Degree Prediction of Malignancy in Brain Glioma using Support Vector Machines

Guo-Zheng Li\textsuperscript{a,*}, Jie Yang\textsuperscript{a}, Chen-Zhou Ye\textsuperscript{a}, Dao-Ying Geng\textsuperscript{b}

\textsuperscript{a}Institute of Image Processing \\& Pattern Recognition, Shanghai Jiaotong University, Shanghai, China 200030

\textsuperscript{b}Hua-Shan Hospital-Medical Center of Fudan University, Shanghai, China 200040

Abstract

The degree of malignancy in brain glioma needs to be assessed by MRI findings and clinical data before operations. The previous attempted to solve this problem with a fuzzy rule extraction algorithm based on fuzzy min-max neural networks. We utilize support vector machines with floating search method to select relevant features and to predict the degree of malignancy. Computation results show that the feature subset selected by our techniques can yield better classification performance, in contrast with the base line method, which generated two rules and got 83.21\% accuracy on the whole data set, our method generates one rule to get 88.21\% accuracy.

Key words: Support vector machines; Feature selection; Brain glioma; Degree of malignancy; Floating search method
1 Introduction

The degree of malignancy in brain glioma [1] decides the treatment, because if it is grade I or II according to Kernohan, the success rate of operation is satisfactory; otherwise, if it is grade III or IV, there will be high surgical risk and poor life quality [2] after surgery which must be taken into account before any further decision. Now, the degree of malignancy is predicted mainly by Magnetic Resonance Imaging (MRI) findings [3] and clinical data [4] before operations. Some features obtained manually are fuzzy values, some features are redundant, even irrelevant, which makes the prediction of the degree of malignancy a hard task. Moreover, brain glioma is severe but infrequent, only a small number of neuroradiologists have the chances to accumulate enough experiences to make correct judgment. Therefore, it is worth generating some rules to predict the degree of malignancy of tumors.

Artificial neural networks (ANNs) are powerful tools in medical data processing field, and have been used widely [5]. However, ANNs are prone to overfitting, and are hard to understand, so several techniques are proposed to reduce the effect of overfitting and to extract rules from ANNs. Zhou et al. [6] proposed the C4.5 Rule-PANE method, which not only got high accuracy, but also got rules to help understand the working mechanism. However, the items

* Corresponding author.

Email addresses: lgz@sjtu.edu.cn (Guo-Zheng Li), jieyang@sjtu.edu.cn (Jie Yang), ye2002@sina.com (Chen-Zhou Ye), gengdy@163.com (Dao-Ying Geng).
of the rules are too many to be used conveniently in practice. Ye et al. [7]
proposed a fuzzy rule extraction algorithm based on fuzzy min-max neural
networks (FMMNN-FRE), although the rules are few and easy to understand,
the accuracy of FMMNN-FRE was not better than that of ANNs.

In the last decade, support vector machines (SVMs) proposed by Vapnik and
his coworkers [8,9], have developed rapidly and reached state-of-the-art perfor-
mance [10]. Compared to ANNs, SVMs have the better generalization perfor-
mance and can obtain a global optimal solution. At the same time, a type
of feature selection algorithms based on SVMs named embedded algorithms
has been proposed [11], to solve the problems involving many irrelevant and
redundant features such as gene selection and text categorization. These em-
bedded algorithms are designed to select features efficiently, but, compared to
wrapper methods [12] the accuracy is sacrificed in some degree. Furthermore,
subset generation methods used in most embedded algorithms are methods
like sequential backward feature selection which can not effectively handle
combined features in the feature selection procedure.

In the literature, floating search method [13] which is a heuristic feature sub-
set generation algorithm, has been proved one of the best subset generation
algorithms [14] when the training sample is comparatively small. At the same
time, the wrapper method [12] which is optimal to specific learning machine
can get better performance than filter methods and embedded methods such
as SVM-RFE. Since the data set of brain glioma is comparatively small, a
method which combines floating search method with wrapper based on SVMs can be applied to it.

In the previous work, degree prediction of malignancy had been solved by Ye et al. [7]. Though high understandable rules have been obtained, accuracy of prediction is not better than that of ANNs. In order to obtain both high accuracy and well understandable rules, we combine floating search method with wrapper method based on SVMs to deal with the prediction of degree of malignancy in brain glioma. The rest of the paper is organized as follows. Section 2 presents a description of our method, which combines floating search method with wrapper method based on SVMs. Section 3 describes the brain glioma data set. Results and comparative study are presented in Section 4. Finally, a discussion on this work is given in Section 5.

2 Method of computation

Wrapper method based on support vector machines and backward floating search method are combined to select relevant features and generate rules to predict the degree of malignancy in brain glioma, the former is used as feature subset evaluation method, and the latter is as feature subset generation method.
2.1 Support vector machines (SVMs)

Support vector machines (SVMs) [8] proposed by Vapnik and his co-workers in 1990s, have been developed quickly during the last decade [10], and successfully applied to diverse domains such as pattern recognition [15], text categorization [16], computer vision [17], biological data mining [18], and medical diagnosis [19,20], etc.

Denoting the training sample as $S = \{(x, y)\} \subseteq \{\mathbb{R}^n \times \{-1, 1\}\}^\ell$, SVMs discriminant hyperplane can be written as

$$y = \text{sgn}(\langle w \cdot x \rangle + b)$$

where $w$ is a weight vector, $b$ is a bias. According to the generalization bound in statistical learning theory [9], we need to minimize the following objective function for a 2-norm soft margin version of SVMs

$$\min_{w, b} \langle w \cdot w \rangle + C \sum_{i=1}^\ell \xi_i^2$$

subject to

$$y_i(\langle w \cdot x_i \rangle + b) \geq 1 - \xi_i, \quad i = 1, ..., \ell,$$

in which, slack variable $\xi_i$ is introduced when the problem is infeasible. The constant $C > 0$ is a penalty parameter, a larger $C$ corresponds to assigning a larger penalty to errors.

By building a Lagrangian and using the Karush-Kuhn-Tucker (KKT) complementarity conditions [21,22], we can obtain the value of optimization problem
(1). Because of the KKT conditions, only those Lagrangian multipliers, $\alpha_i$s, which make the constraint active are non zeros, we denote these points corresponding to the non zero $\alpha_i$s as support vectors (sv). Therefore we can describe the classification hyperplane in terms of $\alpha$ and $b$:

$$y = \text{sgn} \left( \sum_{i \in sv} \alpha_i (x_i \cdot x) + b \right).$$

The SVMs software used in our paper is mySVM package [23]. Description in detail of SVMs can be found at: http://www.kernel-machines.org.

2.2 Wrapper method based on SVMs

Wrapper method [12] is used in this work, which means to utilize the prediction ability of some learning machine, like SVMs in this work, to evaluate the feature subset. Compared to filter methods and embedded methods [11], wrapper model tends to obtain a feature subset with higher accuracy measured by the used learning algorithm. It is defined as

$$W = R_{\text{acc}}(S_w)$$

where $S_w$ is a feature subset candidate, and $R_{\text{acc}}$ is calculated by 10-fold cross validation method. For the training sample $S$, $R_{\text{acc}}$ is defined as:

$$R_{\text{acc}} = \frac{1}{\ell} \sum_{j=1}^{\ell} (\tilde{y}_j = = y_j)$$
where $\ell$ is the number of training examples, $\hat{y}$ is the predicted value by the used learning machines, like SVMs in this work. The feature subset corresponding to the largest $W$ is the most significant one.

2.3 SVM-BFS method

Backward floating search (BFS) proposed by Pudil et al. [13] is a feature subset generation method, and has been proved to be one of the best subset generation algorithms [14]. In this work, in order to solve the degree prediction of malignancy in brain glioma we combine BFS with wrapper based on SVMs and name it as SVM-BFS method. The motivation of this approach is to meet the need of high prediction accuracy and strong comprehensibility.

BFS is based on sequential backward search (SBS) and sequential forward search (SFS), of which SBS eliminates the least important feature one by one, while SFS adds the most important feature one by one. The least important feature $f_x$ means that for a subset $S_w$, there exists $R_{acc}(S_w - f_x) \geq R_{acc}(S_w - f_i)$, $\forall f_i \in S_w$, while the most important feature $f_x$ means that for a subset $S_w$ which satisfies $S_w \cup S_u = S$ and $S_w \cap S_u = \emptyset$, there exists $R_{acc}(S_w + f_x) \geq R_{acc}(S_w + f_i)$, $\forall f_i \in S_u$, where $R_{acc}(S_w \mp f_x)$ means $S_w$ removes or adds $f_x$. FiM$(S_w, S)$ is a function used to find the most important feature for $S_w$ from $S$.

SVM-BFS can be divided into five steps.
Step0: initialization, it uses SBS method to reduce the least two important features and sets the value for the current number and the target number of features, since we want to eliminate the features from the maximum to 1, we set the target number is 1.

Note, in the following steps, keep the number of current feature subsets two less than the number of the total feature set, otherwise go to Step1.

Step1: exclusion, it uses the basic SBS method to remove the feature. In BFS this step is the main module.

Step2: conditional inclusion, it finds the most significant feature, among the excluded features, if it is not the feature just eliminated, the basic SFS is used to add it.

Step3: continuation of conditional inclusion. it continues to find the most significant feature among the excluded features with respect to the subset got in Step2, if $R_{acc}$ of the subset is no higher than that of any other subset with the same number of feature, then goto Step1; otherwise, goto Step3.

Step4: display, if the number of current subset is greater than the target number, go to Step1; otherwise, display the optimal feature subset and the merit of subset with different number of features.

In summary, the pseudo code of SVM-BFS is shown in Table 1 and 2, and the notations are listed in Table 3.
3 Brain glioma data set

The brain glioma data set [7] was gathered by neuroradiologists from Huashan Hospital in Shanghai of China. There are more than 20 items in each case, including symptoms on different features, preoperative diagnosis made by some neuroradiologist and, without an exception, a clinical grade (the actual grade of glioma obtained from surgery). With the help of domain experts, we chose fifteen features, *Gender, Age, Shape, Contour, Capsule of Tumor, Edema, Mass Effect (Occupation), Post-Contrast Enhancement, Blood Supply, Necrosis/Cyst Degeneration, Calcification, Hemorrhage, Signal Intensity of the T1-weighted Image, Signal Intensity of the T2-weighted Image, and Clinical Grade* [4,24]. In some cases, the value of *Post-Contrast enhancement* is unknown. In fact, *Location* and *Size* also help to make the diagnosis, but their complex descriptions can not be well modeled by our algorithms, so we didn’t adopt it. Except *Gender, Age* and *Clinical Grade*, other items are got from MRI of the patient and are described with uncertainty to various extents. In order to predict the degree of malignancy in brain glioma, descriptions of all features are converted into numerical values in Table 4 and 5, of which the
unknown value of \textit{Post-Contrast enhancement} is defined as $-1$.

TABLE 4 about here

TABLE 5 about here

Originally, four grades are used to mark the degree of malignancy, we merge grade I and II into low-grade and grade III and IV to high-grade. According to the grade, all 280 cases of brain glioma are divided into two classes: low-grade and high-grade, in which 169 are of low-grade glioma and 111 are of high-grade ones. There are 126 cases containing missing values on \textit{Post-Contrast enhancement}, and in the other subset of 154 complete cases, 85 cases are of low-grade gliomas and 69 ones are of high-grade ones. Computation will be performed on D280 and D154, of which D280 is the total data set of 280 cases with missing values in \textit{Post-Contrast enhancement}, while D154 is the the data set of 154 completed cases.

More details about this data set can also be referred to [7].

4 Computation results

Firstly, we compute the accuracy of prediction for two data sets with full features, then, we select the most relevant feature subset by SVM-BFS, and compute the accuracy by different learning algorithms on two data sets with the selected feature subset. At last, rules are generated to help the neuroradi-
ologists predict the degree of malignancy in brain glioma. The accuracy will be compared with that of FMMNN-FRE on the whole training data set with the selected feature subset.

Before the computation, we perform a preprocessing step to normalize the data to the quadrant of [-1, 1] using an affine transformation.

4.1 Feature subset selected by SVM-BFS

4.1.1 Accuracy on the total feature set

To compare the classification ability of the methods of SVMs, artificial neural networks, FMMNN-FRE on both the training data sets with the total features, we use the 10-fold cross validation method to give the results in Table 6. The neural networks with weight decay in Bayesian learning frame(BNN)[25] is used in this work to be compared with SVMs and FMMNN-FRE, because this type of neural network has a regularizing term, whose coefficients are computed in Bayesian learning frame, the neural network can overcome the overfitting problem and can be insensitive with the number of nodes in the hidden layer, which has been implemented in MATLAB[26].

[Table 6 about here]

From Table 6, it can be seen that SVMs are slightly superior to the other two algorithms on both data sets.
4.1.2 Feature selection by SVM-BFS

With the proposed algorithm in Section 2, the relevant feature subsets are selected on both data sets. Results of the prediction accuracy on the selected feature subset with different number are shown on Fig 1.

[Figure 1 about here]

From the results of feature selection on 280 cases data set, it can be seen that a subset of six features gets the highest accuracy, though there are three subsets which contain different number of features get the same accuracy, we consider the subset which contains more than six features has redundant features, so the subset with six features is chosen as the selected feature subset on 280 cases, which contains Gender, Age, Capsule of Tumor, Post-Contrast Enhancement, Blood Supply, and Hemorrhage. Similarly, on 154 cases data set, we also select a subset with six features, which are Gender, Age, Post-Contrast Enhancement, Blood supply, Necrosis/Cyst Degeneration, and Signal Intensity of the T2-weighted Image. Totally, we select eight features on both data sets.

Accidentally, Ye et al. [7] also got six features for each data set, and totally 8 features. There are four features which are identical with ours, which are Age, Post-Contrast Enhancement, Blood supply and Hemorrhage. In addition, we use four features of the rest, which are Gender, Capsule of Tumor, Necrosis/Cyst Degeneration and Signal Intensity of the T2-weighted Image, while
Ye et al. use other four different features, which are *Edema, Mass Effect, Calcification*, and *Signal Intensity of the T1-weighted Image*.

In fact, according to our experiences, all features are relevant to the degree of malignancy in brain glioma, but different combinations of features can yield different prediction accuracy.

### 4.1.3 Comparison of classification accuracy on the selected feature subsets

Results of classification accuracy on the selected feature subset by SVM-BFS are calculated using SVMs and BNN with the 10-fold cross validation method, and shown in Table 7. At the same time, results on the selected feature subset by FMMNN-FRE, are shown in Table 8.

(Table 7 about here)

(Table 8 about here)

From Table 6, 7 and 8, it can be seen that 1) Classification accuracy by SVMs on the selected feature subset is higher than that by FMMNN-FRE; 2) Accuracy by SVMs on the selected feature subset is higher than that on the total feature set; 3) Accuracy by BNN on the selected feature subset is slightly higher than on the total feature set; 4) Accuracy by SVMs on the feature subset by SVM-BFS is higher than that on the feature subset by FMMNN-FRE.
4.2 Rules generated on the selected feature subset

We train SVMs on two data sets and get one rule for each data set. On all 280 cases, the rule is,

\[ g = 4.0508 + [-0.5762 - 1.7209, -0.1762, -0.3244, -1.2274, -0.4096]^* \]

\[ [\text{Gender}, \text{Age}, \text{Capsule of Tumor}, \text{Post} - \text{Contrast Enhancement}, \text{Blood Supply}, \]

\[ \text{Hemorrhage}]', \]

(2)

where a positive value means the degree is of low grade, while a negative one means it is of high grade, and \( a' \) means the transpose of \( a \). This rule is quantitative, and means that the tumor degree of \textit{young} people with female \textit{Gender}, intact \textit{Capsule of Tumor}, absent \textit{Post-Contrast Enhancement}, normal \textit{blood Supply}, and absent \textit{Hemorrhage} will be of low grade, otherwise, it will be of high grade.

On 154 cases, the rule is,

\[ g = 5.1603 + [-0.4092, -2.9368, -0.9734, -1.1179, -0.3506, 0.1928]^* \]

\[ [\text{Gender}, \text{Age}, \text{Post} - \text{Contrast Enhancement}, \text{Blood supply}, \]

\[ \text{Necrosis/Cyst Degeneration, Signal Intensity of the T2 - weighted Image}]', \]

(3)

where a positive value means the degree is of low grade, while a negative one means it is of high grade, and \( a' \) means the transpose of \( a \). It means
the tumor degree of young people with female Gender, absent Post-Contrast Enhancement, normal blood Supply, absent Necrosis/Cyst Degeneration and hypointense or hypointense accompanied by isointense Signal Intensity of the T2-weighted Image will be of low grade, otherwise, it is of high grade.

These two rules are conducted from two related data sets, and they are unanimous. In a word, the two rules can be fused into one that the tumor degree of young people with female Gender, intact Capsule of Tumor, absent Post-Contrast Enhancement, normal blood Supply, absent Necrosis/Cyst Degeneration, absent Hemorrhage and hypointense or hypointense accompanied by isointense Signal Intensity of the T2-weighted Image tends to be low grade, otherwise it tends to be high grade.

Classification accuracy of the above two rules are calculated on the respective data sets. Results are shown in Table 9. From Table 9, it can be seen that rules got by SVMs on the features selected by SVM-BFS can get higher accuracy than those by FMMNN-FRE do. Although we get higher accuracy than FMMNN-FRE and used different features, our rules have no conflicts with rules got by Ye et al. [7].
From the computation results, it can be seen that SVM-BFS get higher classification accuracy and more information than the FMMNN-FRE algorithm. This is due to support vector machines with better generalization ability and a different feature subset.

Compared with FMMNN-FRE, there are three distinct specialties of SVM-BFS. One is that with the concept of margin, SVMs can tolerate the noise and the fuzzy value in the data set, while FMMNN-FRE use the fuzzy set method to treat the fuzzy value in the data set, the former is easier to be implemented.

The second specialty is that with the concept of margin, SVMs realize the principle of data dependent structure risk minimization, in which they make use of the relation between the target function and the data set, and do not depend on the dimension of data set, at the same time, they minimize the structure risk of both complexity and minimising loss. All these greatly improve the generalisation performance.

The third specialty is that feature selection methods are used to select the most relevant feature subset. Compared with FMMNN-FRE, a rule extraction method, backward floating search method can eliminate the redundant features and handle the combined features efficiently which are more important in prediction, to help the learning machine get better generalization perfor-
In addition, although SVM-BFS does not take steps to treat the missing values in Post-Contrast Enhancement, accuracy on the data set of 280 cases is nearly equal to that on the data set of 154 cases. However, FMMNN-FRE treat the missing values using hyperbox, but accuracy on the data set of 280 cases is 3% less than that on 154 cases. This advantage of SVM-BFS should be owed to the second speciality of SVM-BFS, the excellent generalization performance.

From the rules (2) and (3), it can be seen that the feature subset of Age, Post-Contrast Enhancement, Blood supply, Hemorrhage, Gender, Capsule of Tumor, Necrosis/Cyst Degeneration and Signal Intensity of the T2-weighted Image, seem to be more useful than other subset even the full feature set for the degree prediction of malignancy in brain glioma. These features used in our subset have already been verified, i.e. Gender is in [3], Age in [7,3], Capsule of Tumor, Hemorrhage in [7], Post-Contrast Enhancement in [27] Blood supply in [7], Necrosis/Cyst Degeneration and Signal Intensity of the T2-weighted Image in [4]. The rules got in this work are also in accord with the experiences of human experts[3,2]. In fact, the combination of features but not the individuals are more important for the prediction, so the features need not be explained one by one from the view of statistics.

At last, we can conclude that 1) combining support vector machines and feature selection methods can get another different feature subset, based on which
higher accuracy and qualitative information for the degree prediction of malignancy have been got, since this subset does not conflict with the results in [7], this paper may be complementary to the research of Brain Glioma; 2) support vector machines with feature selection has the potential to be powerful tools in computer aided medical diagnosis problems like the degree prediction of malignancy in brain glioma.

6 Acknowledgments

Thanks to Stefan Rüping for his free SVMs toolbox. Thanks to the National Natural Science Foundation of China for their financial support under the grant number 50174038, 30170274.
References


List of Figures

1. $R_{acc}$ on both data sets with different number of features selected by SVM-BFS, of which a is on the data set of 280 cases and b is on the data set of 154 cases.
List of Tables

1  SVM-BFS approach 26

2  Continuation part of SVM-BFS approach 27

3  Notations in the pseudo code of SVM-BFS 28

4  Description of features used in brain glioma data set 29

5  Description of features used in brain glioma data set (Cont’) 30

6  Results of accuracy got by 10-fold cross validation method with the total features 31

7  Results of accuracy got by 10 fold cross validation method on the selected feature subset by SVM-BFS 32

8  Results of accuracy got by 10 fold cross validation method on the selected feature subset by FMMNN-FRE 33

9  Classification accuracy got by the rules 34
Fig. 1. $R_{acc}$ on both data sets with different number of features selected by SVM-BFS, of which a is on the data set of 280 cases and b is on the data set of 154 cases.
Table 1

SVM-BFS approach

Input: $S$, $n$

output: $S_o$, $E$

$n_c = n, n_t = 1$ and use SBS to remove two features /* Initialization */

$[S_{n_c-1}, E_{n_c-1}, f_c] = \text{SBS}(S_{n_c}); n_c = n_c - 1; /*Exclusion*/$

$f_b = \text{FiM}(S_{n_n-c-1}, S); /*Conditional inclusion*/$

if $(f_c \neq f_b)$ {

$S_b = \text{SFS}(S_{n-n_c-1}, f_b); n_c = n_c + 1;$

if $(n - n_c = 2)$

$S_{n-n_c-1} = S_b; E_{n-n_c-1} = R_{acc}(S_{n-n_c-1});$

else while(1) /* Continuation of conditional inclusion */

Continuation Part;

}

if $(n_c > n_t)$

goto Exclusion;

else

display $S_o, E_i, i = 1, \ldots, n; /* display */$
Table 2

Continuation part of SVM-BFS approach

\[
f_b = \text{FiM}(S_b, S);
\]

\[
S_t = \text{SFS}(S_b, f_b);
\]

if \((R_{\text{acc}}(S_t) < E_{n-n_c-2})\) {

\[
S_{n-n_c-1} = S_b; \ E_{n-n_c-1} = R_{\text{acc}}(S_{n-n_c-1});
\]

break;

}

else{

\[
S_b = S_t; n_c + +;
\]

if \((n - n_c == 2)\) {

\[
S_{n-n_c-1} = S_b; \ E_{n-n_c-1} = R_{\text{acc}}(S_{n-n_c-1});
\]

break;

}

}
<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>([S_u, E_u, f_u] = \text{SBS}(S_w))</td>
<td>to eliminate the least important feature (f_u) from (S_w) and get the subset (S_u) as well as the (E_u = R_{\text{acc}}(S_u))</td>
</tr>
<tr>
<td>(S_u = \text{SFS}(S_w, n_w))</td>
<td>to add the most important feature to (S_w) and get the subset (S_u)</td>
</tr>
<tr>
<td>(E)</td>
<td>the merit of the generated feature subsets</td>
</tr>
<tr>
<td>(f_b)</td>
<td>the best feature not in current feature subset</td>
</tr>
<tr>
<td>(f_c)</td>
<td>the feature being removed currently</td>
</tr>
<tr>
<td>(n)</td>
<td>the number of total feature set</td>
</tr>
<tr>
<td>(n_c)</td>
<td>the number of current subset</td>
</tr>
<tr>
<td>(n_t)</td>
<td>the number of target feature subset</td>
</tr>
<tr>
<td>(S)</td>
<td>the training set</td>
</tr>
<tr>
<td>(S_b)</td>
<td>(R_{\text{acc}}) got with the subset is higher than that with the current subset</td>
</tr>
<tr>
<td>(S_o)</td>
<td>the output feature subset</td>
</tr>
<tr>
<td>(S_t)</td>
<td>a temporary subset</td>
</tr>
</tbody>
</table>
Table 4

Description of features used in brain glioma data set

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0-Female, 1-Male.</td>
</tr>
<tr>
<td>Age</td>
<td>used directly.</td>
</tr>
<tr>
<td>Shape</td>
<td>1-Round, 2-Ellipse, 3-Irregular.</td>
</tr>
<tr>
<td>Contour</td>
<td>1-Clear, 2-Partially Clear, 3-Blur.</td>
</tr>
<tr>
<td>Capsule of Tumor</td>
<td>a hypointense ring surrounding the tumor on T1-weighted and T2-weighted</td>
</tr>
<tr>
<td></td>
<td>images. 1-Intact, 2-Partially, 3-Absent.</td>
</tr>
<tr>
<td>Edema</td>
<td>an area surrounding the tumor with hypointense signal on T1-weighted</td>
</tr>
<tr>
<td></td>
<td>image and hyperintense signal on T2-weighted image. 0 - Absent (no such area),</td>
</tr>
<tr>
<td></td>
<td>1 - Light (width of the area ≤ 2cm), 2 - Middle (2cm &lt; width of the area ≤ 1/2 width of the hemisphere at the same side), and 3 - Heavy (width of the area &gt; 1/2 width of the hemisphere at the same side).</td>
</tr>
<tr>
<td>Mass Effect</td>
<td>0 - Absent (adjacent tissues are not compressed), 1 - Light (enlargement of gyrus can be observed; adjacent sulcus is occluded), 2 - Middle (adjacent tissues and ventricle is shifted), and 3 - Heavy (middle line structure is shifted to the opposite direction).</td>
</tr>
<tr>
<td>Feature</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Post-Contrast</td>
<td>-1-unknown, 0-Absent, 1-Homogeneous, 2-Heterogeneous.</td>
</tr>
<tr>
<td>Blood Supply</td>
<td>abnormal flow voids suggesting abnormal vascularity. 1 - Normal, 2 - Middle (one or two linear flow voids), and 3 - Affluent (flow voids in clusters).</td>
</tr>
<tr>
<td>Necrosis/Cyst</td>
<td>0-Absent (no cystic lesion), 1-Present (one or more cystic lesions).</td>
</tr>
<tr>
<td>Degeneration</td>
<td></td>
</tr>
<tr>
<td>Calcification</td>
<td>spot-like or patch-like abnormal areas with hypointense signal on both T1-weighted image and T2-weighted images. 0-Absent, 1-Present.</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>0-Absent, 1-Acute, 2-Chronic.</td>
</tr>
<tr>
<td>Signal Intensity of</td>
<td>1 - Hypointense Only (relative to the gray mater), 2 - Isointense or Isointense accompanied by Hypointense, and 3 - Hyperintense or Hyperintense accompanied by Isointense or/and Hypointense.</td>
</tr>
<tr>
<td>the T1-weighted</td>
<td></td>
</tr>
<tr>
<td>Image</td>
<td></td>
</tr>
<tr>
<td>Signal Intensity of</td>
<td>1 - Hyperintense Only, 2 - Isointense or Isointense accompanied by Hyperintense, and 3 - Hypointense or Hypointense accompanied by Isointense or/and Hyperintense.</td>
</tr>
<tr>
<td>the T2-weighted</td>
<td></td>
</tr>
<tr>
<td>Image</td>
<td></td>
</tr>
</tbody>
</table>
Table 6

Results of accuracy got by 10-fold cross validation method with the total features

<table>
<thead>
<tr>
<th>Data set</th>
<th>$R_{acc}$ (%)</th>
<th>Std. dev.(%)</th>
<th>Highest(%)</th>
<th>Lowest(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVMs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D280</td>
<td>85.70</td>
<td>6.52</td>
<td>96.43</td>
<td>71.43</td>
</tr>
<tr>
<td>D154</td>
<td>84.96</td>
<td>9.97</td>
<td>100.00</td>
<td>73.33</td>
</tr>
<tr>
<td>BNN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D280</td>
<td>85.62</td>
<td>7.68</td>
<td>100.00</td>
<td>73.33</td>
</tr>
<tr>
<td>D154</td>
<td>84.00</td>
<td>7.76</td>
<td>100.00</td>
<td>73.33</td>
</tr>
<tr>
<td>FMMNN-FRE†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D280</td>
<td>83.21</td>
<td>5.31</td>
<td>89.29</td>
<td>75.00</td>
</tr>
<tr>
<td>D154</td>
<td>86.37</td>
<td>8.49</td>
<td>100.00</td>
<td>73.33</td>
</tr>
</tbody>
</table>

†Results are collected from [7] and values are predicted by FMMNN-FRE, in which feature selection has been implicitly performed.
Table 7

Results of accuracy got by 10 fold cross validation method on the selected feature subset by SVM-BFS

<table>
<thead>
<tr>
<th>Data set</th>
<th>$R_{acc}$ (%)</th>
<th>Std. dev.(%)</th>
<th>Highest(%)</th>
<th>Lowest(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVMs D280</td>
<td>87.14</td>
<td>5.38</td>
<td>96.43</td>
<td>78.57</td>
</tr>
<tr>
<td>D154</td>
<td>88.33</td>
<td>8.07</td>
<td>100.00</td>
<td>80.00</td>
</tr>
<tr>
<td>BNN D280</td>
<td>86.07</td>
<td>6.61</td>
<td>96.42</td>
<td>78.57</td>
</tr>
<tr>
<td>D154</td>
<td>84.75</td>
<td>9.28</td>
<td>100</td>
<td>68.75</td>
</tr>
</tbody>
</table>
Table 8

Results of accuracy got by 10 fold cross validation method on the selected feature subset by FMMNN-FRE

<table>
<thead>
<tr>
<th>Data set</th>
<th>$R_{acc}$ (%)</th>
<th>Std. dev.(%)</th>
<th>Highest(%)</th>
<th>Lowest(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVMs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D280</td>
<td>85.36</td>
<td>8.66</td>
<td>96.43</td>
<td>64.28</td>
</tr>
<tr>
<td>D154</td>
<td>84.37</td>
<td>9.44</td>
<td>93.75</td>
<td>66.67</td>
</tr>
<tr>
<td>BNN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D280</td>
<td>85.35</td>
<td>8.65</td>
<td>96.42</td>
<td>67.85</td>
</tr>
<tr>
<td>D154</td>
<td>85.04</td>
<td>5.49</td>
<td>93.33</td>
<td>73.33</td>
</tr>
<tr>
<td>FMMNN-FRE‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D280</td>
<td>83.21</td>
<td>5.31</td>
<td>89.29</td>
<td>75.00</td>
</tr>
<tr>
<td>D154</td>
<td>86.37</td>
<td>8.49</td>
<td>100.00</td>
<td>73.33</td>
</tr>
</tbody>
</table>

‡please refer to Table 6.
Table 9

Classification accuracy got by the rules

<table>
<thead>
<tr>
<th>method</th>
<th>data set</th>
<th>total(%)</th>
<th>high-grade(%)</th>
<th>low-grade(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM-BFS</td>
<td>D280</td>
<td>88.21</td>
<td>86.98</td>
<td>90.09</td>
</tr>
<tr>
<td></td>
<td>D154</td>
<td>88.96</td>
<td>88.41</td>
<td>89.41</td>
</tr>
<tr>
<td>FMMNN-FRE‡</td>
<td>D280</td>
<td>84.64</td>
<td>76.58</td>
<td>89.94</td>
</tr>
<tr>
<td></td>
<td>D154</td>
<td>87.66</td>
<td>85.88</td>
<td>89.86</td>
</tr>
</tbody>
</table>

‡Results are collected from [7].
Summary

The degree of malignancy in brain glioma needs to be assessed by Magnetic
Resonance Imaging findings and clinical data before operations. Since these
data contain redundant features, fuzzy values and missing values, it is difficult
to predict the degree of malignancy. The previous work attempted to solve this
problem with a fuzzy rule extraction algorithm based on fuzzy min-max neural
networks, whose computation results show it did not treat the missing values
well, and the accuracy can not beyond that of neural networks. Since support
vector machines can obtain better performance than neural networks do, we
combine support vector machines with backward floating search method to
select relevant features and to predict the degree of malignancy. Computation
results show that the feature subset selected by our techniques can yield better
classification performance, In contrast with the base line method, which gen-
erated two rules and got 83.21% accuracy on the whole data set, our method
generates one rule to get 88.21% accuracy.