ Max}$ ) and the maximum classification error rate ( $E \text{ Max}$ ) are determined by applying 1-NN classifier in the initial solution. The minimum number of incorrect predictions ( $P \text{ Min}$ ) is 0 while the minimum classification error rate ( $E \text{ Min}$ ) is 0%.

### 3.2. Initial solution

Feature selection vector is represented by a 0/1-bit string where 0 shows the feature is not included in the solution while 1 shows the feature is included. All features are included in the initial solution.

### 3.3. Neighborhood solutions

Neighbors are generated by randomly adding or deleting a feature from the feature vector of size  $n$ . For example, if 11 001 is the current feature vector, then the possible neighbors with a candidate list size of 3 can be 10 001, 11 101, and 01 001. Among the neighbors, the one with the best cost (i.e., the solution which results in the minimum value of (1)) is selected and considered as a new current solution for the next iteration.

### 3.4. Tabu moves

A tabu list is maintained to avoid returning to previously visited solutions. With this approach, if a feature (move) is added or deleted at iteration  $i$ , then adding or deleting the same feature (move) for  $T$  iterations (tabu list size) is tabu.

### 3.5. Aspiration criterion

Aspiration criterion is a mechanism used to override the tabu status of moves. It temporarily overrides the tabu status if the move is sufficiently good. In our approach, if a feature is added or deleted at iteration  $i$  and this move results in a best cost for all previous iterations, then this feature is allowed to add or delete even if it is in the tabu list.

### 3.6. Termination rule

The most commonly used stopping criteria in TS are

- (i) after a fixed number of iterations,
- (ii) after some number of iterations without an improvement in the objective function value,
- (iii) when the objective reaches a prespecified objective value.

In our algorithm, termination condition is a predefined number of iterations.

### 3.7. Intensification of the search

For intensification, the search is accentuated in the promising regions of the feasible domain. Intensification is based on some intermediate-term memory. Since, the solution space is extremely large (initial feature vector  $n > 100$ ), it is important to intensify the search in the promising regions by removing poor features from the search space. The following steps are proposed to intensify the search.

*Step 1.* Store  $M$  best solutions in intermediate memory for  $T_1$  number of iterations.

*Step 2.* Remove features that are not included in the best  $M$  solutions for  $N$  times.

*Step 3.* Rerun the tabu search with the reduced set of features for another  $T_2$  iterations.

*Step 4.* Repeat Steps 1, 2, and 3 until the optimal or near-optimal solution is achieved.

The values of  $M$  and  $N$  can be determined empirically through experiments. As an example, assume that the following four best solutions as shown in Figure 3 are found by tabu search during  $T_1$  iterations. Features  $f_1$  and  $f_3$  are always used while feature  $f_5$  is never used for good solutions. For  $N = 2$ , the reduced feature set consists of only  $f_1, f_2, f_3,$  and  $f_6$ . Thus, tabu search will search for the near-optimal solutions in reduced search space to avoid visiting nonpromising regions.

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$
	1	0	1	0	0	1	0
	1	0	1	1	0	1	0
	1	1	1	0	0	0	1
	1	1	1	0	0	1	0
$\Sigma$	4	2	4	1	0	3	1

FIGURE 3: An example showing intensification steps for tabu search.  $\Sigma$  is the number of occurrences of each feature in the best solutions.

#### 4. EXPERIMENTS AND DISCUSSION

The tabu search in combination with 1-NN classifier is tested on two data sets reported in [2, 3]. In order to offset any bias due to the different range of values for the original features, the input feature values are normalized over the range [1, 11] by using (5) [23]. Normalizing the data is important to ensure that the distance measure allocates equal weight to each variable. Without normalization, the variable with the largest scale will dominate the measure:

$$x'_{i,j} = \left( \frac{x_{i,j} - \min_{k=1,\dots,n} x_{(k,j)}}{\max_{k=1,\dots,n} x_{(k,j)} - \min_{k=1,\dots,n} x_{(k,j)}} * 10 \right) + 1, \quad (5)$$

where  $x_{i,j}$  is the  $j$ th feature of the  $i$ th pattern,  $x'_{i,j}$  is the corresponding normalized feature, and  $n$  is the total number of patterns.

The first data set consists of textured multispectral images taken at 16 spectral channels (from 500 nm to 650 nm) [2, 24]. 592 different samples (multispectral images) of size  $128 * 128$  have been used to carry out the analysis. The samples are routinely viewed at prostatic section seen at low power (x 40 objective magnification) by two highly experimented independent pathologists and labelled into four classes: 165 cases of stroma, 106 cases of BPH, 144 cases of PIN, and 177 cases of PCa. The size of the feature vector is 128 (16 bands \* 8 features(2 structural + 6 Haralick)).

The second data set is derived from prostatic nuclei extracted from prostate tissue [3]. Nuclei are imaged under high power (x 100 objective magnification). These prostatic nuclei are taken at 33 spectral channels (from 400 nm to 720 nm). 230 different images of size  $256 * 256$  have been used to carry out the analysis. The samples are labelled into 3 classes: 63 cases of BPH, 79 cases of PIN, and 88 cases of PCa. The size of feature vector is 266 (33 bands \* 8 features(3 statistical+5 Haralick)+2 morphology features).

Table 1 shows the classification error for the first data set which is reported in [2] with data reduction by using principal component analysis and classification using classical

TABLE 1: Classification error reported in [2].

Classified as	BPH	PCa	PIN	STR	Error
BPH	100	3	2	1	5.66%
PCa	1	174	1	1	1.69%
PIN	1	4	133	6	7.63%
STR	11	1	1	152	7.87%
Overall	—	—	—	—	5.57%

TABLE 2: Classification error by using feature selection through tabu search.

Classified as	BPH	PCa	PIN	STR	Error
BPH	101	1	0	4	4.71%
PCa	1	174	2	0	1.69%
PIN	0	2	141	2	2.77%
STR	2	2	0	161	2.42%
Overall	—	—	—	—	2.90%

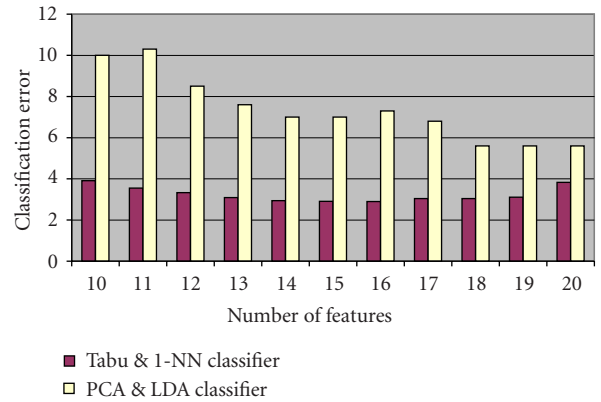


FIGURE 4: Features versus classification error for data set 1.

linear discrimination method. Table 2 shows the minimum classification error obtained by using fuzzy tabu search and classification results using leave-one-out 1-NN classifier. The classification accuracy has been increased for all cases. The overall classification error has been reduced to 2.90% as compared to 5.57% reported in [2]. Furthermore, the number of features used is only 16 out of the available 128 as compared to 20 obtained using principal component analysis in [2].

Figure 4 shows features vector size versus classification error rate. Our proposed algorithm is capable to obtain superior classification accuracy for various feature sizes when compared with technique used in [2]. Even the classification accuracy for the feature vector size of 10 is better than reported.

Table 3 shows the classification error for the second data set which is reported in [3] with data reduction by using principal component analysis and classification using classical quadratic discrimination method. Table 4 shows the minimum classification error obtained by using fuzzy tabu search and classification results using leave-one-out 1-NN classifier. The classification accuracy has been increased for all cases.

TABLE 3: Classification error reported in [3].

Classified as	BPH	PCa	PIN	Error
BPH	60	2	1	4.76%
PCa	3	82	3	6.82%
PIN	1	2	76	3.79%
Overall	—	—	—	5.10%

TABLE 4: Classification error by using feature selection through tabu search.

Classified as	BPH	PCa	PIN	Error
BPH	62	1	0	1.58%
PCa	1	87	0	1.14%
PIN	0	0	79	0%
Overall	—	—	—	0.91%

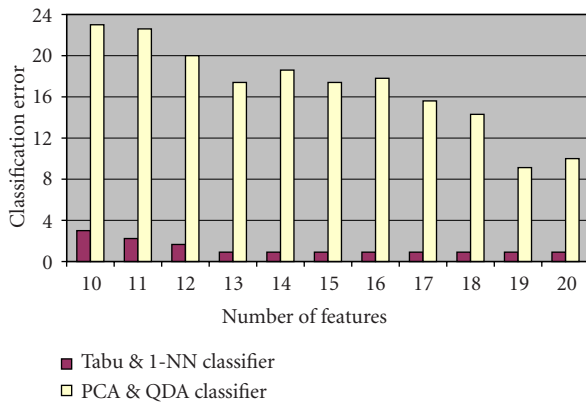


FIGURE 5: Features versus classification error for data set 2.

The overall classification error has been reduced to 0.91% as compared to 5.1% reported in [3]. The number of features used is only 13 out of the available 266 as compared to 25 obtained using principal component analysis in [3].

Figure 5 shows features vector size versus classification error rate. Our proposed algorithm is capable to obtain superior classification accuracy for various feature sizes when compared with technique used in [3]. Even the classification accuracy for feature vector size of 10 is better than reported.

#### 4.1. Quality of solutions by tabu search

Since the features are randomly added or deleted during tabu search, it is imperative that tabu search algorithm has the ability to find optimal or near-optimal solutions during different runs. Table 5 shows different solutions obtained by tabu search for different runs. If tabu search is executed with the same input and settings, it is highly likely that an acceptably accurate solution will be reached as shown in Table 5.

Figure 6 shows the fuzzy membership function, the number of incorrect predictions, the number of features, and classification error versus the number of iterations during the solution search space using tabu search. The figure clearly shows how well focused is tabu search on the good solution space. From the curve, it can be clearly seen that tabu search

TABLE 5: Classification error (%) for different runs using tabu search.

Data sets	1	2	3	4	5	Average
Data set 1	2.90	2.91	2.90	2.90	2.93	2.91
Data set 2	0.91	0.91	1.13	0.91	0.91	0.95

rapidly converged to the feasible/infeasible region border in all these objectives. Figure 7 depicts different objective functions versus the number of iterations after reducing the size of feature set by using intensification technique as mentioned in Section 2. From these curves, it can be seen that the searching for the best solutions is now limited to only good solution space, that is, fuzzy membership is in the range between 0.15 and 0.25 for most of the time while the same fuzzy membership was in the range between 0.2 and 0.4 before intensification.

#### 4.2. Runtime parameters for tabu search

Table 6 shows the tabu runtime parameters chosen after experimentation with different values. The values of  $M$  and  $N$  as mentioned in Section 3.1 are 50 and 10, respectively. Table 7 shows the results of the test for different values of  $M$  and  $N$ . A large value of  $M$  and a small value of  $N$  cause tabu search to converge slowly and search for the near-optimal solutions in almost the same search space as without intensification. On the contrary, a small value of  $M$  and a large value of  $N$  make the search less intensive. In practice, medium values of  $M$  and  $N$  are more appropriate for intensification.

#### 4.3. A discussion on computation time

The execution time to reduce the dimensionality of feature vector using tabu search and 1-NN classifier is much higher than reducing the dimensionality using PCA and LDA (10–12 hours as compared to few seconds). This can be explained by the fact that the former belongs to the class of iterative heuristics where for each iteration, different solutions are explored while the latter is based on only single solution.

However, the feature selection using TS/1-NN is an offline procedure which is used to find the best subset of features while keeping the classification error rate low. Once the feature selection procedure finds the best subset of features, 1-NN classifier is used to determine the class of the new sample (a multispectral image) which provides an online diagnosis decision for the pathologist. Thus, the main advantage is that the feature selection leads to savings in measuring features as most of the features are discarded. The selected features have retained their original physical meanings while for feature extraction the transformed features have no clear physical meaning [4].

Table 8 shows the computation time comparison when determining the class of new sample. The execution time using feature extraction is much higher for both data sets when compared with feature selection. As clear from the table, by using feature selection, the cost of classification is reduced by limiting the number of features which must be measured for feature extraction.

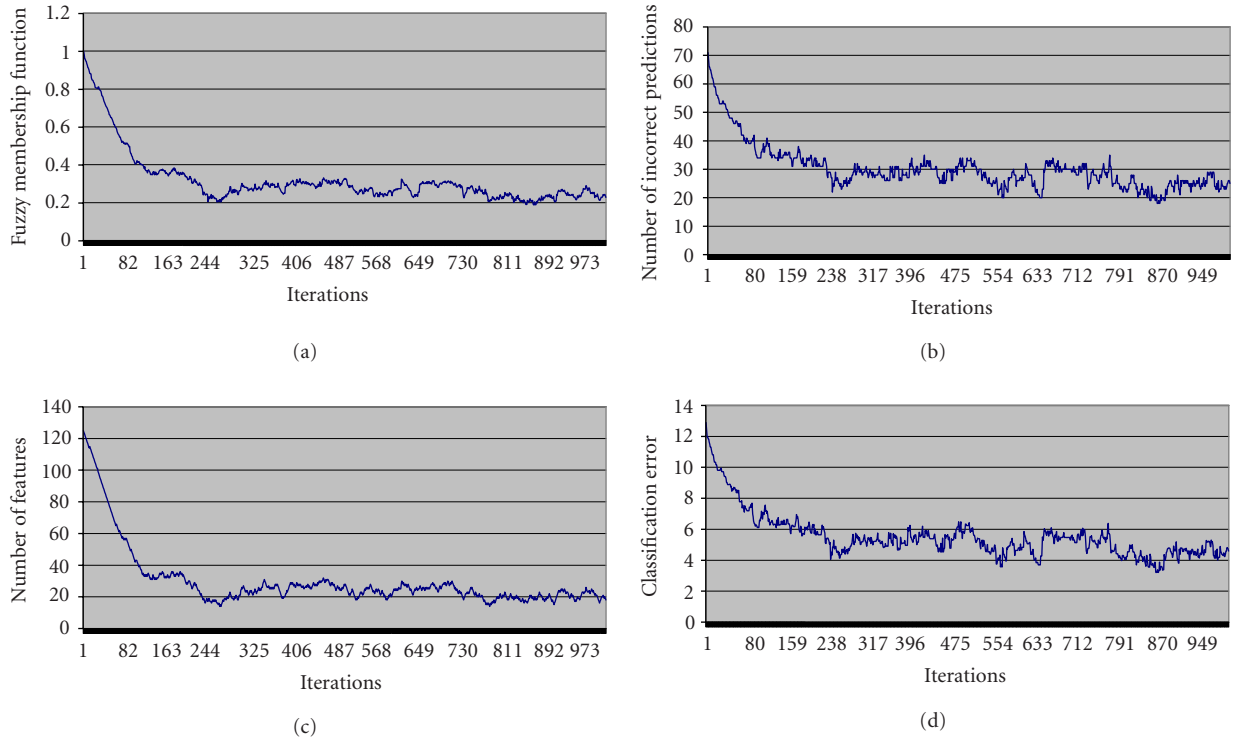


FIGURE 6: Objective functions versus number of iterations for data set 1 before intensification.

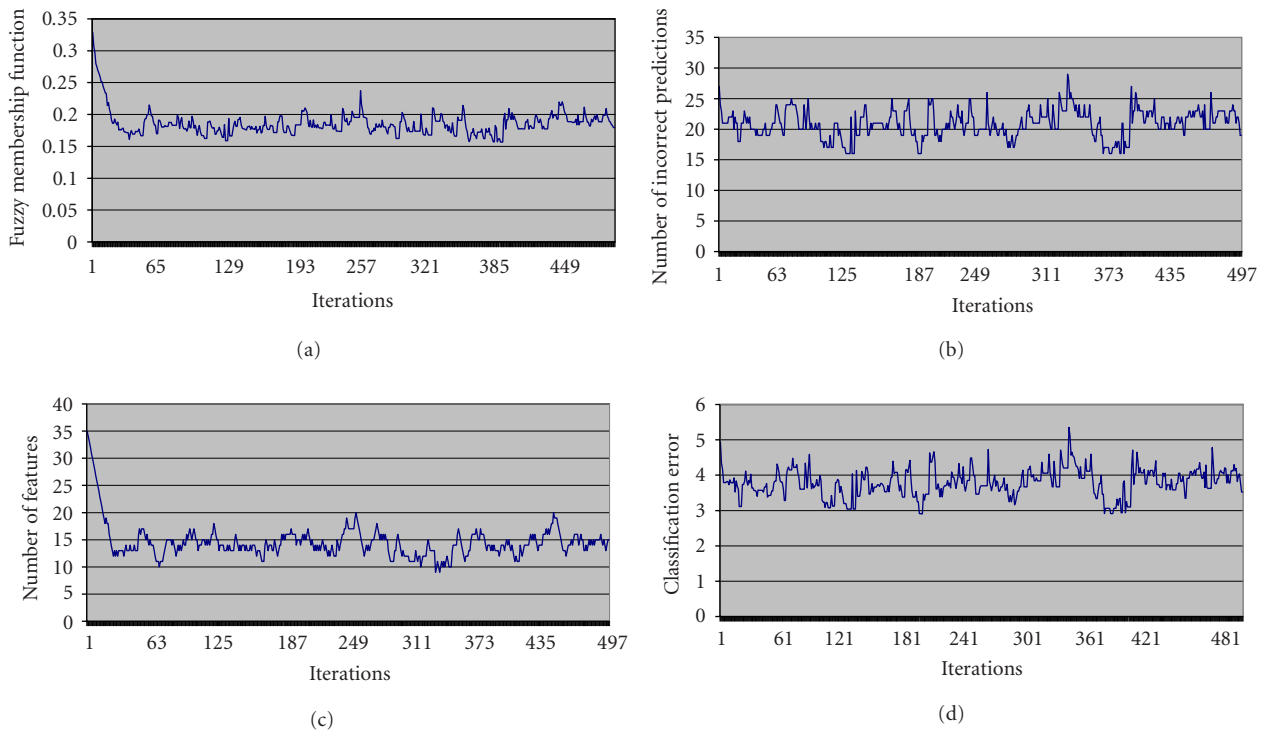


FIGURE 7: Objective functions versus number of iterations for data set 1 after intensification.

TABLE 6: Tabu runtime parameters.  $P$  = parameters,  $V^*$  = number of neighborhood solutions,  $T$  = tabu list size, and  $I$  = number of iterations.

P	Before intensification	After intensification
$V^*$	10	8
$T$	10	8
$I$	1000	500

TABLE 7: Comparison of different values of  $M$  and  $N$ .  $E$  = classification error rate and  $F$  = number of selected features.

Parameters		Data set 1		Data set 2	
$M$	$N$	$E$	$F$	$E$	$F$
10	2	2.94	15	1.18	16
30	6	3.09	14	1.13	16
50	10	2.90	16	0.91	13
70	14	2.90	16	0.91	17
90	18	3.04	17	0.91	18

TABLE 8: Computation time comparison.  $M$  = measuring features cost, DR = data reduction,  $C$  = classification, and  $T$  = total time.

Algorithm	Data sets	$M$ (ms)	DR (ms)	$C$ (ms)	$T$ (ms)
Feature extraction	Data set 1	85.0	0.047	0.0016	85.05
	Data set 2	654.6	0.12	0.0027	654.72
Feature selection	Data set 1	48.2	—	0.11	48.31
	Data set 2	159.4	—	0.042	159.44

## 5. CONCLUSION

In this paper, a tabu search method with intermediate-term memory is proposed for feature selection problem of large feature vector size. The cost of feature subset is measured by leave-one-out correct-classification rate of 1-NN classifier. The feature selection using tabu search has proved to be effective for the classification of prostate needle biopsies. Results have indicated a significant boost in the performance both in terms of minimizing the features and maximizing the classification accuracy. This method is quite generic and can be used with different classifiers of considerable dimensionality. Furthermore, the proposed tabu search progressively zoomed towards a better solution subspace as time elapsed, a desirable characteristics of approximation iterative heuristics. Future research is to exploit the parallelism inherent in tabu search to speed up the search.

## REFERENCES

- [1] J. N. Eble and D. G. Bostwick, *Urologic Surgical Pathology*, Mosby-Year Book, Saint Louis, Mo, USA, 1996.
- [2] M. A. Roula, J. Diamond, A. Bouridane, P. Miller, and A. Amira, "A multispectral computer vision system for automatic grading of prostatic neoplasia," in *Proc. IEEE International Symposium on Biomedical Imaging (ISBI '02)*, pp. 193–196, Washington, DC, USA, July 2002.
- [3] M. A. Roula, A. Bouridane, F. Kurugollu, and A. Amira, "A quadratic classifier based on multispectral texture features for prostate cancer diagnosis," in *Proc. 7th International Symposium on Signal Processing and Its Applications (ISSPA '03)*, vol. 2, pp. 37–40, Paris, France, July 2003.
- [4] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 1, pp. 4–37, 2000.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2001.
- [6] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000.
- [7] T. M. Cover and J. M. Van Campenhout, "On the possible orderings in the measurement selection problem," *IEEE Trans. Syst., Man, Cybern.*, vol. 7, no. 9, pp. 657–661, 1977.
- [8] A. K. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 2, pp. 153–158, 1997.
- [9] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [10] H. Zhang and G. Sun, "Feature selection using tabu search method," *Pattern Recognition*, vol. 35, no. 3, pp. 701–711, 2002.
- [11] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letters*, vol. 10, no. 5, pp. 335–347, 1989.
- [12] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000.
- [13] H. Zhang and G. Sun, "Feature selection using tabu search method," *Pattern Recognition*, vol. 35, no. 3, pp. 701–711, 2002.
- [14] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [15] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letters*, vol. 10, no. 5, pp. 335–347, 1989.
- [16] S. M. Sait and H. Youssef, *General Iterative Algorithms for Combinatorial Optimization*, IEEE Computer Society Press, Los Alamitos, Calif, USA, 1999.
- [17] F. Glover, "Tabu search I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.
- [18] F. Glover, "Tabu search II," *ORSA Journal on Computing*, vol. 2, no. 1, pp. 4–32, 1990.
- [19] S. M. Sait and H. Youssef, *General Iterative Algorithms for Combinatorial Optimization*, IEEE Computer Society Press, Los Alamitos, Calif, USA, 1999.
- [20] J. M. Mendel, "Fuzzy logic systems for engineering: a tutorial," *Proc. IEEE*, vol. 83, no. 3, pp. 345–377, 1995.
- [21] H. J. Zimmerman, *Fuzzy Set Theory and Its Applications*, Kluwer Academic, Norwell, Mass, USA, 3rd edition, 1996.
- [22] H. Youssef, S. M. Sait, and S. A. Khan, "Topology design of switched enterprise networks using a fuzzy simulated evolution algorithm," *Engineering Applications of Artificial Intelligence*, vol. 15, no. 3-4, pp. 327–340, 2002.
- [23] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 2, pp. 164–171, 2000.
- [24] M. A. Tahir, A. Bouridane, F. Kurugollu, and A. Amira, "Feature selection using tabu search for improving the classification rate of prostate needle biopsies," in *Proc. 17th IEEE International Conference on Pattern Recognition (ICPR '04)*, vol. 2, pp. 335–338, Cambridge, UK, August 2004.



**Muhammad Atif Tahir** received a B.E. degree in computer engineering from NED University of Engineering and Technology, Karachi, Pakistan, and an M.S. degree in computer engineering from King Fahd University, Dhahran, Saudi Arabia. He is currently a Ph.D. research student in School of Computer Science, Queen's University of Belfast, UK. His main research interests are in custom computing using FPGAs, image/signal processing, pattern recognition, QoS routing, and optimization heuristics.



**Ahmed Bouridane** obtained an "Ingéniorat d'Etat" degree in electronics from the National Polytechnic School of Algiers (ENP), an M.Phil. degree in VLSI design for signal processing from the University of Newcastle Upon Tyne, UK, and a Ph.D. degree in computer vision from the University of Nottingham, UK. He held several positions in R&D before joining Queen's University of Belfast where he is now a Reader in computer science. His research interests are in high-performance image/signal processing, image/video watermarking, custom computing using FPGAs, computer vision, and high-performance architectures for image/signal processing.



**Fatih Kurugollu** received his B.S., M.S., and Ph.D. degrees in computer science from the Istanbul Technical University, Istanbul, Turkey, in 1989, 1994, and 2000, respectively. He is currently a Lecturer of computer science at Queen's University, Belfast, UK. His research interests include soft computing for image and video object segmentation, hardware architectures for image and video applications, and image watermarking.



**Abbes Amira** obtained an "Ingéniorat d'Etat" degree in electronics and a DEA degree in image and speech processing from the National Polytechnic School of Algiers (ENP), and a Ph.D. degree in computer science from Queen's University, Belfast, UK. He is currently a Lecturer of computer science at Queen's University, Belfast. His research interests are in design and implementation of digital image and signal processing algorithms, and custom computing using field programmable gate arrays.

