Feature selection based segmentation of multi-source images : application to brain tumor segmentation in multi-sequence MRI

Nan Zhang

Numéro d'ordre: 2011-ISAL-0079 Année 2011

THÈSE

présentée devant

L'Institut National des Sciences Appliquées de Lyon

pour obtenir

LE GRADE DE DOCTEUR

ÉCOLE DOCTORALE: ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE
FORMATION DOCTORALE : SCIENCES DE L'INFORMATION, DES DISPOSITIFS ET DES SYSTÈMES

par

ZHANG Nan

Feature Selection based Segmentation of Multi-Source Images: Application to Brain Tumor Segmentation in Multi-Sequence MRI

Soutenue le 12 septembre 2011

Jury :

Yue-Min ZHU	Directeur de recherche CNRS	Co-Directeur de thèse
Su RUAN	Professeur	Co-Directeur de thèse
Qing-Min LIAO	Professeur	Co-Directeur de thèse
Pierre BEAUSEROY	Professeur	Rapporteur
Da-Tian YE	Professeur	Rapporteur
Yue-Sheng ZHU	Professeur	Rapporteur

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **http://sakura.cpe.fr/ED206** M. Jean Marc LANCELIN Insa : R. GOURDON	M. Jean Marc LANCELIN Université Claude Bernard Lyon 1 Bât CPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 Fax : lancelin@hikari.cpe.fr
E.E.A.	**ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE** **http://www.insa-lyon.fr/eea** M. Alain NICOLAS Insa : C. PLOSSU ede2a@insa-lyon.fr Secrétariat : M. LABOUNE AM. 64.43 – Fax : 64.54	M. Alain NICOLAS Ecole Centrale de Lyon Bâtiment H9 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60 97 Fax : 04 78 43 37 17 eea@ec-lyon.fr Secrétariat : M.C. HAVGOUDOUKIAN
E2M2	**EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION** **http://biomserv.univ-lyon1.fr/E2M2** M. Jean-Pierre FLANDROIS Insa : H. CHARLES	M. Jean-Pierre FLANDROIS CNRS UMR 5558 Université Claude Bernard Lyon 1 Bât G. Mendel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 04.26 23 59 50 Fax 04 26 23 59 49 06 07 53 89 13 e2m2@biomserv.univ-lyon1.fr
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTE** Sec : Safia Boudjema M. Didier REVEL Insa : M. LAGARDE	M. Didier REVEL Hôpital Cardiologique de Lyon Bâtiment Central 28 Avenue Doyen Lépine 69500 BRON Tél : 04.72.68 49 09 Fax :04 72 35 49 16 Didier.revel@creatis.uni-lyon1.fr
INFOMATHS	**INFORMATIQUE ET MATHEMATIQUES** http://infomaths.univ-lyon1.fr M. Alain MILLE	M. Alain MILLE Université Claude Bernard Lyon 1 LIRIS - INFOMATHS Bâtiment Nautibus 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72. 44 82 94 Fax 04 72 43 13 10 infomaths@bat710.univ-lyon1.fr - alain.mille@liris.cnrs.fr
	MATERIAUX DE LYON	M. Jean Marc PELLETIER

Matériaux	M. Jean Marc PELLETIER Secrétariat : C. BERNAVON 83.85	INSA de Lyon MATEIS Bâtiment Blaise Pascal 7 avenue Jean Capelle 69621 VILLEURBANNE Cédex Tél : 04.72.43 83 18 Fax 04 72 43 85 28 Jean-marc.Pelletier@insa-lyon.fr
MEGA	**MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE** M. Jean Louis GUYADER Secrétariat : M. LABOUNE PM : 71.70 –Fax : 87.12	M. Jean Louis GUYADER INSA de Lyon Laboratoire de Vibrations et Acoustique Bâtiment Antoine de Saint Exupéry 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél :04.72.18.71.70 Fax : 04 72 43 72 37 mega@lva.insa-lyon.fr
ScSo	**ScSo*** M. OBADIA Lionel Insa : J.Y. TOUSSAINT	M. OBADIA Lionel Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.77.23.88 Fax : 04.37.28.04.48 Lionel.Obadia@univ-lyon2.fr

Abstract

Magnetic Resonance Imaging (MRI) is widely applied to the examination and assistant diagnosis of brain tumors owing to its advantages of high resolution to soft tissues and none of radioactive damages to human bodies. Integrated with medical knowledge and clinical experience of themselves, the experienced doctors can obtain the sizes, locations, shapes and other pathological characteristics of brain tumors according to the information in MRI images to make scientific and reasonable therapeutic treatment.

Because there are several MRI examinations for every patient in the whole therapeutic treatment, each of which can give 3-dimensional data in multiple sequences, it is a large amount of data to be dealt with for the doctors. Long time of hard work will inevitably lead to mistakes in the diagnosis of the tumor contours for the doctors. Moreover, it is subjective for the doctors to determine the state of the diseases according to their medical knowledge and clinical experiences. Therefore, developing an automatic or a semi-automatic computer-aided diagnosis system is meaningful in real medical treatments, which can release the workload of doctors and improve the accuracy by giving objective results.

This problem is a hot point in the research field of biomedical engineering and a lot of algorithms have been proposed to try to solve it. But unfortunately it is still unsolved due to the limitations of low accuracy, efficiency, applicability and robustness of existing algorithms. In this thesis, a semi-automatic brain tumor detection and classification framework from tumor region segmentation to tissue classification is proposed with Support Vector Machine (SVM) as its classifier. Through fusion of input data, extraction of feature vectors, feature selection, primary classification of brain tumor and contour refinement, the final tumor detection and classification can be fulfilled. Multi-kernel SVM is also introduced in our proposed system to be fit to multiple MRI sequences and to improve segmentation accuracy. In addition to the multi-kernel SVM, adaptive training is designed to follow-up the changes of tumors during several MRI

examinations. By adaptive training, the system can obtain the properties of tumors after the first detection and classification and then separate the tumors in the subsequent MRI examinations automatically.

The proposed system is evaluated on 13 patients with 24 examinations, including 72 MRI sequences and 1728 images. Compared with the manual traces of doctors as the ground truth, the average classification accuracy reaches 98.9%. The system utilizes several novel feature selection methods to test the integration of feature selection and SVM classifiers. Also compared with the traditional SVM, the neural network and a level set method, the segmentation results and quantitative data analysis demonstrate the effectiveness of our proposed system.

Key words: brain tumor detection; Support Vector Machine (SVM); feature selection; contour refinement; following-up

Table of Content

CHAPTER 1 INTRODUCTION

1.1 Background

With the growing extent of aging population, cancer has become a global public health problem. According to the World Cancer Research Fund's latest statistics, cancer is the world's first cause of death. In the worldwide each year, 12.7 million people are diagnosed with cancer, and 7.6 million people died of cancer [1]. Meanwhile, the annual incidence of cancer continues to rise. By 2030, every year there will be 26 million new cases, and the death toll will reach to 1.7 million people.

As a kind of cancer, brain tumor is a very malignant and harmful disease. It has a high incidence, and high mortality, which ranks the fifth in the whole tumors and is just below the stomach cancer, uterine cancer, breast cancer and esophageal cancer [2-3]. Diagnosis and treatment of brain tumor cost a longer period, usually one examination every a few months. The doctors need to check on the stage of the disease in last examination as referral, and to develop the treatment plan for the next therapy. Under the existing medical conditions, in addition to surgery and radiation therapeutic methods, there are no more effective treatments, and the patients' condition can be alleviated and controlled, but extremely difficult to cure, which will cause great burdens in both mentality and economy for the patients [4-5]. In this sense, the treatment of cancer is a major social problem in both economic and financial aspects. A good solution to this problem will have important social and practical significance.

For early detection and treatment of brain tumors, some medical image-based diagnosis methods are widely applied to clinical practice. Through a variety of medical imaging, the doctors can obtain and understand the patient's condition more clearly and intuitively, and propose various types of scientific and rational treatment plannings for the patients.

Commonly used medical imaging methods are Computed Tomography (CT), Positron Emission Tomography (PET), CT / PET, Magnetic Resonance Imaging (MRI), and so on [6].

CT uses the radioactive rays to penetrate the human body, and the imaging is based on the different characteristics reflecting to the rays of different tissues. PET needs to inject with radioactive drugs in the human body, and the drugs will flow to all the cells, tissues and organs with the blood in the whole body. The absorbed radiation will be metabolized and released by different tissues to form different rays which can be received for a specific imaging. CT/PET refers to the combination of CT and PET scans which are carried out on the same plane to form a fused image with the machine. Both CT and PET examination have radioactive hazards, and PET examinations are too expensive. Compared with all these imaging modalities above, MRI is the most cost-effective [7].

The schematic diagram of MRI equipment and inspection is shown in Figure 1.1 [8].

Figure 1.1 The schematic diagram of MRI equipment and inspection.

In the MRI imaging process there will be not any instruments entering and any medication injected into the human body. There is not any radiation damage to the human body, and the whole process is quite safe. In addition, MRI imaging has high-resolution and accurate positioning of soft tissues, and is sensitive to the characteristics of diseases, thus it is especially suitable for the diagnosis of brain diseases [2, 9].

After obtaining the MRI images, in radiotherapy or surgery situations, the doctors need to clearly grasp the patient's condition according to the relevant images and to fully prepare for the disease treatment based on the medical information. General speaking, according to the obtained MRI images and their clinic knowledges and experiences, physicians can fulfill the manual segmentation of images to reasonably grasp the size, shape, location, structure distribution and other lesions of the tumor regions [10].

However, under the existing medical conditions, all of the above work can only be done manually by the doctors themselves. On one hand, this is undoubtedly a great burden for the doctors to work on a large number of MRI examination data from many patients. Doctors' hard work will inevitably make errors in the classification results. On the other hand, the manual segmentation results are very subjective, that is to say, different doctors may lead to significantly different segmentation results, and the segmentation results from the same doctor but at different moments may be also

different, which undoubtedly brings a certain degree of risk to the diagnosis of the patients.

There are mainly two common types of brain tumors, meningiomas and gliomas. The characteristics of meningioma are relatively simple on the medical images, in which there is a clear distinction border between normal tissues and tumors. So meningiomas is easy to be segmented and removed directly through the surgical operations. The characteristics of gliomas are invasive, that is, tumors and other abnormal tissues will be invasive and spread into normal tissues to some degree, and all the tissues are close mixed together, so it is very difficult to clearly distinguish them manually by the doctors. Common practice in real clinical surgery is to appropriately excise the determined tumor tissues through the operation, and then to extract the remaining tissues and cells near the cut boundary to accomplish the pathological analysis in order to test the components and the proportion of cancer cells, which determines the surgical excision extent of the abnormal tissues and the treatment program of the next step [8-10].

The detection and segmentation of brain tumors are of great significance, and also there are problems and risks in the process. In this paper will research on the detection and analysis of the complex glioma and present a semi-automatic segmentation and detection system on brain tumors as to the difficulties in the current computer-aided analysis means. The framework can decrease the degree of interaction from human bodies, reduce the workload of doctors, establish a sound mechanism for medical image processing using medical knowledge, and provide a relatively accurate results which can be accepted by most physicians as objective references to assist doctors to diagnose and treat disease. Meanwhile, the system can adaptively track the patient's condition effectively to assess the scientificalness and rationality of the treatment. If a large number of clinical data can be used for statistical analysis and experimental validation, the system also has practical significance.

1.2 Magnetic Resonance Imaging Theory

Nucleus in the tissues of human body is composed of protons and neutrons. Protons have a positive charge, and neutrons are not charged, thus the whole nucleus is shown as positively charged. Charged nucleus rotates around its own axis rapidly, and from the law of electromagnetic induction, high-speed spin will produce a vertical magnetic moment. According to the different natures of spin, the nuclei can be divided into two different types: magnetic nuclei and nonmagnetic nuclei. The former denotes the nuclei which can generate a magnetic moment by the spin, and the latter denotes those which can not generate a magnetic moment. If the numbers of protons or neutrons in a nucleus is odd, then the nucleus belongs to the magnetic nuclei; otherwise the nucleus is non-magnetic nuclei if and only if the numbers of protons and neutrons inside an atomic nucleus are both even. Only the magnetic nuclei can be used for the magnetic

resonance imaging [11].

In theory, the magnetic neclei of all elements can be used for MRI imaging, but in the current imaging mode, only hydrogen nuclei just with one neutron proton (not the other isotope of hydrogen atoms) are mainly used. This is mainly due to [12]:

1. in the body hydrogen atom has the highest molar concentration: the human body contains a lot of water and hydrocarbons (such as sugar, protein and fat, etc), so the amount of hydrogen nuclei in the human body is the highest, more than 2/3 of the total human nuclei. Sufficient hydrogen nuclei will be very beneficial in generating a strong magnetic resonance imaging signal, and the actual magnetic resonance imaging also mainly relies on hydrogen nuclei in the water and fat.

2. hydrogen nuclei has the maximal susceptibility: compared with the nuclei of other elements in the human body, the susceptibility of the hydrogen nuclei is on the highest level. Higher degree of magnetization will generate higher signal intensity in the magnetic resonance imaging.

3. the biological characteristics of hydrogen nuclei are obvious: because there are a lot of water and hydrocarbons in the body, hydrogen nuclei exist in the various tissues of the body. Using hydrogen nuclei for magnetic resonance imaging can describe the different characteristics of a variety of different tissues in the human body with images and thus hydrogen nuclei can lead to a strong biological representation.

The typical MRI imaging system mainly consists of five parts: the main magnet, gradient systems, RF system, computer systems and other auxiliary equipment [11], as shown in Figure 1.2 [13].

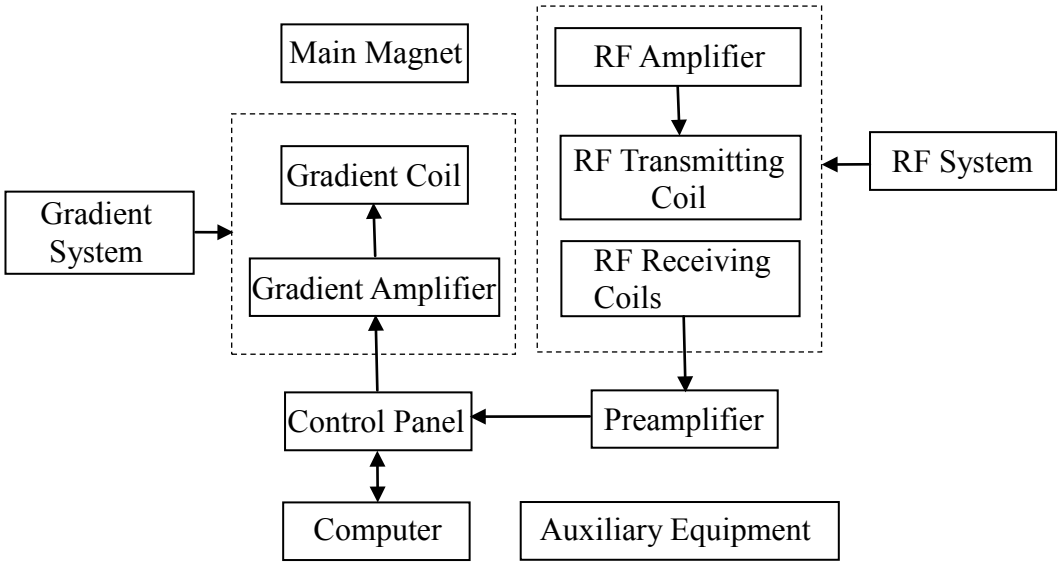


Figure 1.2 Structure of MRI imaging System

The main magnet is a device to produce a magnetic field, and its performance will directly affect the imaging quality of MRI. Generally speaking, the main magnet can be divided into two defferent types: permanent magnet (made of permanent magnet

material) and electromagnetic magnet (twined by coils). Gradient system is used to generate the linear gradient magnetic field, to code the spatial orientation of MRI signal and the switch of gradient field will generate the MR echo. RF system mainly consists of Radio Frequency coils which are used to receive the generated MR echo. Computer system controls the entire operations of the MRI imaging system, including pulse excitation, signal acquisition, data operations and image display. Supplemented by some multi-function softwares, computer system can obtain other complex capabilities, such as data mining and analysis, three-dimensional modeling and so on. Auxiliary system mainly refers to the auxiliary equipments that ensure the normal operation of the MRI imaging system [11].

The detailed imaging process of MRI can be described as:

After the human body enters the main magnetic field, it can produce magnetization phenomena: the magnetic moments from the spin of hydrogen nuclei themselves will deflect along the orientation of main magnetic field. Most magnetic moments will be forward the direction of the magnetic field and into a low-level energy state, and a small number of hydrogen nuclei will be against the direction of the magnetic field and remain high-level energy state. The total magnetic moment vector (called the longitudinal macro-magnetic moment) maintains the same direction with the main magnetic field.

Macro-magnetic moments of different tissues are concerned with the number of hydrogen nuclei in tissues. The higher the content of hydrogen nuclei is, the stronger the longitudinal macro-magnetic moment will be produced. However, the level of macro-magnetic moment vector is still much smaller than that of the main magnetic field, and both of them are in the same direction. Superimposed with each other, the macro-magnetic moment vector will be completely submerged in the main magnetic field. The strength of the pulse signal released from the longitudinal magnetic moment is too weak to be received by the RF coil. Thus it is not feasible to distinguish different tissues according to different levels of macro vectors caused by differnet amount of hydrogen nuclei.

In order to distinguish different tissues, RF pulses are applied by the imaging system to the human body in the main magnetic field. The energy of pulses will be delivered to the hydrogen nuclei in a low energy level to make them achieve the transition to the high energy level by energy absorption. This phenomenon is the so called magnetic resonance phenomenon. The macro-magnetic moment vector deflects under the input pulses. The stronger the energy of RF pulse is, the greater the angle of deflection is. After the absorption, the energy of hydrogen nuclei is in a high and unstable level. Once the RF pulse is removed, the energy will be released automatically and the hydrogen nuclei will return to the initial steady state. In the recovery process, electromagnetic pulses will be released and received by the RF coil for imaging.

The recovery process of macro-magnetic moment vector after the disappearance of pulses is also called the relaxation process, which is divided into two types, namely,

lateral relaxation (transverse magnetization component decreases to disappearance) and longitudinal relaxation (longitudinal magnetization component increases gradually back to the initial value). Both of the relaxation correspond to T2 and T1 characteristics of the tissues respectively. Different tissues have a significant difference in characteristics of T2 and T1 [11-12].

Given different excitation pulse sequences, MRI can form different weighted sequences to reflect the different characteristics and biological properties of tissues. Commonly used MRI sequences are T1-weighted, T2-weighted, PD-weighted (Proton Density) and FLAIR (FLuid Attenuated Inversion Recovery) weighted sequences. T1-weighted images reflect the ability about the recovery speed of magnetic moment in the longitudinal relaxation, while T2-weighted images reflect the decay speed of magnetic moment in the transverse relaxation. PD-weighted images reflect the differences among protons (hydrogen nuclei) density (content per unit volume). After the suppression of free water in the T2 weighted images to get better display of the tumor characteristics, FLAIR-weighted images can be obtained [11-12].

Examples of T1-weighted, T2 weighted and FLAIR-weighted images of the same patient in a same imaging process are shown in Figure 1.3 [2].

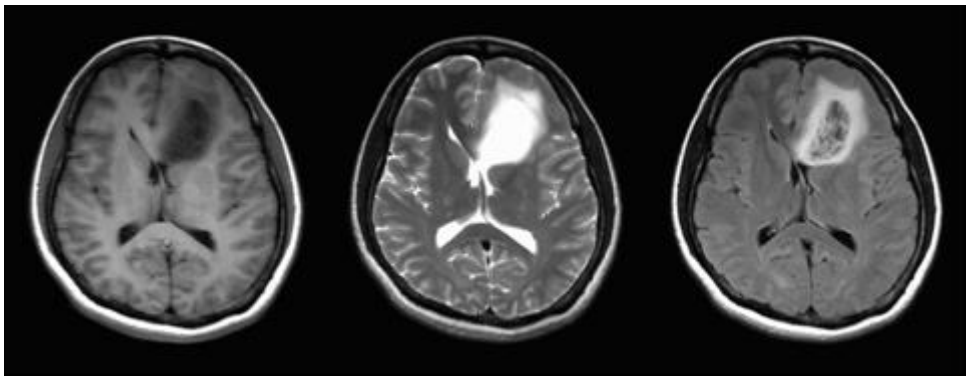


Figure 1.3 Examples of MRI weighted images

(from left to right: T1-weighted, T2 weighted and FLAIR-weighted images)

MRI imaging sequences are composed of multiple slices, of which the positions and thickness can be chosen randomly, as shown in Figure 1.4 [14]. The red, blue and green rectangles refer to commonly used imaging directions to the MRI slices. Different weighted image sequences contain a different number of slices. Generally speaking, T1-weighted sequence contains the most slices (usually 124 slices), while T2 weighted, PD-weighted and FLAIR-weighted sequences usually contain the same number of images as 24.

Brain tissues in MRI images can always be divided into two main types: normal tissues, including gray matter, white matter and cerebrospinal fluid (CSF), and abnormal tissues, usually containing tumor, necrosis, cystic degeneration and edema. Necrosis is in the tumor region in general, while generally cystic change and edema are located near the tumor border. These two types of tissues often overlap with the normal tissues and

they are not easily to be distinguished. The gray contrast of major tissues in different MRI sequences is shown in Table 1.1 [2].

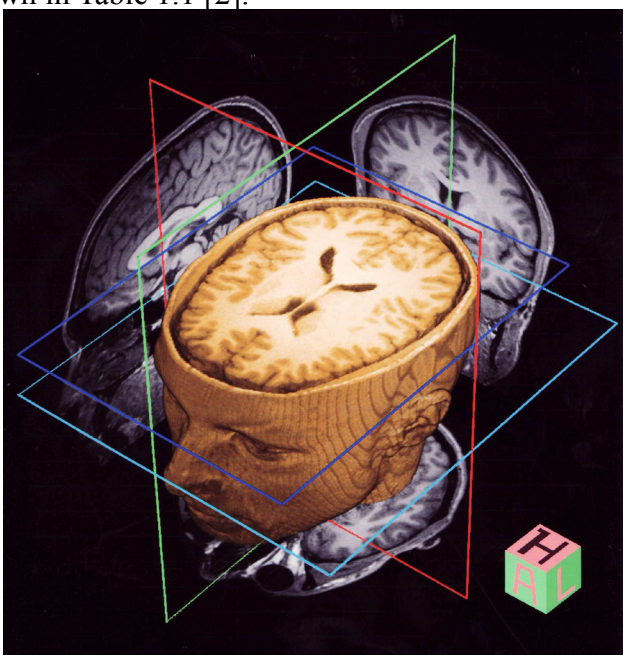


Figure 1.4 The selection directions of MRI slices

Table 1.1 The gray distribution of different tissues in multiple MRI sequences

	Gray Matter	White Matter	CSF	Tumor	Edema
T1-weighted	−+	+	−−−	−	−−
T2-weighted	+−	−−	++	+	+
FLAIR	+−	−−	−−	+	−

In Table 1.1, the negative sign denotes the low gray value, and the positive sign denotes the high gray value. The larger the number of negative signs is, the darker the relevant region is. The corresponding region is much lighter while the number of the positive sign is larger. It should be noted that Table 1.1 can just indicate the gray value contrast in the internal tissues of the same weighted sequence, and different MRI sequences can not be comparable due to the different imaging parameters, imaging environment, imaging equipment and the patient's specific case. In addition, the tumor area typically includes a variety of abnormal tissues, so its intensity distribution will be uneven, which can also be used as one of the prerequisites for the classification of the internal tissues of the tumor region [2].

1.3 Related Works

MRI images can reflect details of different features to provide an important basis for the diagnosis and treatment of illness for patients. However, there are still some restrictions in computer-based analysis of MRI medical images, such as: the differences of imaging equipment, imaging environment and imaging parameters among patients; the redundancy, noise and other interference factors inevitably from the formation of the images; the large amount of image data from multiple sequences to be dealt with; the uniform patients's conditions among large individuals; the lack of available priori knowledge; the complex structure of MRI images, including different tissues in the internal and external region of the tumors; and the lack of clarity aliasing of the tumor borders due to the the invasive characteristics of gliomas.

The variety of restriction above limit the degree of automation in a MRI-based medical image processing systems, that is to say, to develop an accurate and efficient fully automatic system is almost impossible. The issue is also a hot and difficult problem in the international interdisciplinary fields of medicine and information, and it has been active in the forefront of research. A lot of work around this issue are carried out, and a variety of innovative or improved methods are conducted to try to solve this problem.

At present a successful method is to develop a semi-automatic system with the combination of human interaction, which can improve the automatic level of the system on one hand, and reduce the excessive interaction involved on the other hand. Through the auxiliary means (pre-or post-processing steps) the system can take full advantages of multi-modality MRI sequence data and directly optimize the classifiers to improve classification accuracy and efficiency.

1.3.1 Basic Image Processing Methods

The traditional analysis on MRI images can ignore the medical information implied in the images, and deal with them directly as a general image processing problem to operate. Corresponding to each two-dimensional MRI slice, using some image processing methods to grasp regions with the same or similar characteristics to achieve the separation. Regions with coherent characteristics are not necessarily consecutive, and there may be the case where the region is composed of blocks. And other regions of different characteristics are distinguished from other parts to complete the preliminary analysis of the two-dimensional slice. It can be observed that, the traditional tumor detection in MRI images is a typical image analysis problem.

A variety of traditional image processing algorithms can be applied to try to resolve this problem. The most common algorithm is the threshold-based method [15], which is always integrated with histogram analysis [16] to first obtain the overall histogram distribution of the image, then to select the appropriate threshold based on the distribution of the binary image, and finally to finish segmenting the image area

supplemented post-processing processings of morphological operations [16], such as hole filling, boundary improvement, and so on. However, this method is too simple. If the histogram of the image does not contain obvious peaks or valleys, it will be difficult to select the optimal threshold. A commonly automatic threshold selection method is called Otsu Threshold [17], but its accuracy can not still reach a higher level, that is because the algorithm itself is too simple and can not be adaptive to the complex situations of images.

Considering the characteristics of the the whole tumor area, such as the similarity, uniformity or gradual change among the gray levels, region segmentation methods can be used to achieve the detection of tumors. Commonly used algorithms are region growing algorithm and regional separation and merger algorithm [18] .

Region growing algorithm requires a certain amount of pre-selected seed points, then the growth area is expanded by determining the similarity of the gray values of the pixels within the scope of the seed points, and ultimately achieves the purpose of image segmentation. Algorithm is limited by the stopping criteria. In general, if the criterion does not meet the conditions of similarity, the algorithm will automatically stop. But in MRI images, the characteristics of the tumor borders and normal tissues are closely similar, which leads to the situation that the similarity condition of neighborhood and the seed points can always be satisfied. Therefore, it is generally necessary to manually set more stringent criteria to make the algorithm stop.

Regional separation and merger algorithm extends the seed points to the seed regions and the segmentation algorithm is just the opposite with the region growing processing. First, the initial image is randomly divided into a number of non-overlapping regions, then these areas will continue to be divided into much minor parts more detailedly, and finally these similar parts are merged together in accordance with the very close similarity criterion. Because the image has been precisely refined to the small extent, the characteristics of the adjacent pixels and regions are much closer to each other, and the merger operation will be more accurate and simple. But the algorithm does not limit the detailed extent of the image segmentation, and the merger guidelines are also required artificial selection, all of which will influence the algorithm's performance to a certain extent.

Altas-based segmentation methods, also called registration-based methods, utilize the same area characteristics of the images. The normal human brain tissue indicates a symmetric structure, therefore the algorithm pre-create multiple templates to construct a template library. The image needed segmenting will be registered to the templates one by one through different linear, non-linear or combination maps to establish the corresponding relationship between the segmented image and the templates, so as to achieve the purpose of segmentation and classification [19-22]. Registration-based method is more suitable for segmentation of normal brain tissues, but because of the need to establish templates in advance, the algorithm is sensitive to the initial templates, the accuracy of which will influence the segmentation accuracy to a great extent. That is

the limitation of this type of algorithms.

Regional algorithms use the approximate characteristics (similarity) between adjacent pixels, and oppositely the corresponding boundary algorithms use the properties of huge differences around the boundary pixels, such as the jumping intensity, complex contour, gray gradient, the level of the frequency spectrum, and so on. Commonly used boundary detection algorithms are based on a variety of edge detection operators [16], such as the Canny operator, Sobel operator and so on. These operators are equal to discrete square templates with variable sizes of 3×3, 5×5, etc. The templates move pixel by pixel in the entire image, and the combination of the central pixels of the templates which correspond to the convolution result is the relevant border. Edge detection always obtains discrete contour points, and we must utilize a variety of connectivity techniques to get a complete outline of the border. Various edge detection operators have their advantages and disadvantages themselves, but all of them are generally sensitive to noise and slightly lack of accuracy.

1.3.2 Basic Pattern Recognition Methods

The basic elements of MRI images are the pixels, and each pixel contains a variety of features about image properties, including the basic gray-scale [23], a variety of features derived from the characteristics of the basic features, such as texture features (including mean value, standard deviation, and the commonly used Gray Level Concurrence Matrix (GLCM) to reflect statistical characteristics [24-25], and so on), mathematical transformation features (such as wavelet transform features [26-28]) and so on. Various features are related to the corresponding meaningful information in the relevant field, which can indicate a specific physical meaning through certain numerical values and its scope of change. Pattern recognition algorithms research on the pixels and include features as the study objectives. By analyzing and comparing different characteristics, the corresponding pixels are classified into different categories.

Commonly used pattern recognition algorithms are clustering, Bayesian probability model, linear and nonlinear discriminant classification methods [29]. Clustering mainly uses the gray values of the image, and merges data points of different types according to the clustering criteria based on similarity. A typical unsupervised clustering algorithm is the nearest neighbor method, in which the data is clustered into the same category as that of the nearest point. The improved algorithms are K-nearest neighbor (KNN) method (the data is clustered into the same category as that of the K nearest points), edited nearest neighbor (nearest neighbor method becomes to a two-step process: the first step is the pre-classification of data to remove the misclassification data, and the next step is to classify using the KNN criterion to improve accuracy), and so on. Nearest neighbor method is relatively simple: one point is just compared with another data point in classification. The algorithm is not reliable enough and can not handle the situation of complex boundary, or of points very far from the center of the class when the distribution of the elements is long and narrow, the same as some types of its improved

algorithms.

Another commonly used clustering algorithm is C-means clustering (C-means), in which all the centers of the classes are prefixed in advance. During classification, the distances between each data point and all the centers are computed, and the data is classified using the nearset distance criterion. Class centers will automatically be updated as the change of the classification of data sets and its containing elements. The limitations of C-means clustering is that the number of classes needs to be pre-set and can not be changed in the clustering process. This algorithm is relatively simple and can not handle the situation of the existence of outliers (data is very far from the center of this class and is wrong classified). The improved algorithms of C-means clustering, such as ISODATA, have also been widely applied [29].

Bayesian probability model-based classification is to cluster data points into the corresponding categories by calculating the posterior probability of data points, and decision criterion used is always the minimal risk criterion. This algorithm can be easily extended to the case of multi-dimensional data and multi-class classification, but the number of data sets and the probability distribution of all the elements are needed to be known in advance, otherwise it is impossible to calculate the probability.

Linear discriminant algorithms mainly refer to that the classification surface is a plane or a hyperplane, which can be described using the linear equations. Given input data set, the algorithm can obtain the normal vector and offset intercept vector of the plane by optimizing the equation of the plane. Commonly used linear discriminant criterion is Fisher criterion, in which the linear plane equation is optimized based on the mean value, the between-scatter and within-scatter among various classes of samples. Linear discriminant function is simple, and the direct use of linear discriminant classification usually tends to cause great errors. Fisher criterion is carried out in the original input space, therefore, it is difficult to optimize the corresponding surface equation for linear inseparable data in the space.

In complex situations, it is necessary to extend the linear discriminant method to a nonlinear discriminant one in order to adapt other linearly inseparable cases with overlapping elements. Nonlinear discriminant criterion is commonly piecewise linear discriminant method, the classification hyperplane of which is composed of several sub-planes, each of which corresponds to a linear separable subset. It is necessary to know the number of classes in piecewise linear function to better design the classifier. Directly using the high-order discriminant function can also solve the situation that the input data are not linear separable, but the calculation process and the classification surface are too complex.

1.3.3 Fuzzy Theory-based Method

By introdung membership function into the traditional set theory, we can obtain the Fuzzy Set Theory [30]. Membership function indicates the degree of the elements belonging to a particular class. The same element can belong to different categories in

different levels, and the sum of the corresponding values of all membership functions is 1. Finally the element is determined that belongs to the category which has the largest value of the membership function. That is the classificaiton criterion in fuzzy set theory-based algorithms.

Introduing the concept of fuzzy set theory into the field of traditional pattern recognition and integrating the traditional pattern classification methods and fuzzy theory, a new kind of classification method can be produced [24]. Commonly used methods are fuzzy C-means clustering (FCM), fuzzy K nearest neighbor method, fuzzy connectedness, and so on.

FCM algorithm transforms the hard decisions in the traditional C-means clustering to fuzzy decisions, and the value of membership function and the center of each category need to optimize together, that is, the updating process combines the classified elements and membership function instead of the traditional operations which just rely on the data points [31]. Similar to C-means clustering, the number of classes needs to be pre-set, and the update process may be not necessarily able to converge, in which the stop criteria may need to be set manually. FCM can obtain fuzzy results, therefore it is necessary to add deblurring treatment to gain the final output. If a broader membership condition can be introduced (the sum of the values of the membership functions, which corresponds to different categories, of the same element is greater than 1 after the update process), it can improve the FCM clustering algorithm. Relaxing the conditions to a broader one will lead to the decrease of the sensitivity on the pre-set number of classes, but the algorithm is still sensitive to the initial cluster centers.

Membership function used in Fuzzy K-nearest neighbor method is equivalent to the weights on the K nearest surrounding elements, and the weights depend on the distances between the elements.

Fuzzy connectedness algorithm is often combined with the region growing algorithms [32-34], which is equivalent to use the membership function to weighting on the similarity of the pixels in region growing. The procedure of the algorithm is as follows: first, select the seed points, and calculate the fuzzy connectedness between each pixel and each seed point with membership function; then use the region growing algorithm for image segmentation. Although the measure of similarity among the elements is more reasonable after introducing the fuzzy theory, but the algorithm does not significantly improve the sensitivity to the initial values, and the initial choice of seeds still play a greater role on the algorithm's performance.

In addition to integrating with the clustering methods, fuzzy set theory can also be used in combination with the Markov Random Field theory (MRF), in which the image segmentation is based on the gray values of pixels of their own and the impact of neighborhood information of pixels simultaneously [35-37], that is so called fuzzy Markov Random Fields (fMRF). The algorithm is now used to solve the classification of normal tissues in T1-weighted MRI images [38]. As the preliminary work of distinguishing the tumor, the fuzzy Markov Random Field algorithm can divide the

T1-weighted images into several connected isotropic regions in order to achieve the purpose of distinguishing normal tissues. As to MRI images of multiple sequences and multi-band, we can apply different membership functions to different sequences, in order to improve the accuracy of MRI image segmentation [39-40]. Through the fusion methods, the best results of each band are synthesized to achieve the global optimum, and the fusion theory can also be used in multi-kernel SVM.

1.3.4 Deformable Model-based Method

In pattern classification problems, using the relevant constraints within the image, the size, location, shape of the object to be split as prior knowledge, combining with the overall and regional boundary properties of tissues in the image, and building the appropriate model to improve the classification accuracy is the basic idea of deformable model-based classification methods [41-42]. Deformable model approaches are more suitable to analyze the tissues with complex boundary, lower contour smoothness and various properties in different individuals which meanwhile vary significantly with time. Currently, the deformable model based method has received a wide range of applications in medical image analysis processing [41,43-50].

In accordance with its principles, the deformable model based method can be divided into two main categories, namely parametric deformation model (based on the parameters of image) and geometric deformation model (based on the geometric properties of the image) [51]. Parametric deformation model needs to be given the initial curve profile and be defined the energy function. The energy funtion is composed of two parts: the internal energy and external energy, the former is able to control the smoothness of the curve and and the latter facilitate the evolving of curve. Optimization process of the energy function curve is the evolving process, the final convergence of the curve is the boundary contour corresponding to the regions to be segmented. Commonly used Active Contour Model (ACM), such as the Snake algorithm, is a typical parametric deformable model, of which the limitations is its sensitivity to the initial curve and that the convergence process maybe need to manually set the stop criterion to control the speed of convergence, and that ultimately the convergence results may fall into the local optimum. Considering the limitations of the Snake algorithm, some scholars propose the improved Snake algorithm [52] and Global Active Contour Model algorithm (Global ACM) [53-55]. Both of them are not sensitive to the initial values, and the Global ACM algorithm converts the energy function to be optimized into a convex function, which means the function extreme values surely contain the global optimum value in the mathematical sense.

Geometric features can be used as prior information in Geometric Deformable Models. Using the prerequisite that the normal tissues are symmetric in the brain, the corresponding geometric features and structural information can be extracted [56], which is a basis to complete and improve the segmentation accuracy. Commonly used Geometric Deformable Model algorithm is Level Set algorithm, which has been widely

used in medical image processing [57-64]. The principle of Level Set algorithm is to assume the region to be separated as the cross section of a high dimensional surface, which is described by analytic equations (mostly implicit partial differential equations). When the surface's equation is 0, the solution of the equation will corresopond to the boundary curves of the region to be split. The solution set is also called the zero level set. In the whole process, the region to be segmented has always been maintained as the zero level set of a high -dimensional surface. Level Set algorithm transforms the curve evolution in a plane to surface evaluation in high-dimensional space, and the continuous deformation of the cross-section boundary curve will converge to the final optimal contour border. In Level Set Algorithm the surface equation is given by the implicit functions, therefore this non-parametric method is not sensitive to the initial values and more suitable for the case of topology change in tumor detection. Surface evolution is also an optimization process, the speed of which is controlled by the image information. Generally speaking, one surface can only be used for the evaluation of the boundary curve of one region.

Deformable model-based method can be used in conjunction with each other to construct the hybrid model of better performance [64] or in integration with other theories to apply to different applications, such as face detection [65], three-dimensional reconstruction [66], etc.

1.3.5 Tumor Detection in Multiple MRI sequences

Using the different characteristics and information provided by the multi-spectral or multi-sequence MRI images, data can be effectively fused in the data layer [67-71] and decision-making layer [72-74] to extract features for tumor detection. Algorithms in the decision-making layer are also applied in face recognition and multi-modality data fusion [75-78]. Although the data fusion will introduce more redundancy, noise and increase the computation time, sufficient data can reduce the randomness and arbitrariness of classification in order to improve the quality of cancer detection [67, 69-71,79-81].

In [67], according to the fuzzy set theory, tumor information in different sequences are modeled using the membership functions on each MRI sequence, and then all the obtained fuzzy sets are segmented together. The difficult point of this algorithm lies in the selection of membership function. A parametric smoothing model is proposed in [69], and the intensities of each category in T1-weighted images, T2-weighted images and PD-weighted images are handled by the Expectation Maximization (EM) algorithm. The final tumor segmentation results are synthesized by the fusion of models of each category. A hierarchical genetic clustering algorithm is exploited in [70], which integrates the fuzzy learning vector quantization algorithm, called Hierarchical Genetic Algorithm with a fuzzy Learning Vector Quantization network (HGALVQ), to deal with T1-weighted images, T2-weighted images and PD-weighted images. The algorithm is optimized by hierarchical genetic algorithm, and the optimization criterion is selected as

the minimization the weighted error value and the complexity of local competitive network, the former of which is defined as the mean distance between the feature vector and its corresponding original image, and the latter of which is number of active nodes in the network. But the converged number of categories in algorithm is very large. In [71], the PSPTA algorithm (Piecewise Triangular Prism Surface Area) is used to extract fractal features, the Self-Organizing Map (SOM) [79] is used to achieve the feature fusion. However, the accuracy of the system changes in a wide range so it is not very robust. In [80], a algorithm composed of a probabilistic model and active contour models for segmentation is applied to deal with T1-weighted images, T2-weighted images and T1-enhanced images. The algorithm relies on the extraction of the multi-dimensional features and the brief description of the natural information in the feature, but some of the assumptions used can not apply to all patients. In [81], two kinds of features, gray values of tissues and prior probability based on the alignment, are extracted by registration with the templates, and then data from manual separation are used as prior knowledge to train and learn a statistical classification model. The algorithm is very sensitive to the initial value of manual segmentation.

In order to achieve better performance, the methods described above are often used in combination [64,82-84], for example, in [64], a hybrid deformation model combined with shape, texture, image models and learning algorithm is proposed to track the changes of heart and brain tumors. The paper [82] proposes a geometric probability model which is combined with registration and spatial prior knowledge to detect a variety of tumors and tissues. The paper [83] proposes a two-step algorithm for the detection and classification of different kinds of tumor in clinic practice. The initial segmentation region is obtained by the fuzzy set theory, and its contour is improved by a deformation model under some strict spatial constraints. In [84], a altas-based flexible transformation is proposed, which is the effective integration of registration-based method and model transformation method.

In addition to the combination of methods and theories, some new algorithms are gradually began to use [85-86]. For example, in [85], the Fluid Vector Flow algorithm (FVF) is proposed as a kind of active contour models, which can be used to detect a large concave area and has been used for tumor tracking in 2D image. But the algorithm is time-consuming and unefficient. In [86], a method to transform the direct detection of tumors is proposed, in which the probability map of brain images is calculated first based on the pattern matching under the nearest neighbor criterion, and then the segmentation of original input gray image is transformed to the segmentation of the probability map .

The methods described above require some prior knowledge or related assumptions, which can not work for the tumor detection of all the patients, such as the shape, size, intensity, texture and location of the tumor. Furthermore, all the information above can not be obtained in advance in the first examination of the patient, which will increase the difficulty of tracking the tumor.

The Support Vector Machine (SVM) is a successfully parametric method, which has been widely used to get accurate results in many multiple-class pattern recognition applications. As introduced in [87-88], SVM, which fits to classify data of high dimensions and from multiple sources particularly, extends the use of kernels which are crucial to incorporate priori knowledge into practical applications.

A simplification is one-class SVM, which is derived from two-class situation SVM; in one-class SVM the training points just involve the class to be separated from others [89]. Other recent developments have shown the benefits of multi-kernel SVM [90-91]. Multi-kernel SVM has more potential for fusion of the data from heterogeneous sources at the expense of computation complexity. It has been proven in [92] that, when the kernel function can be decomposed into a large sum of individual and basic kernels which can be embedded in a directed acyclic graph, the penalty functions can be explored by sparsity-inducing norms such as the l^1-norm.

1.4 Work of This Paper

Currently, there are still many outstanding issues in the medical images-based, particularly MRI images-based brain tumor detection, such as the huge amount of data in MRI examination, heavy burden of the computation, data of high dimensionality, complex characteristics of tissues, the difficulty in classifier design, large differences among individual patients, the infeasibility of establishing a fully automated system, no prior knowledge of the patient's condition to use, a variety of tissues in the tumor region, the overlapping of the boundaries between normal tissues and abnormal tissues which is difficult to be distinguished and so on.

As to these problems and some limitations of existing methods, this paper proposes a semi-automatic SVM-based system to achieve the detection and track of the brain tumors, which can classifies the tumors and abnormal tissues gradually. The system consists of three subsystems, including three main abilities of classification, contour refinement and tracking. The input data of the system are the fusion of T2-weighted images, PD-weighted images and FLAIR-weighted images. The feature vectors are extracted using the characteristics of the weighted sequence of their own and that of correlation between several sequences. Feature selection can fuse and reduce the data effectively and multi-kernel SVM is applied for a variety of data input to improve the accuracy of tumor detection. At the same time the system can effectively track the condition of patients after the first detection of tumors and give reasonable diagnosis conclusion as a reference.

This paper contains five sections, which is organized as follows: Chapter 1 is the introduction, which introduces the background and significance of the subject, the principle and theory of magnetic resonance imaging, and the summary of existing tumor detection algorithms. Chapter 2 proposes a method to integrate the feature selection and classifier's design, and describes in detail all the feature selection methods used in the

experiments of this paper. Chapter 3 describes the proposed tumor detection system and the three subsystems of classification, contour refinement and trackingare also described in detail. Chapter 4 lists the results of this paper and discusses a variety of situations, including the validity of features, the selection of key parameters, the comparison among methods, and the evaluation of the system's performance. Chapter 5 concludes the full text, proposes the improvement goals of the system and finally states the future development of this subject.

CHAPTER 2 FEATURE SELECTION

2.1 Feature Selection Theory

Feature selection is an important data processing step in pattern recognition and classification problems, and it has been successfully applied to many fields [93-97]. The procedure of feature selection is as follows: in accordance with pre-designed selection criteria, the most important features of the given input data are selected by the optimal operations under the prefixed criterion and the remaining features are removed from the input to reduce the data amount.

Although the applied classification fields are different, the ultimately obtained signal is overlapped with a lot of interference and noise in the specific pattern recognition and classification problems, due to the inevitable signal loss and introduced interference in the process of equipment's acquisition, imaging environment, data transmission and conversion. In addition to the traditional denoising algorithms, feature selection, as an important pre-processing operation, plays an important role in the data pre-processing:

1. Feature selection can eliminate the redundancy, interference, noise and less important data in input. According to the definition of feature selection, this process can choose the data based on some certain criteria to eliminate all the factors which are not relative to the classification problems, and to effectively fuse the important data and therefore greatly reduce the data amount.
2. Feature selection can improve the accuracy of the classifier. After the feature selection operation, a large number of non-relevant data which contain many interferential components are removed. Only the most important features are remained for training, which makes the obtained classification model much better to improve the applicability of the model and its ability for solving the problem, and finally to achieve higher classification accuracy.
3. Feature selection can improve the operational efficiency. After the feature selection, the training sample data greatly decrease and the computational complexity is reduced in a relatively lower degree (mainly determined by the algorithm, thus changes on the computational complexity by reducing the amount of data is just in a relative sense) to reduce the computation time.

Because of the importance of feature selection in the pattern classification problems feature selection has been a hot point and difficult issues in this research area. How to determine an effective feature selection criterion, to extract the important features and to verify the validity, reliability, applicability, robustness and computational

efficiency of the criterion by experiments are the key problems to be solved [98-103].

Commonly used feature selection framework are shown in Figure 2.1 [95].

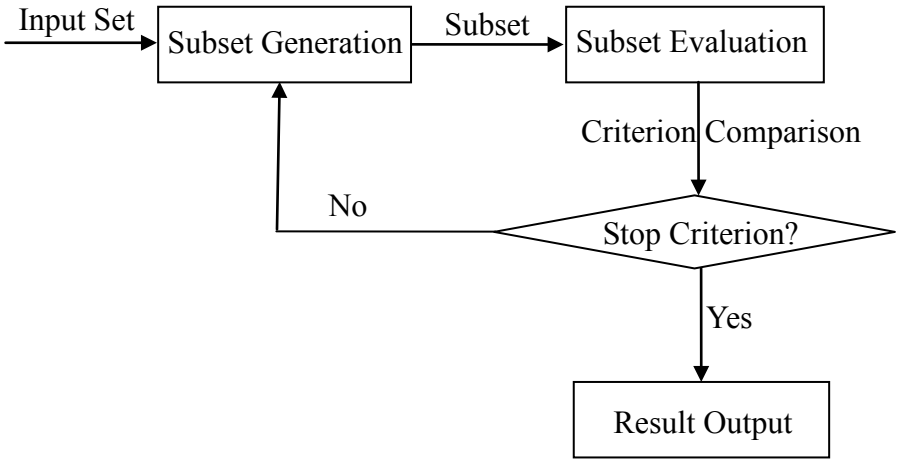


Figure 2.1 A typical framework of feature selection.

A typical feature selection process is divided into 3 steps [95]:

1. Subset Generation: this is the process to conditionally extract the feature subset from the training feature vector matrix prepared for being analyzed by the classifier or a single vector according to some certain criteria. The feature subset will be considered as the result of one step of the feature selection operation to evaluate its performance of feature selection.

2. Subset Evaluation: feature selection can be fulfilled in two directions: one is to begin with a feature subset which contains just one element and to increase the capacity of the subset element by element; the other is to begin with a universal set which contains all the elements of the feature subset and to decrease the capacity of the subset element by element. No matter what kind of direction is used for the feature selection, the effectiveness of the current subset is needed to evaluate once the feature subset changes in a time. The evaluation process of feature selection is to be completed by the pre-set criterion.

3. The determination of the stop criterion: each feature subset after assessing is needed to be compared with the stopping criterion to verify if the characteristics of the current subset have attained a pre-set standard. If so, feature selection will stop automatically and the current subset will be considered as the final output; otherwise this process will continue to repeat again and again until a feature subset which meets the stopping criterion appears. This is an optimization process, and the algorithm sometimes does not automatically converge due to the restrictions of the feature selection criteria and the stopping criteria. Therefore, the stopping criteria maybe need to be set manually (such as setting a limited number of iteration steps). This also indicates that the feature selection criteria and the stopping criteria

will significantly affect and restrict the efficiency, performance and accuracy of the feature selection operations.

As a number of MRI sequences present a three-dimensional structure and each MRI sequence also contains multiple images, which leads to a large amount of data, dimensionality reduction of data by feature selection is particularly necessary. In this paper, a SVM-based tumor detection framework is designed to analysis and deal with the MRI sequences. The difficulties lie in the design of the SVM classifier, the selection of the kernel function and its parameters. Therefore, the algorithm to integrate the feature selection and classifier design is proposed in this paper as a pre-processing step, the key of which is trying to improve the accuracy and efficiency of the system while reducing the dependence of SVM to the kernel function parameter, the difficulty of classifier design and the influence of parameters on the classifier design by the feature selection process.

For these purposes of the effective fusion of data and the verification of the proposed algorithm to reduce the difficulties of parameter selection in classifier design by the feature selection process, we selected 6 novel feature selection methods in the experimentations of tumor detection. The performances and effectivenessof the feature selection methods above is evaluated and compared by visual results and quantitative analysis.

The feature selection methods will be introduces in detail in the next section.

2.2 Feature Selection Algorithms

2.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a classic feature selection method [104], which uses the projection transformation to project the input data from the original space into the new space. The axises in the new space correspond to the projection direction and all the axes maintain the orthogonal relationship. Projection direction can be obtained from the diagonalization transformation of the raw data, therefore the projectieddata will be all located in the axis in the new space.

The value of each new axis can be considered as the energy of the data, then the sum of all the coordinates is all the included energy of the original input data. Align the coordinate values in a descending order to choose the first n values, the sum of which reaches 95% of the total energy and discard the remaining coordinates, therefore, the axes the remained coordinates correspond to are the principal directions in the new space after the projection.

The ratio of the energy can be manually set, which always changes in a range from 85% to 99%. Reflected in the projected data, the larger eigenvalues are retained in the transformed diagonal matrix, the small ones are removed, and the sum of the retained eigenvalues reaches the proportion of the total values. Although the vast majority of

energy reserves, the new space just contains only a very few of axes, which are the corresponding principal components. PCA is significantly based on the principle of energy concentration to achieve the purpose of data reduction through the projection transformation.

The implementation of PCA is often based on the SVD decomposition, the key of which is to obtain the spatial projection transformation matrix. Each row of the matrix above corresponds to the eigenvectors of the matrix after the projection. Set $\mathbf{x}_i$ as the input sample vector, $\boldsymbol{\mu}$ as the mean vector for all the input sample vectors, which is composed of the mean values of all the various components arranged in the order in accordance with the corresponding vector, see equation (2-1).

$$\mu_j = \frac{1}{M}\sum_{i=1}^{M} x_{ij} \qquad (2\text{-}1)$$

Where μ_j and x_{ij} are the j-th element of $\mathbf{x}_i$ and $\boldsymbol{\mu}$ respectively and M is total number of training samples.

Define:

$$\mathbf{X} = [\mathbf{x}_1 - \boldsymbol{\mu}; \mathbf{x}_2 - \boldsymbol{\mu}; \cdots; \mathbf{x}_M - \boldsymbol{\mu}] \qquad (2\text{-}2)$$

The semicolon indicates the separation between the row vectors, which is the row vector $\mathbf{x}_i - \boldsymbol{\mu}$ of the matrix $\mathbf{X}$, respectively.

The covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{X}$ is:

$$\boldsymbol{\Sigma} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}} = \frac{1}{M}\mathbf{X}\mathbf{X}^{\mathrm{T}} \in \Re^{M\times M} \qquad (2\text{-}3)$$

Let the eigenvalues of $\boldsymbol{\Sigma}$ be λ_i, then the row vector of the projection transformation matrix is the eigenvector $\mathbf{u}_i$ of $\boldsymbol{\Sigma}$, which is orthonormalized and corresponds to λ_i. $\mathbf{u}_i$ can be solved through matrix $\mathbf{R}$ and its eigenvector $\mathbf{v}_i$.

Define $\mathbf{R}$ as:

$$\mathbf{R} = \frac{1}{M}\mathbf{X}^{\mathrm{T}}\mathbf{X} \in \Re^{M\times M} \qquad (2\text{-}4)$$

$\mathbf{R}$ and $\boldsymbol{\Sigma}$ contain the same eigenvalues of λ_i, and there is a mathematical relationship between $\mathbf{u}_i$ and $\mathbf{v}_i$:

$$\mathbf{u}_i = \mathbf{X}\mathbf{v}_i / \sqrt{\lambda_i} \qquad (2\text{-}5)$$

Therefore, the projection transformation matrix is:

$$\mathbf{T} = [\mathbf{u}_1; \mathbf{u}_2; \cdots; \mathbf{u}_i; \cdots; \mathbf{u}_n] \qquad (2\text{-}6)$$

In PCA-based data classification, the first step is to calculate the projection transformation matrix to project the original input data to the new space, and then to

train the classifier model by using the data after dimensionality reduction. Test data must be dealt with by the same dimension reduction transformation to obtain the same data format as the training samples and the classification model can be finally applied.

PCA algorithm is quick, easy and simple (without high complexity). After the dimension reduction of PCA algorithm, only a very small part of the original input data is preserved, which correspond to a finite number of the largest eigenvalues of the matrix $\mathbf{R}$ in the equation (2-4). Projection transformation just changes the form of data, but not the energy of data. Therefore, the eigenvalues after transformation contain most of the energy in the original image, that is to say, there is no energic loss of information in the original image, which will not influence the classification to a great degree.

The disadvantage of PCA algorithm is its low accuracy. As the complexity of data increases, its accuracy can not be surely guaranteed.

2.2.2 Kernel PCA

Kernel Principal Component Analysis (Kernel PCA, or KPCA) is an effective extention of PCA algorithm [105], which is also widely used in various fields of pattern recognition and classification [106-108]. The main idea of Kernel PCA is to apply PCA directly to the kernel space or feature space. First define the kernel function as:

$$k(\mathbf{x}_i, \mathbf{x}_j) = < \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) > = (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \qquad (2\text{-}7)$$

Φ is to map transformation from the original input data to a higher-dimensional space (ie the kernel space or feature space), $\mathbf{x}_i$ and $\mathbf{x}_j$ are the original input data, $<\cdot>$ denotes the inner product of two vectors. From Equation (2-7), the result of kernel function is a scalar value, which can be measured as the distance between two vectors to some extent. The closer the distance is, the higher the similarity between vectors is. The basic definition of the kernel function can be applied to all kernel function-based methods in the next sections of this paper.

Refered to the SVD method on the implementation of PCA, the input vector becomes to $\Phi(\mathbf{x}_i)$ in high dimensional feature space, and the corresponding covariance matrix is:

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^{M} \Phi(\mathbf{x}_i)[\Phi(\mathbf{x}_i)]^{\mathrm{T}} \qquad (2\text{-}8)$$

The projection transformation matrix of PCA in the feature space is composed of normalized eigenvectors of the matrix $\mathbf{C}$. To solve the eigenvalues and eigenvectors, the equation as follows is needed to be solved:

$$\tilde{\lambda}\tilde{\mathbf{u}} = \mathbf{C}\tilde{\mathbf{u}} \qquad (2\text{-}9)$$

$\tilde{\lambda}$ and $\tilde{\mathbf{u}}$ are the eigenvalues and eigenvectors of the matrix $\mathbf{C}$ respectively. It is more distinguishable to use the superscript.

Since any vector in the linear space can be expressed as a linear combination of other vectors, the eigenvectors $\tilde{\mathbf{u}}$ can be expressed by a linear combination of each $\Phi(\mathbf{x}_i)$. Suppose α_i is the coefficients of the linear combination, then:

$$\tilde{\mathbf{u}} = \sum_{i=1}^{M} \alpha_i \Phi(\mathbf{x}_i) \qquad (2\text{-}10)$$

Pre-multiply $\Phi(\mathbf{x}_k)$ at the both ends of the Equation (2-9), and substitute Equations (2-7), (2-8) and (2-10) into (2-9), then we can get:

$$M\tilde{\lambda}\sum_{i=1}^{M} \alpha_i \cdot k(\mathbf{x}_k, \mathbf{x}_i) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \cdot k(\mathbf{x}_k, \mathbf{x}_i) \cdot k(\mathbf{x}_i, \mathbf{x}_j) \qquad (2\text{-}11)$$

Construct the inner product matrix $\mathbf{K}$ and the coefficient vector $\boldsymbol{\alpha}$. Each element of $\mathbf{K}$ is an inner product of two input vectors, ie, $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = <\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)>$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_M)$, then Equation (2-11) can be rewritten as:

$$M\tilde{\lambda}\mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha} \qquad (2\text{-}12)$$

Pre-multiply $\mathbf{K}^{-1}$ at the both ends of Equation (2-12):

$$M\tilde{\lambda}\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha} \qquad (2\text{-}13)$$

From (2-13), this equation is equal to solve the eigenvalues $M\tilde{\lambda}$ of the matrix $\mathbf{K}$ which corresponds to the eigenvectors $\boldsymbol{\alpha}$. All the eigenvectors corresponding to the non-zero eigenvalues can form the projection transformation matrix after normalization.

In the realization of Kernel PCA algorithm, first calculate the inner product of the various samples according to the given input samples to construct the inner product matrix $\mathbf{K}$; and then calculate its eigenvalues and eigenvectors to obtain the projection transformation matrix; finally, obtain the principal component projection in the feature space for any input test sample. The whole procedure is very similar to PCA.

Kernel PCA algorithm tries to solve the complex classification problem by mapping from the low-dimensional space into high-dimensional feature space. In general, the difficulty of classification will greatly decrease, and data which can not be separated in low-dimensional space can be linearly separable in high-dimensional space. Algorithm introduces an inner product function, which ensure the feature space must be a Hilbert space and the operations in this space can be completely achieved through the inner product instead of calculation and solution of the unknown mapping. The restriction of the algorithm is its need to pre-selected the kernel function and its corresponding parameters. The selection process will also greatly affect the effectiveness of the algorithm.

2.2.3 Kernel Class Separability

Kernel Class Separability (KCS) is the extention of commonly used traditional class separability in pattern recognition and classification into a kernel space or feature

space [109]. Given an input data set, first calculate of its Between-Scatter Matrix $\mathbf{S}_B$ and the Within-Scatter Matrix $\mathbf{S}_W$. The optimization criterion of KCS is generally defined as the ratio of the traces or the determinants of the Between-Scatter Matrix $\mathbf{S}_B$ and the Within-Scatter Matrix $\mathbf{S}_W$ ($|\mathbf{S}_B|/|\mathbf{S}_W|$ or $\mathrm{tr}(\mathbf{S}_B)/\mathrm{tr}(\mathbf{S}_W)$). The higher the ratio is, the greater the distance between categories is. The classification is easier, and the features from the optimization are more important.

Let Φ be the map for transforming the data from the input space to the high dimensional feature space. Similar to Kernel PCA, KCS is also based on the principle that the inseparable data in original input space can surely be linearly separable in a high-dimensional feature space after a certain map projection. It is not necessary to calculate the map and directly introducing the kernel function can transform the KCS criterion into the Hilbert space criterion.

Using the same definition of the kernel function as in Equation (2-7), the trace of the Between-Scatter Matrix can be expressed as:

$$\begin{aligned}\mathrm{tr}(\mathbf{S}_B^{\phi}) &= \mathrm{tr}[\textstyle\sum_{i=1}^{l} n_i(\mathbf{m}_i^{\phi}-\mathbf{m}^{\phi})(\mathbf{m}_i^{\phi}-\mathbf{m}^{\phi})^{\mathrm{T}}] \\ &= \textstyle\sum_{i=1}^{l} n_i(\mathbf{m}_i^{\phi}-\mathbf{m}^{\phi})^{\mathrm{T}}(\mathbf{m}_i^{\phi}-\mathbf{m}^{\phi})\end{aligned} \qquad (2\text{-}14)$$

In the Equation (2-14), the superscript Φ is used to express the variables in the feature space and distinguish with the sample data in original input space. tr denotes the trace of the matrix; $\mathbf{m}_i^{\phi}$ and $\mathbf{m}^{\phi}$ are the mean vectors in the feature space of samples of the $i-th$ class and of all the classes; n_i is the number of total samples in the $i-th$ class; l is the total number of classes.

Expand the Equation (2-14) and define the inner product matrix $\mathbf{K}$, the element of which represents the inner product of two vectors $\Phi(\mathbf{x}_i)$ and $\Phi(\mathbf{x}_j)$. The symbol $\mathbf{K}_{A,B}$ denotes the elements of the matrix $\mathbf{K}$, which can meet the condition that $\mathbf{x}_i \in A$ and $\mathbf{x}_j \in B$, where A and B denote the sample set. Equation (2-14) can be transformed to:

$$\mathrm{tr}(\mathbf{S}_B^{\phi}) = \sum_{i=1}^{l} \frac{\mathrm{Sum}(\mathbf{K}_{D_i,D_i})}{n_i} - \frac{\mathrm{Sum}(\mathbf{K}_{D,D})}{n} \qquad (2\text{-}15)$$

Where, D_i and D denote the dataset of the samples from the $i-th$ class and all the classes, and the relationship between them needs to meet $D = \cup_{i=1}^{l} D_i$; function $\mathrm{Sum}(\cdot)$ is to sum all the elements of the matrix.

Accordingly, the Within-Scatter Matrix can be expressed as:

$$\begin{aligned}\mathrm{tr}(\mathbf{S}_W^{\phi}) &= \mathrm{tr}[\textstyle\sum_{i=1}^{l}\sum_{j=1}^{n_i}(\phi(\mathbf{x}_{ij})-\mathbf{m}^{\phi})(\phi(\mathbf{x}_{ij})-\mathbf{m}^{\phi})^{\mathrm{T}}] \\ &= \textstyle\sum_{i=1}^{l}\sum_{j=1}^{n_i}(\phi(\mathbf{x}_{ij})-\mathbf{m}^{\phi})^{\mathrm{T}}(\phi(\mathbf{x}_{ij})-\mathbf{m}^{\phi})\end{aligned} \qquad (2\text{-}16)$$

where $\mathbf{x}_{ij}$ is the $j-th$ element of the $i-th$ class.

Similarly, the Equation (2-16) can be expressed as:

$$\mathrm{tr}(\mathbf{S}_W^{\phi}) = \mathrm{tr}(\mathbf{K}_{D,D}) - \sum_{i=1}^{l} \frac{\mathrm{Sum}(\mathbf{K}_{D_i,D_i})}{n_i} \qquad (2\text{-}17)$$

Add the Equation （2-15）and（2-17）, the total scatter is:

$$\begin{aligned} \mathrm{tr}(\mathbf{S}_T^{\phi}) &= \mathrm{tr}(\mathbf{S}_B^{\phi}) + \mathrm{tr}(\mathbf{S}_W^{\phi}) \\ &= \mathrm{tr}(\mathbf{K}_{D,D}) - \frac{\mathrm{Sum}(\mathbf{K}_{D,D})}{n} \end{aligned} \qquad (2\text{-}18)$$

Define the feature selection criterion is the ratio of the traces of the Between-Scatter Matrix $\mathbf{S}_B$ and the Within-Scatter Matrix $\mathbf{S}_W$:

$$J^{\phi} = \frac{\mathrm{tr}(\mathbf{S}_B^{\phi})}{\mathrm{tr}(\mathbf{S}_W^{\phi})} \qquad (2\text{-}19)$$

The function $\mathrm{tr}(\mathbf{S}_B^{\phi})/\mathrm{tr}(\mathbf{S}_W^{\phi})$ and $\mathrm{tr}(\mathbf{S}_B^{\phi})/\mathrm{tr}(\mathbf{S}_T^{\phi})$ has the same monotonicity, therefore, although the largest values of the two functions are different, the optimization procedure is exactly the same. To optimize the criterion （2-19）is equal to optimize the criterion （2-20），and finally the feature selector vector can be obtained.

$$J^{\phi} = \frac{tr(\mathbf{S}_B^{\phi})}{tr(\mathbf{S}_T^{\phi})} \qquad (2\text{-}20)$$

Now we will appropriately simplify the feature selection criterion based on some restrictions.

Obviously,

$$\mathrm{Sum}(\mathbf{K}_{D,D}) \geq \mathrm{tr}(\mathbf{K}_{D,D}) \qquad (2\text{-}21)$$

Integrated with Eqution (2-21), and then

$$\mathrm{tr}(\mathbf{S}_T^{\phi}) \leq \mathrm{tr}(\mathbf{K}_{D,D}) - \frac{\mathrm{tr}(\mathbf{K}_{D,D})}{n} \qquad (2\text{-}22)$$

Suppose $\sqrt{k_s(\mathbf{x},\mathbf{x})}$ is the radius of the classification hyperplane in the feature space, then

$$\mathrm{tr}(\mathbf{K}_{D,D}) - \frac{\mathrm{tr}(\mathbf{K}_{D,D})}{n} = (n-1)k_s(\mathbf{x},\mathbf{x}) \qquad (2\text{-}23)$$

The radius of the separating hyperplane is a constant, Equation (2-23) denotes a fixed value. The denominator of the criterion is a constant value which can be removed without affecting the monotonicity and optimization process of the function. After the amplification in Equation (2-22) by inequality, the final feature selection criterion can be approximated to optimize:

$$J^{\phi} = \mathrm{tr}(\mathbf{S}_B^{\phi}) \qquad (2\text{-}24)$$

From the mathematical analysis above, it can be observed that the feature selection criterion can be approximated by its lower bound $\mathrm{tr}(\mathbf{S}_B^{\phi})$. This means that only optimizing the lower bound function of the feature selection criteria can achieve the feature selection performance to greatly reduce the computation. In pattern recognition theory, the greater the scatter is, the farther the distance between different classes is, and the more easily data between different categories can be distinguished, in which the simplification process is the same as the pattern classification theory.

Using KCS criterion to achieve feature selection, the result of the optimization will be a feature selector. The elements of this vector have only two values: 0 and 1, the element 1 corresponds to the features on these positions are preserved, while the element 0 corresponds to the features of these locations are removed. The dimensions and structures of the feature vectores in feature selector and training vectors in matrix are identical, both of which are multiplied by the corresponding position (element to element), to complete the feature component selection. the feature vector after selection and the initial feature vector have the same dimension, but the number of non-zero components greatly reduces, the number of zero components greatly increases, and the sparsity of vectors is significantly enhanced.

Choose the kernel function as the Gaussian Radial Basis Function (RBF):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}) \qquad (2\text{-}25)$$

$\mathbf{x}_i$ and $\mathbf{x}_j$ are the input vectors; $\|\cdot\|$ denotes the norm of the vector (2-norm is commonly used); σ is the only parameter can be tuned in the RBF kernel function, which can influence the performance to a great degeree. It has been proven theoretically that [109], the KCS-based feature selection can be effectively integrated with the SVM classifier design, both of which can share the same kernel functions and their parameters. Feature selection and classification process do not separately conducted, but similar performance can be achieved to the situation of separate parameters selection.

Define $\otimes$ as the operation of multiplication element by element and $\boldsymbol{\alpha}$ as the feature selector. The vector after feature selection can be described as:

$$\mathbf{x}(\boldsymbol{\alpha}) = \mathbf{x} \otimes \boldsymbol{\alpha} \qquad (2\text{-}26)$$

The inner product of vectors after the feature selection can be described as:

$$\begin{aligned} k(\boldsymbol{\alpha} \otimes \mathbf{x}, \boldsymbol{\alpha} \otimes \mathbf{z}) &= \exp[-\sum_{i=1}^{d} \frac{\alpha_i^2 (x_i - z_i)^2}{2\sigma^2}] \\ &= \exp[-\sum_{i=1}^{d} \eta_i (x_i - z_i)^2] \end{aligned} \qquad (2\text{-}27)$$

In Equation (2-27), $\mathbf{x}$ and $\mathbf{z}$ are the input vectors, d is dimension of the feature vector, where α_i, x_i, z_i and η_i are the $i-th$ element of the vectors $\boldsymbol{\alpha}$, $\mathbf{x}$,

$\mathbf{z}$ and $\boldsymbol{\eta}$ respectively, and also $\eta_i = \alpha_i^2 / 2\sigma^2$.

After the feature selection, the inner product calculation of feature vectors is discrete. once σ is determined, α_i and η_i will become proportional. The final feature selector needs to achieve the binaryzation opreation to 0 or 1, so the optimization of $\boldsymbol{\alpha}$ is equivalent to the optimization of $\boldsymbol{\eta}$. The larger the value of η_i is, the more important the $i-th$ feature element is. $\boldsymbol{\eta}$ can be obtained by the traditional gradient descent algorithm, the initial value of which can be set as a random vector with the mean value of 0 and the magnitude varying in a small range at the level of 10^{-4}. The binaryzation operation needs to set the important threshold T. Only feanture elements which are greater than T can be retained and the feature selection can be achieved.

The KCS criterion integrates the feature selection with the design of SVM and both of them share the same kernel functions and parameters, which reduces the influence of the kernel functions and the parameters to the kernel-based methods and the difficulty of system to a certain extent.

2.2.4 Kernel F-Score

F-Score is a basic method and used to measure the discrimination between two sets which are entirely composed of samples of real values [110].

Traditional F-Score needs to be calculated based on samples. Given m training feature vectors $\mathbf{x}_k$, $k = 1, 2, \cdots, m$, if the numbers of samples in the positive and negative categories are n_+ and n_- respectively, the F-Score Corresponding to the $i-th$ element is:

$$F(i) = \frac{(\overline{x}_i^{(+)} - \overline{x}_i)^2 + (\overline{x}_i^{(-)} - \overline{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \overline{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \overline{x}_i^{(-)}} \qquad (2\text{-}28)$$

In the Equation (2-28), $\overline{x}_i$, $\overline{x}_i^{(+)}$ and $\overline{x}_i^{(-)}$ are the mean values of the $i-th$ element in total samples, total positive samples and total negative samples respectively; $x_{k,i}^{(+)}$ is the $i-th$ element in the $k-th$ positive sample and $x_{k,i}^{(-)}$ is the $i-th$ element in the $k-th$ negative sample.

Using F-Score criterion for the feature selection, it is necessary to calculate the F-Score of each corresponding feature element, and then to calculate the mean value of all the F-Scores, finally to compare all the F-score with an important threshold of the mean value one by one to selection the elements. If greater than this threshold, the element is considered as important and retained, otherwise it will be removed directly.

Kernel F-Score-based feature selection (KFFS) is an extension of traditional F-Score algorithm, and it has been successfully used in the field of medical detection, such as for heart disease and gene sequences [111]. The main difference of Kernel F-Score and F-Score algorithms is the introduction of kernel function in KFFS, which transforms the data processing and computing from the original input space to the

feature space.

The kernel functions in kernel F-Score are mainly selected as the linear or Gaussian RBF kernel function. The processing of KFFS contains two steps: first to map the data or features from input space to the feature space through the kernel functions; second, to calculate the orresponding F-Scores of each element in the high-dimensional feature space according to the Equation. Using the mean value of all F-Scores as a threshold, the elements which are greater than the threshold are retained, the others will be removed.

Kernel F-Score feature selection method is always integrated with the SVM classifier to deal with the classification of medical data together [111], but the choice of the threshold will significantly affect the performance of the classifier. The traditional threshold selection method is too simple, which just uses the mean value of all the F-Scores and is not robust. Therefore, it is necessary to develop a more appropriate threshold selection method.

2.2.5 Support Vector Machine Recursive Feature Elimination

Support Vector Machine Recursive Feature Elimination (SVM-RFE) is a feature selection algorithm for backward elimination of elements based on the dichotomous SVM classification [112], which was first used in the cancer classification problem to select a series of relevant features [113].

The criterion of Optimal Brain Damage (OBD) is first proposed in the treatment from a specific medical problem, which uses the difference of the cost function of the classifier before and after the feature selection as the feature selection criterion [114]. The criterion can be approximated by the second order term of the Taylor expansion of the cost function, as shown in Equation (2-29).

$$c_f = \frac{1}{2} \cdot \frac{\partial^2 L}{\partial (\omega^f)^2} (D\omega^f)^2 \qquad (2\text{-}29)$$

Where L is the corresponding cost function of the used classifiers; ω^f refers to the weight of the feature element f; different feature elements correspond to different weights. Expand the given cost function of the classifier and the Taylor expansion of the cost function will change if removing one element of a feature. OBD feature selection criterion uses c_f as the criterion to approximate the difference of the Taylor expansion before and after the removal of the feature element.

The feature selection criterion of SVM-RFE uses the improved OBD criterion and in theory it has proven that the improvement can achieve superior performance to the original OBD criterion [115]. For the binary SVM, the coefficients $\boldsymbol{\omega}$ of its normal vectors of the hyperplane determine the maximal distance between the hyperplane and the Support Vectors, so OBD criterion can be seen as the removal of those elements which have the least influence to the 2-norm of $\boldsymbol{\omega}$.

According to previous analysis, the difference of the Taylor expansion before and

after the removal of the feature element can be approximated by c_f as:

$$\begin{aligned} c_f &= \left| \|\omega\|^2 - \|\omega^{(-f)}\|^2 \right| \\ &= \frac{1}{2} | \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i^* \alpha_j^* y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i^{*(-f)} \alpha_j^{*(-f)} y_i y_j K^{(-f)}(\mathbf{x}_i, \mathbf{x}_j) | \end{aligned} \qquad (2\text{-}30)$$

Where the coefficient α_i^* is obtained from the training of SVM; K is the same kernel function as defined in Equation (2-7); the symbol $-f$ denotes the removal of the feature element f in the process of feature selection. Therefore, $\alpha_i^{*(-f)}$ and $K^{(-f)}(\mathbf{x}_i, \mathbf{x}_j)$ are the coefficients and kernel function of SVM after the removal of the feature element f.

When using the SVM-RFE feature selection criterion, the initial feature subset is the universal set of the feature vector. In each step of feature selection, the new weights $\boldsymbol{\omega}$ are needed to calculate by and comparing the training samples and the existing feature elements. Then remove the feature elements one by one to calculate the criterion function c_f, and remove the element corresponding to a minimum of c_f. The algorithm continues the iterative optimization process until the chosen feature vector meets the to the pre-set requirements or the algorithm reaches the manually-set stop conditions.

From the calculation procedure of SVM-RFE algorithm, its calculation amount is very large. The corresponding weights must be updated every time, and the feature elements are needed to compare one by one. The alrorithm can achieve high accuracy. In the calculation, SVM-RFE uses the optimized coefficients from the training of the kernel functions and the classifier, therefore the algorithm can be effectively integrated with the SVM classifier, which is consistent with the proposed integration method of feature selection and classifier design.

SVM-RFE-based feature selection algorithm is proposed from the binary classification problem of SVM and the commomly used kernel functions are linear or Gaussian RBF kernel function, which are used to accelerate the training process or be fit for the non-linear situation respectively. This algorithm can also be easily extended to the case of multi-class classification problem by the strategy of one-versus-one [116], but the strategy is too simple and not much applicable to the multi-class problem which will lead to a worse classification result. Because there are still not better strategies on the multi-class problems of SVM, this algorithm is also difficult to adapt to various fields of classification.

2.2.6 Correlation based Filter Selection

Filter-based feature selection methods mainly obtain the optimal subset of feature vectors by using the search strategy. The selection process and the classification process are independent and the main consideration in the selection process is the relationship among variables to be predicted and that between the variables and all the classes [117]. The simplest filter-based method is described as follows: align the feature elements according to their importance and the variables ranked fronter in the sequence will be considered as important variables to be retained. The importance of the feature elements need to measure and quantify to the value by a variety of manually set criteria, and the the important threshold and the number of feature elements to be retained in the selection process required also need to pre-fixed.

Correlation based Filter Selection (CFS) is a filter-based method, which can get a group of feature elements which are closely related to the corresponding categories by quantifying and comparing the correlation between the various features and the corresponding categories and that between the various feature elements. Meanwhile, the algorithm can remove the redundancy which has a weak correlation or is irrelevant with the various features. Using the information theory, the procedure to calculate the correlation is as follows:

The correlation between two feature elements x_u and x_i is measured by the uncertainty factor $U(x_u \mid x_i)$, which can be described as the ration of the mutual information $I(x_u; x_i)$ and the entropy $H(x_u)$ [118], that is:

$$U(x_u \mid x_i) = \frac{I(x_u; x_i)}{H(x_u)} = \frac{H(x_u) - H(x_u \mid x_i)}{H(x_u)} \qquad (2\text{-}31)$$

In the Equation (2-31), $H(x_u \mid x_i)$ is the conditional entropy. Suppose each feature element be the discrete random variables, and based on the information theory:

$$H(x_u) = -\sum_{m=1}^{t_u} P(x_{uv_m}) \log P(x_{uv_m}) \qquad (2\text{-}32)$$

$$H(x_u \mid x_i) = -\sum_{n=1}^{t_i} P(x_{ij_n}) \sum_{m=1}^{t_u} P(x_{uv_m} \mid x_{ij_n}) \log P(x_{uv_m} \mid x_{ij_n}) \qquad (2\text{-}33)$$

Where t_u and t_i are the number of values of feature elements x_u and x_i, that is to say, x_u can get t_u values from v_1 to v_{t_u}, and x_i can get t_i values from j_1 to j_{t_i}. The symbols $P(x_{uv_m})$ and $P(x_{uv_m} \mid x_{ij_n})$ are expressed as follows in Equations (2-34) and (2-35):

$$P(x_{uv_m}) \triangleq P(x_u = v_m) \qquad (2\text{-}34)$$

$$P(x_{uv_m} \mid x_{ij_n}) \triangleq P(x_u = v_m \mid x_i = j_n) \qquad (2\text{-}35)$$

Suppose the given sample set have the same dimensions, and the criterion (or the search strategy) of CFS for searching the optimal subset can be expressed as:

$$G_S = \frac{k\overline{r}_{ci}}{\sqrt{k + k(k-1)\overline{r}_{ii'}}} \qquad (2\text{-}36)$$

Where k is the number of feaure elements of the selected subset in this search process, $\overline{r}_{ci}$ is the mean value of the correlation between each feature element in the subset and the classes, $\overline{r}_{ii'}$ is the mean value of the correlation between the feature elements in the feature subset.

Uncorrelated or weakly correlated feature elements will lead to a small value of G_S, therefore, the search will follow the direction of a larger G_S value. The advantages of CFS-based algorithm are that the feature selection process is completely independent with the classification and that it can conveniently entended to the situation of only one class of samples (positive samples) with a strong applicability. But the limitation of CFS is that the process of feature selection depends on the choice of the search strategy. If the strategy is not selected properly, the search time may be very long but it can not get the optimal solution.

2.3 Summary

This chapter introduces the general concept, significance and process of feature selection. For the proposed SVM-based tumor detection system, some feature selection algorithms related to the kernel functions are also introduced in detail to verify proposed integration of feature selection and classifier design. The next chapter will introduce the proposed brain tumor detection framework.

CHAPTER 3 TUMOR DETECTION IN MRI IMAGES

3.1 The Framework of the Proposed System

As to the brain tumor detection problem in MRI images of multiple sequence, this paper proposes a two-step framework, from the total tumor detection to the tissue classification, that is, first to achieve the total detection of brain tumor region, and then to extract the abnormal tissues' details in the obtained tumor region. The whole tumor region in the initial detection is the basis for the classification of tissues and its accuracy will greatly affect the accuracy of tumor classification. If the accuracy of the tumor detection is high, the tissue classification can also achieve high accuracy in general as long as the classification algorithm is mature, and vice versa, a bad algorithm will limit the performance of the system.

A SVM-based semi-automatic framework on brain tumor detection is proposed in this paper, as shown in Figure 3.1.

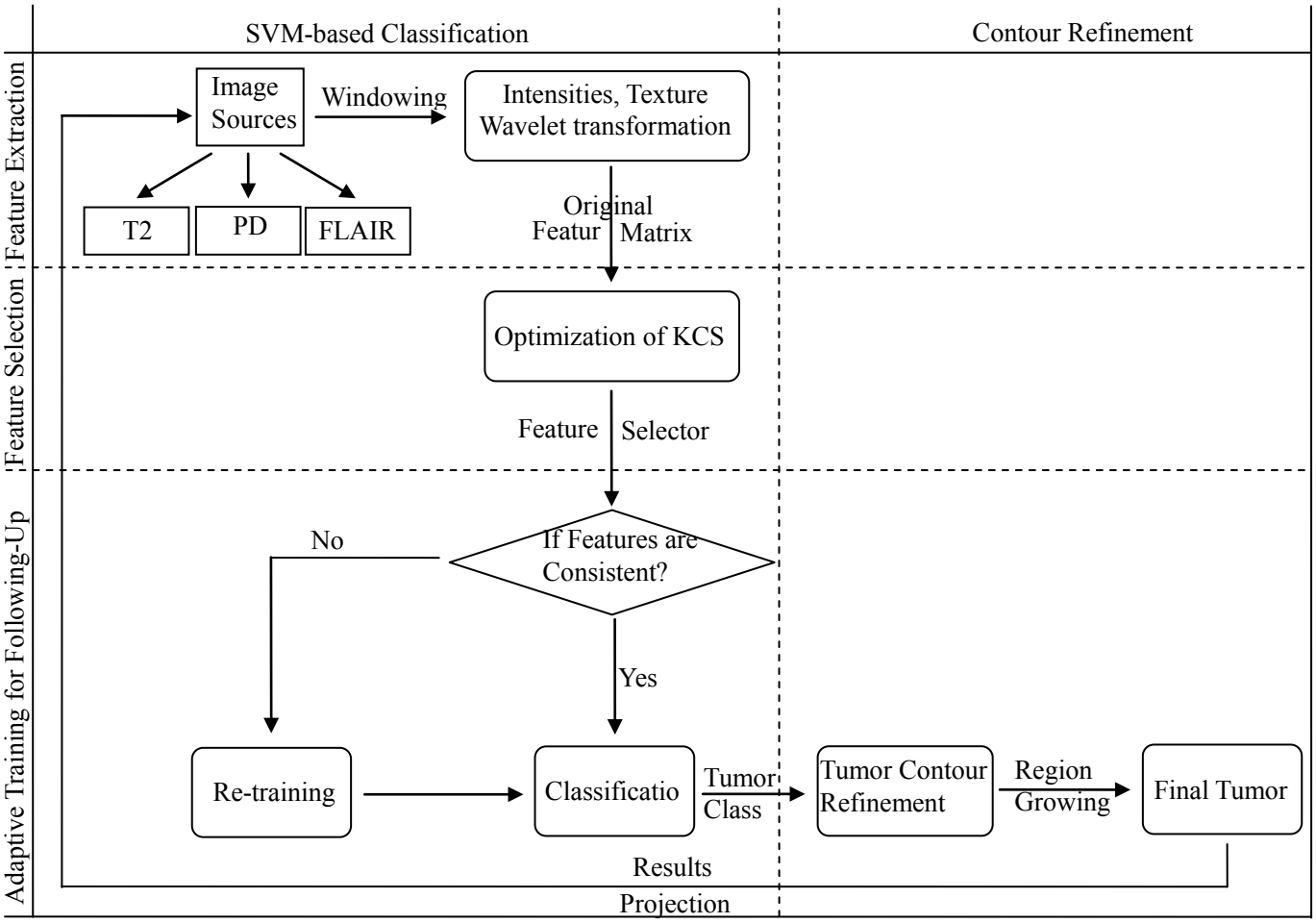


Figure 3.1 the framework of the proposed system.

According to Figure 3.1, the proposed framework for the detection of brain tumors is mainly composed of three parts, namely, the SVM-based tumor classification subsystem, the region-growing-based contour refinement subsystem and the adaptive training-based following-up subsystem. The classification subsystem is to complete the initial detection of the whole tumor based on SVM; the contour refinement subsystem is to improve the quality of the initial contour of the detected tumor region and the classification accuracy; the tracking subsystem is used for the following-up and automatic analysis of the patient's data in the following MRI examinations and the evaluation to the treatment effect in last examination based on adaptive training.

The processing flow of the system is as follows:

After obtaining the MRI examination data of patients, it is necessary to register them in order to ensure that the pixels at the same positions (corresponding to the same coordinates in the MRI slice) in all the different MRI sequences corresponds to the same anatomical brain tissue. Multiple sequences of MRI images can be meaningful and the data fusion makes sense only after the registration to maintain the exact correspondence of data.

The algorithm on classification used in this proposed system is a supervised learning classification method and the classifier used in the system is SVM. The design of the classifier needs a training the learning process, therefore the next task after registration is to construct the training matrix. This process is divided into two steps, the selection of sample points and the generation of feature vectors.

The selection of sample points is randomly carried out on a random FLAIR slice by a team of doctors with rich experience and medical knowledge. The samples contain two classes inside and outside the tumor region respectively. This operation can not only ensure the correctness and reliability of the random sample points, but also reduce the subjectivity in the selection process. The selected results can be accepted by other doctors to make the following processes more persuasive .

After the selection of the sample points, it is necessary to construct the training matrix of feature vectors for the classifier. The feature vectors are extracted in a square window W with the sample point as its center. Suppose the length of a window's side is a, thus, the window W contains a total of a^2 pixels. All the gray values, the texture features of the mean value and standard deviation of the sampling window and derivative features based on a variety of transformation (such as low-frequency coefficients of two-dimensional wavelet transformation) are extracted to form a feature vector, the dimension of which is the sum of dimensions of all the features above. Set m is the number of texture features, and b is the number of derivative features, then the total dimension of a feature vectors is: a^2+m+b. If the derivative feature is selected as the low frequency coefficients of the two-dimensional wavelet transformation, then $b=(a+1)^2$. Align all the extracted features above in the order of gray levels, texture features and derivative features, and we can get a feature subvector corresponding to some sample point in a weighted MRI image sequence.

Set x_i as the gray value of pixel in image, a^2 as the total number of pixels in the window W for feature extraction and $\sum_W$ as the sum of a certain feature of all the points in the window. The commonly used texture features can be obtained according to the equations as follows:

(1) Mean Value:

$$\overline{m}_w = \sum_W x_i \Big/ a^2 \qquad (3\text{-}1)$$

(2) Standard Deviation:

$$SD_W = \sum_W (x_i - \overline{m}_w)^2 \Big/ (a^2 - 1) \qquad (3\text{-}2)$$

(3) Geometric Mean:

$$G_W = (\Pi_W x_i)^{1/a^2} \qquad (3\text{-}3)$$

(4) Harmonic Mean:

$$H_W = a^2 \Big/ \sum_W (1/x_i) \qquad (3\text{-}4)$$

(5) Skewness:

$$S_W = \sum_W (x_i - \overline{m}_w)^3 \Big/ (a^2 - 1) SD_W^3 \qquad (3\text{-}5)$$

(6) Kurtosis:

$$K_W = \sum_W (x_i - \overline{m}_w)^4 \Big/ (a^2 - 1) SD_W^4 \qquad (3\text{-}6)$$

The feature vector at some sample point is a combination of three feaure subvectors which corresponds to T2-weighted images, PD-weighted images and FLAIR weighted images respectively. The three feature subvectors are aligned in a fixed order. Each feature vector occupies a single row in the initial training matrix prepared for the classifier, respectively. The total number of sample points selected in the first MRI examination is $2N$ (a half in tumor region and and a half outside the tumor region), and then the total dimension of the training matrix is $2N \times [3 \times (a^2 + m + b)]$.

If the value a is larger, the dimensions of the feature vector and the training matrix are both very high, and the extracted features in the data are inevitably mixed with a lot of redundancy and interference to affect not only the accuracy, but also the efficiency of classification. Therefore, it is necessary to reduce the dimension of the training matrix before the classification, to select some of the most important features, to decrease the influence to the classification and to accelerate the classification process by eliminating the redundancy and noise in input data.

In order to be applied to the SVM classifier and validate the proposed method on the integration of feature selection and SVM classifier design, this paper attempts to utilize a variety of feature selection algorithms, including the classic PCA, kernel

function-based methods, Information theory-based methods to verify the feasibility of the proposed system.

The tidy matrix after the feature selection is the real training matrix, which is used to learn and obtain the optimal parameters of the relevant classifier, to construct the classification model and to be applied to all the images. Then we can detect the initial tumor region. All the description above is the main work of the SVM-based classification subsystem.

The region-growing-based contour refinement subsystem is used to improve the contour of the initial tumor region, to reduce the error determination and missing points near the border, to decrease the effect on tumor detection from invasive glioma and finally to obtain the final results of tumor detection.

The adaptive training-based following-up subsystem is mainly used to automatically analyze and evaluate the patient's condition. The classifier can achieve the characteristics of the patients with cancer information through the analysis on the first examination data under the human intervention. The system can segment the tumor region in the data of next MRI examinations automatically by the classifier of the follow-up system, which can reflect the effectiveness of the medical treatments on one hand and evaluate to improve the treatment of the next examinations on the other hand by comparing the size changes of tumor volume between two adjacent examination periods.

The entire treatment procedure of the systems is expressed as mentioned above. In the next section, the three subsystems will be described in detail one by one.

3.2 SVM-based Classification Subsystem

3.2.1 SVM theory

SVM is a classification algorithm for high-dimensional data analysis which is proposed by Vapnik to solve the classification problems of two issues [88]. SVM has been widely used in the fields of medical image processing, text analysis, image retrieval and so on.

SVM is based on the principle that the data in the original input space can be linear separable in a higher-dimensional feature space after a certain mapping. The feature space after mapping is a complete Hilbert space, in which the inner product of data can be calculated as the equivalent funcation value of the unput vectors by introducing the corresponding kernel function. The inner product is the measurement of the distance, which can be expressed as the degree of similarity between two vectors in a certain extent. In general, the closer the distance between two vectors is, the higher the similarity of them is. Due to the kernel function, the computation of distance in the feature space is transformed to the input space without the need to solve the mapping of spatial transformation, therefore, it reduces the computational difficulty to a certain

extent.

Based on the Statistical Learning Theory, SVM can obtain the optimal performance by constructing the optimal classification surface with the largest classification margin.. The schematic diagram of SVM is shown in Figure 3.2 [29, 119].

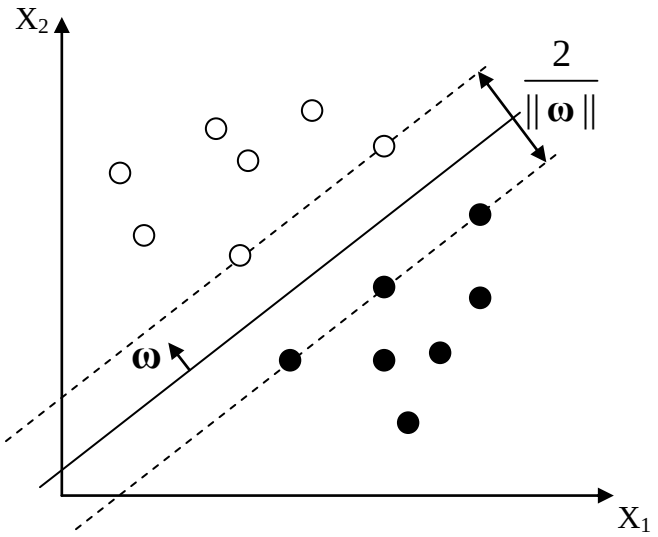


Figure 3.2 The schematic diagram of SVM.

In Figure 3.2, the black and white circles represent the samples of the two linearly separable classes. The symbol $(\mathbf{x}_i, y_i)$ $(i = 1, 2, \cdots, n)$ denotes the sample set consisting of n samples. $\mathbf{x}_i$ is the sample vector and y_i denotes the corresponding class with its values of just ± 1 which corresponds to the positive and negative categories respectively. According to the theory of pattern classification, in high dimensional space the separating hyperplane can be expressed by a linear equation as $\boldsymbol{\omega} \cdot \mathbf{x} + b = 0$, where $\boldsymbol{\omega}$ is the normal vector of the hyperplane, $\mathbf{x}$, generally corresponding to a given sample vector, is a vector as a independent variable of the linear separating hyperplane, and b is the intercept.

Normalize the distance between the samples and the separating hyperplane to make the nearest distance in both classes be 1, then the classification margin of the two classes is $2/\|\boldsymbol{\omega}\|$. To maximize this classification margin is to minimize $\|\boldsymbol{\omega}\|$ or $\|\boldsymbol{\omega}\|^2$.

Then,

$$\phi(\boldsymbol{\omega}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 = \frac{1}{2}(\boldsymbol{\omega} \cdot \boldsymbol{\omega}) \qquad (3\text{-}7)$$

Construct the Lagrange function as:

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}(\boldsymbol{\omega} \cdot \boldsymbol{\omega}) - \sum_{i=1}^{n} \alpha_i \{y_i[(\boldsymbol{\omega} \cdot \mathbf{x}_i) + b] - 1\} \qquad (3\text{-}8)$$

Where α_i is the Lagrange coefficient and $\alpha_i > 0$.

According to the stationary points' method in the function extremum solution, in Equation (3-8) the partial differential about $\boldsymbol{\omega}$ and b are respectively computed to make the corresponding partial differential equations equal to 0, the original problem can be transformed to that of dual problem under the constraints

$$\sum_{i=1}^{n} \alpha_i y_y = 0 \qquad (3\text{-}9)$$

$$\alpha_i \geq 0, \quad i = 1, 2, \cdots, n \qquad (3\text{-}10)$$

To maximize the following function:

$$Q(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \qquad (3\text{-}11)$$

Based on the Kuhn-Tucker, the extreme value problem of the quadratic function under the inequality constraints exists the unique solution similar to the Equation (3-12):

$$\alpha_i \{ y_i [(\boldsymbol{\omega} \cdot \mathbf{x}_i) + b] - 1 \} = 0, \quad i = 1, 2, \cdots, n \qquad (3\text{-}12)$$

According to the Equation (3-12), the corresponding coefficients α_i for the solution of most samples are 0, and those nonzero coefficients is relevant to the Suppoet Vectors. Finally, the optimal classification function is:

$$f(x) = \operatorname{sgn}\{(\boldsymbol{\omega}^* \cdot \mathbf{x}) + b^*\} = \operatorname{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^*\} \qquad (3\text{-}13)$$

Where $\boldsymbol{\omega}^*$, b^* and α_i^* are the coefficients from the optimization, and $\mathbf{x}_i$, y_i are the input vector and its label. The symbol $\operatorname{sgn}(\cdot)$ is the sigh function.

When the input samples are not linearly separable, it is necessary to introduce a relaxation term of $\xi_i \geq 0$ to reduce the requirements of the constraints, which can satisfy most situations:

$$y_i [(\boldsymbol{\omega} \cdot \mathbf{x}_i) + b] - 1 + \xi_i \geq 0, \quad i = 1, 2, \cdots, n \qquad (3\text{-}14)$$

The problem need to be optimized is transformed to the following equation after the introduction of relaxation term:

$$\phi(\boldsymbol{\omega}, \xi) = \frac{1}{2} (\boldsymbol{\omega} \cdot \boldsymbol{\omega}) + C(\sum_{i=1}^{n} \xi_i) \qquad (3\text{-}15)$$

Where C is a specified cost constant to control the extent of the punishment on the wrong classification samples and to achienve the compromise between the ratio of the wrong classification samples and the complexity of the algorithm [29, 88].

The optimization in this case can be understood as: after defining a mapping Φ for the tramsformation of the input samples into the linearly separable feature space, the

linearly non-separable data will change into a linearly separable sample set. Comparing with the analysis in the linearly separable situation, the input samples here are required to replace as $\Phi(\mathbf{x}_i)$. Introducing the kernel function to calculate the inner product to simplify the Equations (3-11) and (3-13), the equation of the SVM classification hyperplane can be finally obtained as:

$$f(x) = \mathrm{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^*\} \quad (3\text{-}16)$$

The Commonly used kernel functions in SVM are:

(1) the linear kernel function:

$$K(\mathbf{x}, \mathbf{x}_i) = g(\mathbf{x} \cdot \mathbf{x}_i) + h \quad (3\text{-}17)$$

(2) the polynomial kernel function:

$$K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} \cdot \mathbf{x}_i) + 1]^q \quad (3\text{-}18)$$

(3) the Sigmond kernel function

$$K(\mathbf{x}, \mathbf{x}_i) = \tanh[\upsilon(\mathbf{x} \cdot \mathbf{x}_i) + c] \quad (3\text{-}19)$$

When using the Sigmond kernel function, SVM is equal to a two-layer neural network, but the weights of neural networks and the number of nodes in the hidden layer need to be calculated and determined by the algorithm itself.

(4) Gaussion Radial Basis Function (RBF):

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\{-\frac{\| \mathbf{x} - \mathbf{x}_i \|^2}{\sigma^2}\} \quad (3\text{-}20)$$

The advantages of SVM is fully theoretical. Despite the theoretical calculation and optimization of SVM is related to a high-dimensional mapping, but in the actual calculations it is not necessary to solve the mapping. Optimization can be achieved just based on the transformation of the inner product function and the calculation in the high-dimensional space by the kernel function, which greatly simplies the complex solution process.

It can be observed from the principle of SVM that, SVM is especially suitable for the high-dimensional input data, and not sensitive to the source of the data, which proves that it isconvenient to easily integrate the analysis of the data with SVM to improve the total classification accuracy.

3.2.2 Multi-kernel SVM

SVM has many extensions, such as one-classification SVM, in which all the interested data will be considered as one class and all the other data as the other class. In the classification, the training is only fulfilled on the data of the first class. This is a simple application of two-class SVM classification [89].

Another important application of SVM is the multi-kernel SVM [90-91]. It is especially suitable for dealing with high-dimensional data of different sources and different types. The classification algorithm does not use a uniform kernel function in order to be able to dig the implicit information in the data and to take full use of the characteristics in the data.

In general, the multi-kernel SVM has two typical forms :

1. for different types of input data from different sources, use different kernel functions or different parameters with a same kernel function to design different classifiers for each input respectively, select the optimal parameters to classify each input one by one to obtain the optimal results. The final result is the integration of all the optimal results of the input data;
2. First, some strategy is adopted to fuse all the input data, and then use different kernel functions or different parameters of a same kernel function to carry out the classification; finally using the selected fusion strategy to obain the integrated results.

The two forms of multi-kernel SVM have very different characteristics. The former takes full use of the properties of their own in different types of data, which are independent. The use of different kernel functions can mine a variety of important and implied characteristics of data and the ultimate classification result is the integration of all the optimal results. The latter fuses the data first to use the implicit information, characteristics or relationship in various data, although different kernel functions are applied to all the parts of the data. The more important point in this processing mode is to use the correlation and mutual influence of data to achieve the classification.

In addition to adapt to multiple types of input data, multi-kernel SVM can also extended to the multi-class classification problem.

At present, the multi-class classification problems on brain tissues' detection mainly depends on the fuzzy theory or the probability models, both of which are not very accurate in general. Some algorithms can just segment edema and normal tissues separately, which indicates the lack of ability to deal with brain tumors.

There is no specific multi-class classification algorithm in traditional SVM. The existing multi-class classification strategies are the simple expansions of two-class classification algorithms, such as a commonly used one versus all strategy, in which it is necessary to construct a two-class classifier for each classification of different data. The interested data will be considered as one class and all of the other data as the other class. Each class has its special classifier to make the margin between this class of samples and the other samples the largest. Then the multi-class classfication can be achieved by determine each category one by one.

The results of each two-class classification can be removed from the data set for the next classification. It is also acceptable not to remove these data. The removal of data from last classification may lead to the previous classification results can not be

amended. If the previous classification is not accurate, the error will be gradually accumulated in the following process, which will lead to worse and more inaccurate results in future.

However, if removing the data which has been previously classified, the same sample maybe classified into different categories, especially for the samples at the edge of the tissues, of which the characteristics are between the various tissues, the classification will be difficult. This kind of repeated classification will greatly influence the performance of the classifiers, since each sample needs to be compared with all other samples at each classification, which leads to a very heavy computational burden.

It can be observed that directly apply SVM to multi-class classification problems will lead to poor results, no matter what situation above appears. That is to say, SVM can not solve the classification problems in the complex situation of multi-sequence MRI images. Some possible solutions are extending the existing classification algorithms to the kernel space or feature space by introducing the kernel functions in other classifiers or introducing other classification algorithms into SVM to improve the accuracy and effect of SVM on multi-class classification.

As to the limitations of existing algorithms, multi-kernel SVM can be well applied to multi-class classification problem, whether the fusion is at the data layer or the decision-making layer. We can use different kernel functions or different parameters in a same kernel function to correspond to the characteristics of different tissues, different data, different treatments. Using the characteristics of each input samples of their own or the correlation between samples can achieve the desired classification.

The training process of multi-kernel SVM is usually to train the parametric model with a single kernel function, and then the parametric model with multiple kernel function. the relevant parameters are continuously optimized in the SVM training process to apply to different input.

3.2.3 Fusion Strategy

Two types of multi-kernel SVM both require the use of some certain fusion strategies in order to take full advantages of multi-sequence MRI data. Corresponding to multi-kernel SVM, the fusion strategy is also applied in two levels, namely, data level fusion and decision-making level fusion, respectively. The former means the fusion of the input data or the extracted feature, while the latter refers to the fusion between different classification results.

Dempter-shafer theory [120] can solve the fusion problems. The theory is the extension of Bayesian probability model, which can measure the accuracy of the uncertainty of the events by defining the setting the trust function and the likelihood function, besides the uncertainty description by the probability. The trust function is the lower bound of the uncertainty, while the likelihood function is the upper bound. They measure the minimal and maximal value of the uncertainty respectively.

The implemention of the Dempter-shafer theory algorithm can be divided into two

steps, according to the general rules of human decision-making process [121]:

1. First consider the available evidences and its impact to decisions at the trust level of trust, to obtain the initial judgement by the objective conditions;
2. Second at the decision-making level, integrated with the existing evidences, add the subjective judgments and synthesize the two factors above to make a decision.

Simple fusion strategy can be the direct use of logical functions, as long as different input data or the results of the different categories have different weights, the results can be fused under various conditions.

Another expanded direction is the a fusion of multi-modality data, that is, the fusion of different types of medical images, such as fusion of MRI images and CT/PET images. This fusion strategy is theoretically feasible, but necessarily in practice may not easily be achieved. In general, patients will not repeat an MRI examination after the CT examination and PET examination is very expensive for the patients, so there are still certain difficulties to achieve in reality.

A more feasible idea is to fuse MRI images and Magnetic Resonance Spectrum (MRS). Difference from MRI images, MRS is functional analysis of the spectrum, which can image “in advance”. In real medical diagnosis, the disease can be shown on the images when it develops to a certain stage or deteriorates to a certain extent. In fact, before the lesions can be shown on the images or in the early period of the disease, the content of some disease-related compounds in human body has changed obviously because of the metabolic abnormalities of tissues (normal compounds are inhibited and their contents reduce while the contents of pathogenic compounds significantly increase). MRS can predict the incidence of lesions in advance by examining the abnormal changes of the elements or compounds. In other words, if the contents of some key compounds corresponding to the cancer in a brain region significantly increase, then we can predict in advance this region is likely to become cancer incidence areas or lesions in the future.

On the other hand, MRI images and MRS spectrum are homologous data and they correspond to each other. If the MRI images have clearly reflected the tumor area, which can be used as the prior knowledge of the tumor in the test, the contents’changes of key compounds are tested in the region by determining the approximate range of the tumor region. Then using the pathological analysis will help to determine the current state and future development trend of tumor disease and to clearly guide the evaluation and determination of the medical treatment; oppositely, the changes of compounds in the brain tumors detected by MRS spectra indicate the incidence area, which is the focus area and need to be imaging significantly for MRI images.

3.2.4 Feature Selection

The system uses the three sequences of T2-weighted images, PD-weighted images

and FLAIR-weighted images from the MRI examinations. We use feature selection method to choose the most important features from the three selected sequences and fuse them effectively (feature selection-based data fusion). Feature selection is an important part of SVM-based classification subsystem, which can not only extract critical information after data fusion, but also be integrated with the SVM classifier to reduce the difficulty of classifier design. The specific methods and their meaning have been discussed in the second chapter detailedly.

3.3 Region Growing-based Contour Refinement Subsystem

The training data for the classifier are selected by the feature selection. Not only the redundancy, interference and noise in the data, but also even some relatively minor information have been removed. The study in this paper is mainly on the gliomas in the brain, of which the properties are the invasion of the boundary tissues (such as cystic degeneration, edema) into the surrounding normal tissues to varying degrees and mixed with the normal tissues together, therefore, the contour of the tumor can not be clearly distinguished. Reflected in the MRI images, it can be observed that the data characteristics of a variety of tissues near the border are similar. With the limitation of the performance of classifiers, all the factors above determine that the initial results obtained from the tumor detection will inevitably include errors, especially in the border of the initial classification, where there are always the situations of errors or missings. Thus, it is necessary to improve the initial tumor contour by some contour refinement algorithms.

In this paper, the region growing-based contour refinement algorithm is proposed. The algorithm uses the criterion of the the maximum likelihood combined with the distance between current boundary points and the initial tumor contour. The algorithm is simple and fast, which can well solve the problem of contour refinement in the tumor boundary.

As to the initial results of tumor detection for a certain MRI sequence, first fit the total tumor region to satisfy the normal distribution. Since the tumor region is already known, the gray value of each pixel in the region is known. Calculate the mean value and standard deviation of the region and then the probability density function can be fit as the Equation (3-21) as follows:

$$l_{\gamma}^{p}(x)=\frac{1}{\sqrt{2\pi\sigma_{\gamma}}}\cdot\exp[-\frac{(x_{\gamma}^{p}-\mu_{\gamma})^{2}}{2\sigma_{\gamma}^{2}}] \qquad (3\text{-}21)$$

Where the parameter γ denotes different MRI sequences, and $\gamma=\mathrm{T2,PD,FLAIR}$; μ_{γ} and σ_{γ} correspond to the mean value and standard deviation of the initial tumor region in the MRI sequence γ; x_{γ}^{p} is the gray value of pixel (or border point) p in the MRI sequence γ; $l_{\gamma}^{p}(x)$ is the density probability of pixel p.

Dilate the initial tumor region outside for one pixel, and calculate the sum of the

the density probabilities of each sequence fot a new pixel p on the boundary according to the Equation (3-21), as shown in Equation (3-22):

$$l_{\mathrm{T2}}^{p}(x)+l_{\mathrm{PD}}^{p}(x)+l_{\mathrm{FLAIR}}^{p}(x)>th \Rightarrow p \in \text{tumor} \qquad (3\text{-}22)$$

The obtained total density probability is compared with the pre-set threshold. If superior to the threshold, then determine the corresponding pixel belongs to the tumor tissues and is retained, and vice versa, to determine that the point does not belong to the tumor area, and it will be removed. Determine the boundary points one by one according to the criterion above, and then we can obtain the new bourdary after the region growing.

Repeat the above process and dilate outside the tumor region one pixel by one pixel. Each dilation is based on the latest contour. Determine the categories of the boundary point by point to update the contour. The whole process automatically and continuously runs, until the border no longer changes.

The threshold th need to be set manually. The initial threshold is selected as the average density probabilities of pixels in the initial tumor area of the entire sequences (each sequence separately). Considering the fact that the points which are farther from the initial tumor contour are less likely to be long to tumor regions, the threshold is set to increase gradually with the distance, so that the decisions will be more and more difficult to ensure that the algorithm can automatically stop and results are accurate and reasonable.

$$th=[1+0.1\times(d-1)]\times t \qquad (3\text{-}23)$$

where t is the original threshold, and d is the distance between the latest contour and the initial border of tumor detection.

After the contour improvement, the final tumor region can be obtained.

3.4 Adaptive Training-based Tracking Subsystem

The proposed brain tumor detection framework is a semi-automatic system, but there is only one human interaction in the system in the first treatment on the data analysis of the patient (random selection of the training sample points by a medical team). The following data analysis in the next MRI examinations will be automatically carried out by the system, and the tracking on the disease of the patients mainly relies on the adaptive training.

The system obtains the characteristics of the patient's tumor and the corresponding data information after the first analysis and processing of the patient's data. As to the data in the next examinations, the system can not only use the existing prior knowledge to automatically detect the tumor, but also to evaluate the changes of tumor size, which can be used to assess the medical treatment to a certain extent and to provide reasonable reference for the treatment of the next period.

The process of the adaptive training-based following-up is as follows:

1. Data Projection: after obtaining the already-registered MRI data of the latest inspection, it is necessary to project the final tumor detection results after contour improvement to the new data as a prior knowledge in this data analysis. However, there is no prior knowledge on the changes in tumor size between several examinations. In general the period for the patient's examination is about 4 months, and the tumor does not change a lot during this period because of the radiation therapy or drug control from the doctor. In order to ensure a high accuracy of the projection (the deviation of tumor regions in two examinations will not be large), it is necessary to make a morphological erosion operation before the projection.

2. Automatic selection of sample points: the sample points for adaptive training are automatically selected by the system. In general, the more the number of sample points is, the more classified information it can provide and the better the automatic classification is. However, too many sample points can also introduce much more redundancy and noise, while greatly increasing the computational burden. Compromise the two factors of the accuracy and efficiency, the system automatically selects twice the number of new sample points than that of the initial sample points for training, that is, to randomly select $2N$ points both inside and outside the tumor. In order to ensure that the new sample points are absolutely correct in the case without human interaction, a series strict constraints are introduced, such as the spatial location restriction (based on the last tumor detection area as a reference), gray level restriction (to meet the consistence of the characteristics of tumors reflected the same in the MRI images) and texture restriction (to meet a uniform texture characteristics in the feature extraction window).

3. Feature Extraction and Selection: repeat the analysis process on patient data in the first examination, extract the corresponding feature vectors and training matrix in a square window with the sample points as the center, and use the preset feature selection method for data dimensionality reduction and screening.

4. Comparison of the feature's consistency: although the two adjacent MRI examinations is only apart about 4 months, in general the tumor will not significant change, but the changes of the objective conditions, such as the imaging equipment models, different imaging parameters, the different examination environment, the inevitable changes of the patient's posture in the examination, and so on, will lead to the differences in characteristics of tissues in the MRI images. Reflected at the data level, the retained important features may be different after the

feature selection under the same criteria. The feature set will directly affect the training of the classifier, which leads to that the classifier used in last tumor detection is not necessarily appropriate for the data in this examination. Therefore, in this paper, a t-test is used to compare the consistency of feature subsets in the training matrix after the two feature selection. If the differences of the feature subset is below 5%, it can be considered that data between the two examinations have the same characteristics and the tumor characteristics do not change obviously, which will not significantly affect the training of the classification model. Thus it is not necessary to re-train the new data and the system can directly use classification model in last period; otherwise we need to train a new model from use the new data to make the classification model be applicable to the new data, and then achieve the tumor detection.

5. Evaluation of tumor volume: it is required to refine the tumor detection contour of the new tumor region to improve the classification accuracy. Define the tumor volume as the total number of pixels in the detected regions of all the slices. The development trend and speed can be obtained by comparing the tumor volume between two adjacent MRI examination periods, which can provide a supplementary reference for the doctor's diagnosis and treatment.

3.5 Summary

This chapter proposes a SVM-based, semi-automatic tumor detection system. The classification framework is composed of the SVM-based classification subsystem, the region growing-based contour refinement subsystem and adaptive training-based follow-up subsystems. The system requires one manual interaction only in the first data analysis, and the other operations can be dealt with automatically.

In addition to the detection of the entire tumor regions, the system can also track the patient's condition and obtain quantitative information on changes in tumor to provide information for the diagnosis and treatment to the doctors.

CHAPTER 4 EXPERIMENTATION AND DISCUSSION

4.1 Experimental Data

All the experimental data of patients used are provided by the Caen Medical Center (France) and Beijing Tiantan Hospital, with a total of 13 patients, 23 examination periods, in which Patient 1 to Patient 3 has 5, 4 and 5 examinations respectively. The intervals between adjacent examinations are about 4 months or so, so these three patients have been supervised for more than one year in order to verify the automatic tracking algorithm proposed in this paper. Patient 4 to Patient 13 contains a period of examination data.

Examination data of each patient contain four MRI sequences respectively: T1-weighted images, T2-weighted images, PD-weighted images and FLAIR images. All the image data are derived from 1.5T magnetic resonance imaging equipment of the General Electric (GE) Co. Ltd., with an axial FSE (fast spin echo) imaging sequence to obtain the MRI sequences, to reduce the sensitivity of the imaging system to the non-uniformity of the magnetic field and to increase the energy deposition of the echo sequence [11]. The obtained T1-weighted images contain 124 slices, while T2-weighted images, PD-weighted images and FLAIR images contain 24 slices, respectively. The image resolution in all slices is 512 × 512, however, the including number of slices in T1-weighted image sequence is quite different from that of T2-weighted images, PD-weighted images and FLAIR images, which leads to the imaging thickness of a slice, the tissue location in the brain and the contained information are not completely corresponding. Therefore, in the experiment, this paper uses the other three MRI sequences except the T1-weighted image sequence as the input data in the experimental test, and the slice thickness is 5.5 mm, the size of the corresponding voxel is 0.47 × 0.47 × 5.5mm^3.

There are multiple examinations in the diagnosis and treatment of a patient, and the examples of the tracking data at the same brain position from different periods are shown in Figure 4.1. Compared with the other two MRI image sequences (T2-weighted images and PD-weighted images), FLAIR images have better visual effect, so Figure 4.1 just indicates the FLAIR image, so does the rest of this chapter. If there are no special instructions, the figures just list the FLAIR-weighted images.

The examples of the original input image sequences of patients are shown in Figure 4.2. In Figure 4.2, the listed three images are all from Patient 1, of which each row corresponds to the same position in the brain, and from left to right the images are the PD-weighted images, T2-weighted images and FLAIR images.

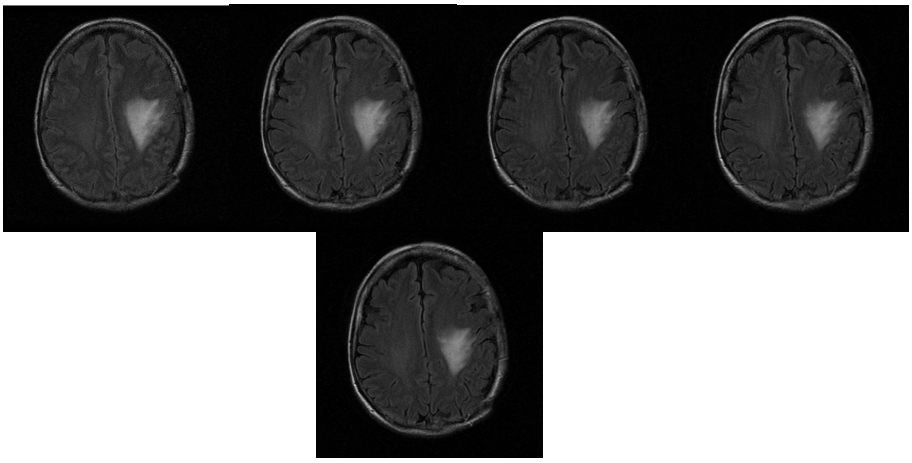


Figure 4.1 The examples of five MRI examinations of Patient 1 (FLAIR images)
(from top to bottom, from left to right corresponding to the first to the fifth period).

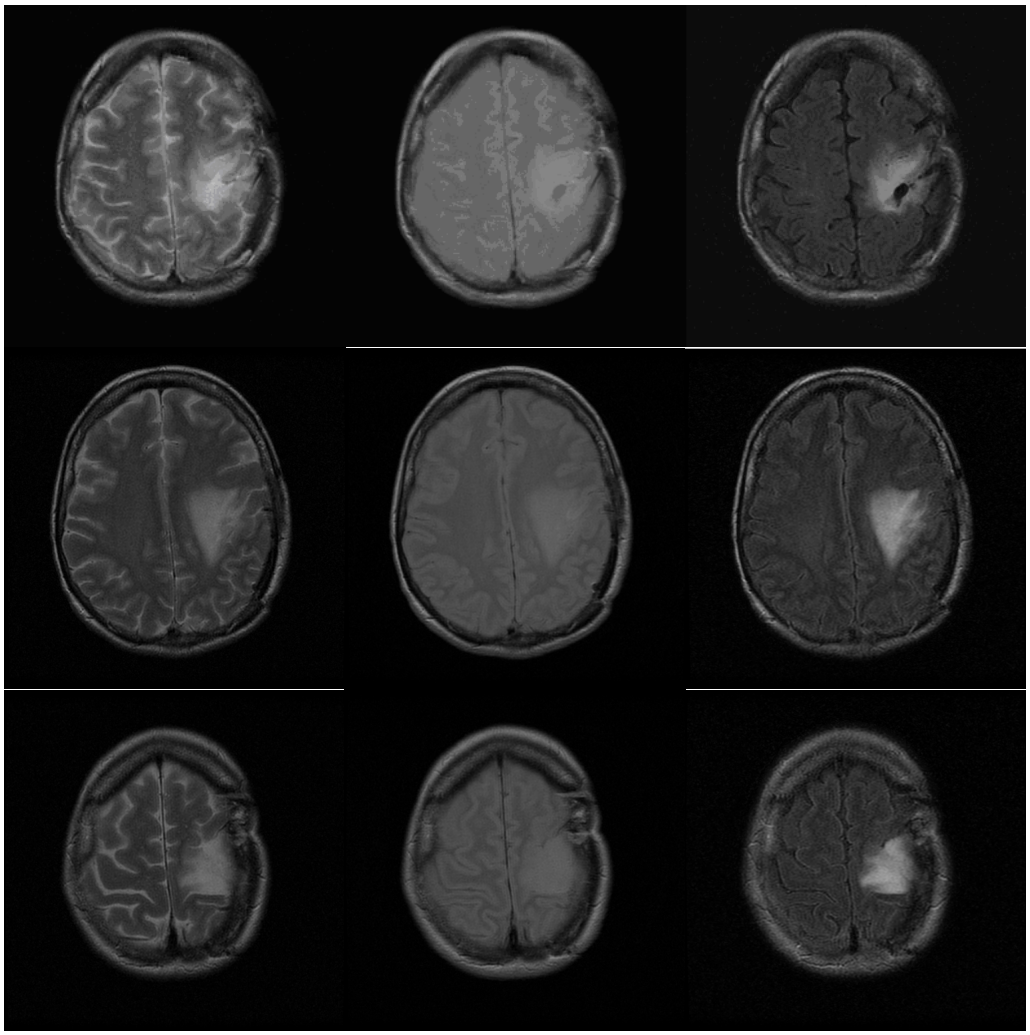


Figure 4.2 examples of multi-sequence MRI images of Patient 1
(from left to right corresponding to T2-weighted, PD-weighted and FLAIR images).

The MRI examples of Patient 2 to Patient 13 are as shown in Figure 4.3.

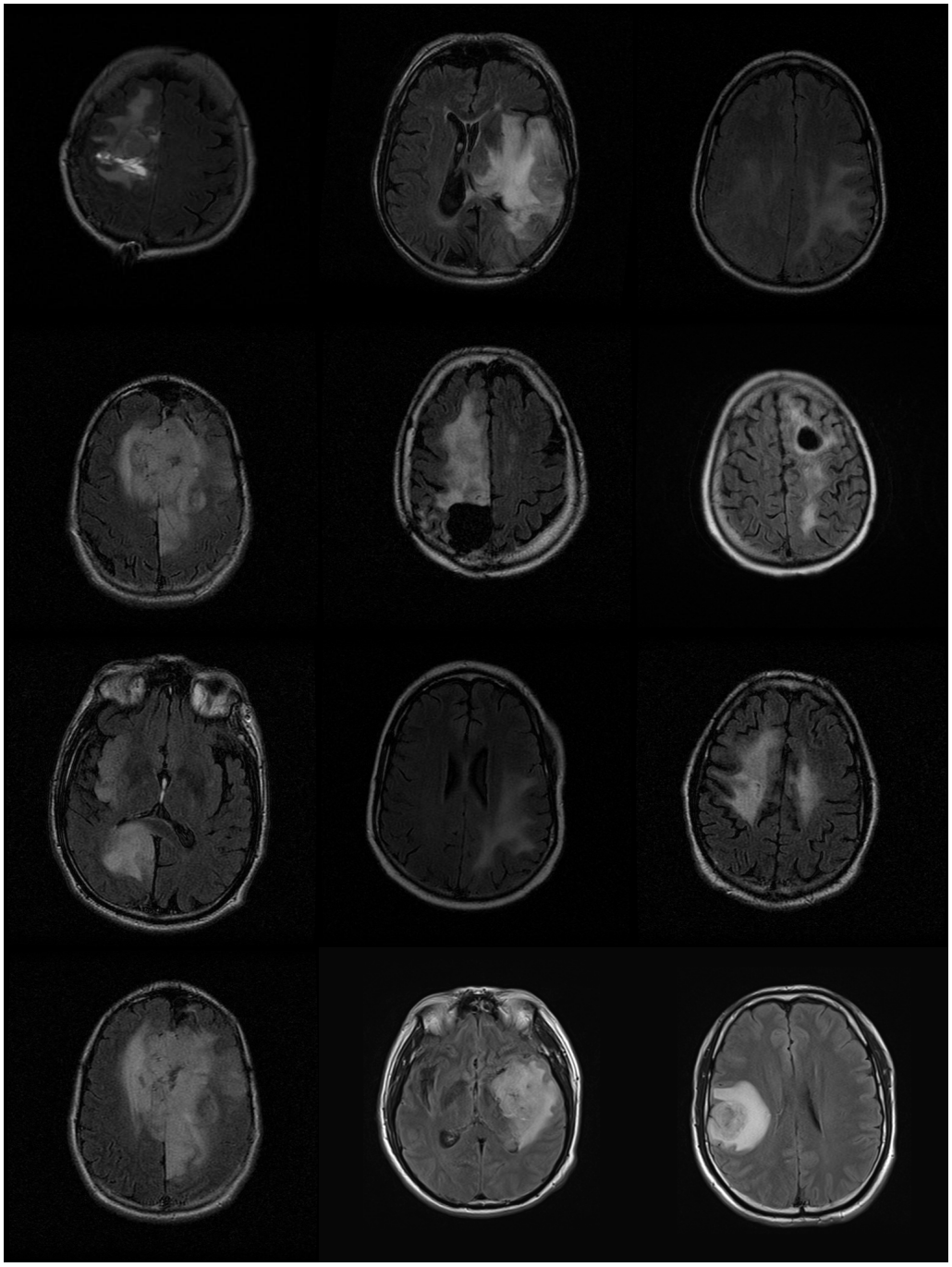


Figure 4.3 The MRI examples of Patient 2 to Patient 13.

4.2 Experimental Results

After obtaining the T2-weighted images, PD-weighted images and

FLAIR-weighted image sequence of specific examinaiton period, the first work is the registration of all the image sequences. The purpose of registration is to make the coordinates of the same pixel correspond to the same anatomical tissues in different MRI sequences in the brain. As long as the location of tissues and the pixels correspond to each otner in the multiple sequence images, the information can achieve the correspondence to ensure that the information fusion makes sense.

Registration is completed by the software SPM (Statistical Parameter Mapping) [122], which is specifically designed to achieve the analysis on brain imaging data sequences. These sequences can be not only from different imaging mechanisms, but also the same imaging sequences with the imaging objective changing over time. At present, the software has been used in the analysis of medical images, such as the functional MRI imaging (fMRI), Proton Emission Tomography (PET) and so on [123].

The Registration by SPM is based on the spatial voxel. First, all the images in the sequence are rearranged and normalized to a standard space to smooth; then build the parametric statistical model for each voxel, and the experimental data are described and the residual error of registration is estimated by using a General Linear Model (GLM) [124]. For different input images, GLM can be associated with the time-domain model or Bayesian probability estimation in order to achieve better registration results.

After the registration is completed, the next work is the segmentation of the tumor region totally. Only to ensure the accuracy of the tumor segmentation can ensure the accuracy of tissue classification within the tumor. It is necessary to select training sample points prepared for the SVM classifier in the examination data of the patient in the first period.. This selection of training sample points is fulfilled manually by a team of experienced doctors relying on their medical knowledge and clinical experience to ensure the accuracy of this process.

Because there are large individual differences among patients and the imaging condition is obviously different, such as the imaging equipment and the objective environment, the data to be dealt with from different patients are not relevant. There is also no priori knowledge which can be used, therefore it is very difficult to develop an automated analysis system with the universal applicability for any patient. In the proposed algorithm, it is the only step requiring manual interaction in the MRI image processing. The classifier can automatically obtain the data characteristics of tumor or the condition of the patients to track the state of disease in all the following examinations with the gained prior knowledge. In the actual experiments, the initial training is based on a total of 60 sample points (30 points within and outside the tumor respectively) which are randomly selected on a random FLAIR slice by the doctors. The system can automatically record the coordinate of their locations of sample points and their respective classes.

The extraction of feature vectors are based on the locations of sample points. Construct a square window with the sample points as the centers and variable side length and extract all the gray values, texture features and the derived features from

mathematical transformation for each window. All the obtained data are arranged into a row vector, which is the feature subvector of a certain weighted image sequence. Then connect the three corresponding feature subvectors in the order of T2-weighted images, PD-weighted images and FLAIR images, to form a final feature vector. Example of a feature vector is shown in Figure 4.4.

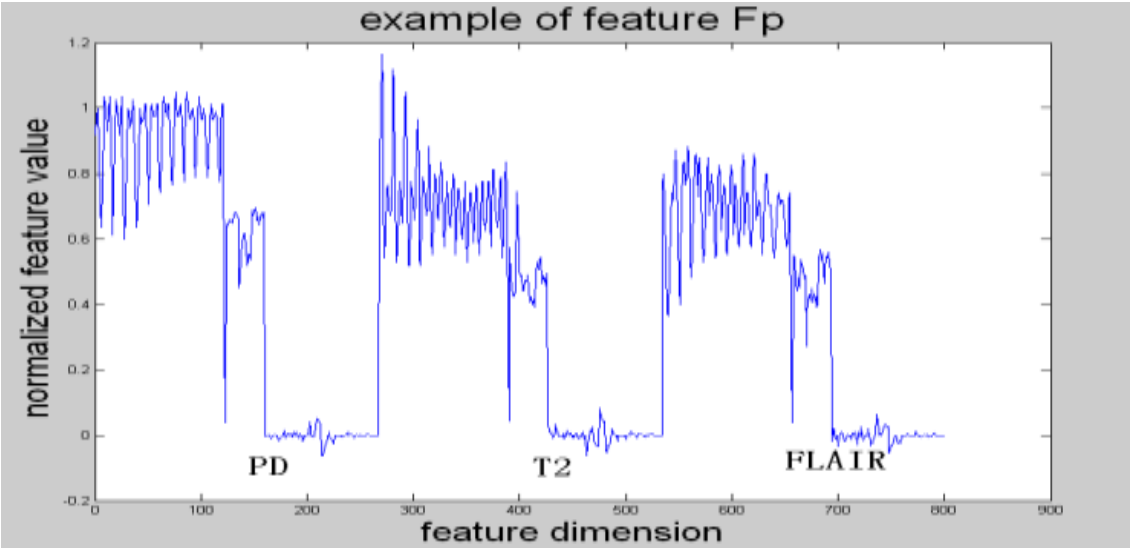


Figure 4.4 Example of a feature vector.

The dimension of the obtained feature vectors by the extraction method above is very high. Suppose the side length of the feature extraction window is N, then all the gray values have the total dimension of $N \times N$; texture features are selected as the mean value and standard deviation of all the gray values in the feature extraction window; the derivative feature is selected as the coefficients of two-dimensional wavelet transformation [125]. The dimension of the feature subvector, composed of the three parts above, is $N \times N + 2 + (N+1)^2$, and the dimension of a total feature vector which is composed of the three feanture subvectors is $3 \times N \times N + 2 + (N+1)^2$. As the number of the selected sample points is 60, the entire dimension of the training feature vector matrix reaches $60 \times 3 \times N \times N + 2 + (N+1)^2$. If N is equal to 11, then the dimension of the training matrix is 60×801.

Because of the limitations of imaging equipment, imaging environment and the position changes during the imaging process, the obtained MRI images will inevitably contain redundancy and noise. Meanwhile, it is impossible for all the features to play the same part in the classification, thus, according to the importance for the classification, the features are divided into two different parts: the important features and the second-important features. Besides the reasons above, large data amount will cause some excessive computational burden. Therefore, it is quite required to carry out the feature selection on the initial training matrix after the feature extraction.

If the side length of the feature extraction window is 11, the texture features are

selected as the mean value and standard deviation of all the gray values in the feature extraction window and the derivative features are selected as the low-frequency coefficients of two-dimensional wavelet transformation, use the Kernel Class Separability as the feature selection criterion and the Gaussian RBF as the kernel function to choose the important features from the extracted feature vector matrix, and then the normailize feature selector with the same dimension as the initial feature vector can be obtained through the optimization and training. Significance threshold is selected as 0.85, and the features of which the optimized values are more than 0.85 will be considered as the important features to be retained, while other features will be considered second important features to be removed. To complete the process of thresholding, we can get the binary feature selection vector (with just two kinds of elements 1 or 0, the former means the elements are retained and the latter means the elements are removed.). The example of ultimate feature selector is shown in Figure 4.5.

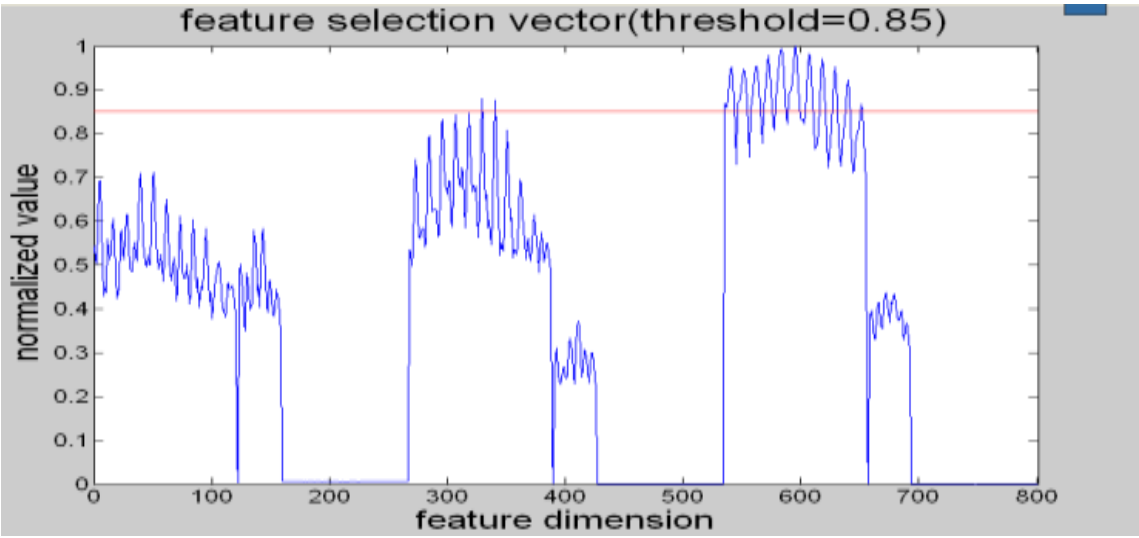


Figure 4.5 The example of feature selector.

Using the final feature vector matrix after the feature selection as the training matrix for the SVM classifier, we can obtain the appropriate classification model. The KCS feature selection method has integrated the training of kernel in both feature selection and SVM design to simplify the training process of SVM parameters and improve the training speed. The training of SVM is carried out by the popular Libsvm software [126], which has higher reliability and stability and is easy to operate.

the obtained training model of the SVM classifier is used for the tumor classification on all the data of patients in this examination period. After obtaining the initial tumor region, the tumor contour is improved by the region growing-based contour refinement algorithm to obtain the final tumor segmentation results. All the segmentation results of all the tumor regions are shown in the FLAIR images in Figure 4.6 and Figure 4.7. Examples shown in Figure 4.6 correspond to the Patient 1 of which

the images are listed in Figure 4.1 and images listed in Figure 4.7 corresponds to the images from Patient 2 to Patient 13 in Figure 4.3.

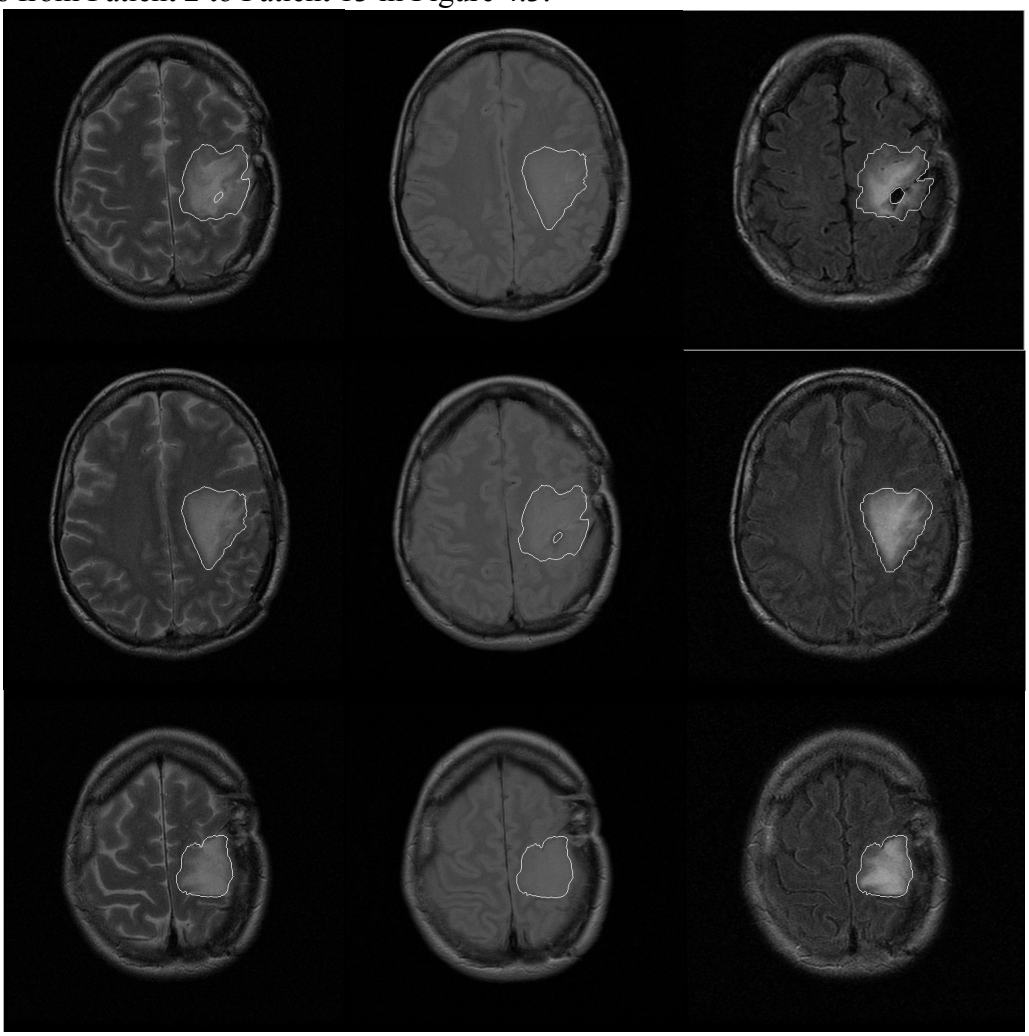


Figure 4.6 Tumor segmentation results of Patient 1

(from left to right are T2-weighted images, PD-weighted images and FLAIR images)

From the above results the effectiveness of the proposed method can be intuitively observed. For the relatively simple cases of tumor segmentation, such as the second case of the patient 1 as shown in Figure 4.6 and the case of Patient 12 in Figure 4.7, in which the shape of the tumor boundary is relatively simple and smooth and there are no various tumor lesions (the distribution of gray values is relatively uniform), the proposed system can well solve them; for the cases that the tumor boundary is complicated, such as Patient 3 and 4 as shown in Figure 4.7 and the cases that there are for other lesions within the tumor region obviously, such as Patient 2 and Patient 13, and so on, the proposed system can also well solve them.

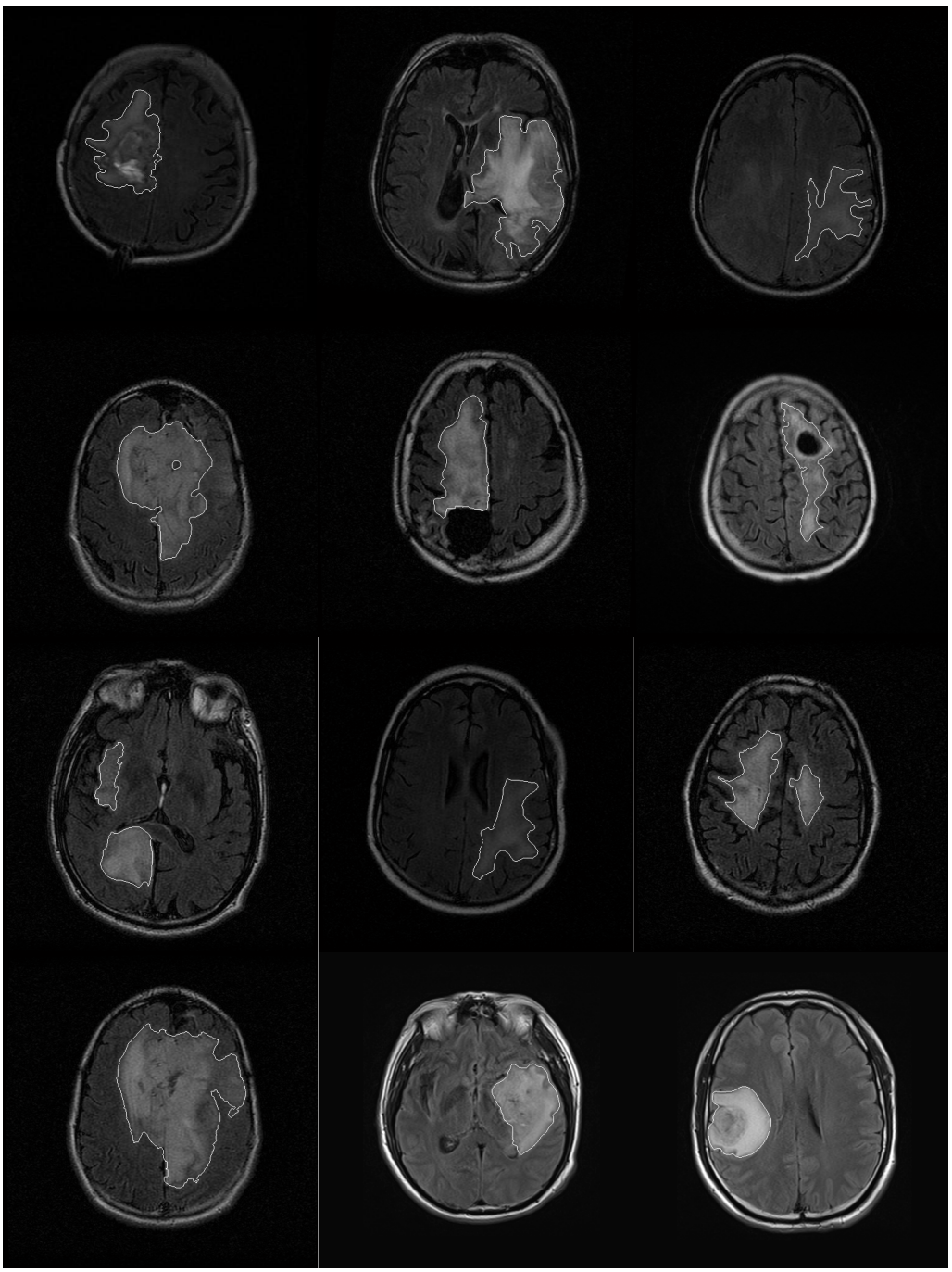


Figure 4.7 Tumor Segmentation results of Patient2 to Patient 13 (FLAIR images)

If there is only one tumor in the brain, the tumor must present a three-dimensional, continuous tridimensional structure. However, reflected in the two-dimensional images, it may appear broken or division in some slices. The occurred separation in the brain images, such as Patient 8 and Patient 10 in Figure 4.7, will increase the difficulty of classification as the shown tumors are more than one. And as the development of the

disease, there may be more serious lesions in the abnormal tissues or even necrosis, which will lead to the holes in the tissues, such as the first case of Patient 1 in Figure 4.6, Patient 5 and Patient 7 in Figure 4.7. It is also possible for the patients to fulfill an surgical operation such as Patient 6 in Figure 4.7, which will also leave holes in the brain. All the sudden jump of the gray value above will affect the classification performance to a certain extent. However, from the above segmentation results it can be observed that the proposed algorithm can also be a good solution for these more complex situations.

4.3 Comparison and Analysis of Feature Selection Methods

In this paper, the integration of the SVM classifier design and the feature selection process is proposed, in which the feature selection criterion is used to optimize and reduce the difficulties of parameters design in SVM classifier by optimizing the feature selection process. SVM utilizes the kernel function to define and describe the distance or similarity between vectors in the high dimensional feature space after mapping. The closer the distance is, the higher degree of similarity between vectors is and the stronger correlation between the features is. Therefore, the key of the SVM classifier design is the selection and optimization of kernel functions and their parameters. This chapter applies all the feature selection methods listed in Chapter 2 to the brain tumor segmentation system proposed in Chapter 3 to carried out a complete MRI image analysis, and obtain the corresponding experimental results and quantitative comparison and evaluation.

Figure 4.8 lists the exampls of Patient 2 with and without the feature selection (Kernel Class Separability).

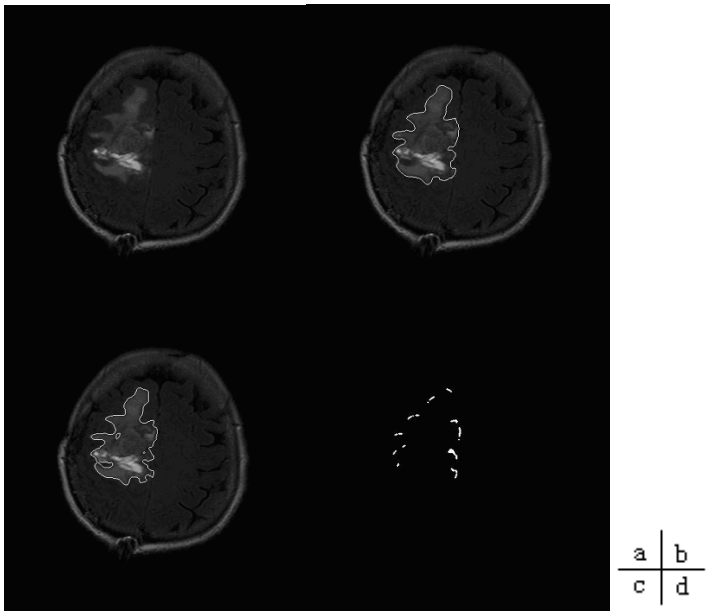


Figure 4.8 Examples of Patient 2 with and without feature selection (FLAIR images)

In Figure 4.8, (a) is the original image; (b) shows the classification results without feature selection; (c) shows the tumor classification results with the feature selection; (d) is the absolute defference between the two images (b) and (C). It can be observed from Figure 4.8 that, after the feature selection, besides the redundancy and noise, some useful information will be appropriately removed from the data. Therefore, the obtained tumor segmentation results reduces after the removal of some second-important information. This is the theoretical basis and preconditions for the proposed region growing-based contour refinement algorithm.

In order to verify the effectiveness of the feature selection methods, a cross-validation experiment is designed in the paper. The initial training feature vectors from 60 sample points are artificially divided into two groups, and there are 30 randomly selected feature vectors of each group. Then one group is directly used for training, or to first conduct the feature selection process and then training. After obtaining the SVM classifier model, the other group is used as test data to verify the validity of the model. As all the test data sets are from the same initial samples, the category labels of them are known. Therefore it is easy to get the classification accuracy of the model on the test samples. The cross-validation experiments are conducted 30 times for each feature selection method, and the average accuracy of the results for various methods is shown in Figure 4.9.

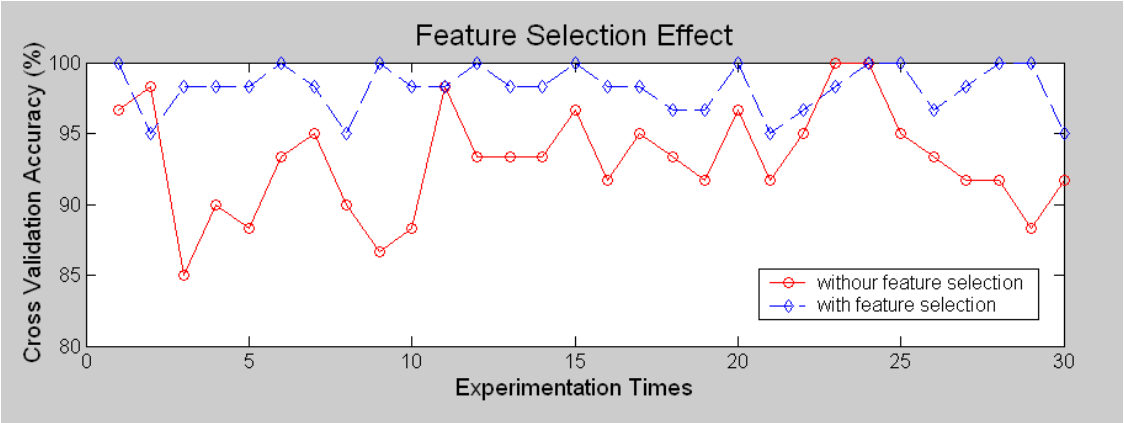


Figure 4.9 The cross-validation experiments of feature selection.

In Figure 4.9, the blue solid line denotes the average accuracy of 30 cross-validation after the the feature selection, while the red dashed line indicates the average accuracy of cross-validation without the feature selection. It can be intuitively observed from Figure 4.9 that, the classification accuracy of the cross-validation experiments is significantly improved after feature selection and its effectiveness is much higher than that in the case without feature selection. This has proven that feature selection can indeed remove the interference and noise in the data effectively, reduce the

influence to he classification performance of the system from the harmful information and to improve the accuracy of data classification.

In order to analyze the effectiveness of feature selection methods quantitatively, we select True Positive (TP), False Positive (FP) and False Negative (FN) as the comparison and evaluation criteria for the tumor classification results. And define the intersection is of the tumor segmentation results and the Ground Truth:

1. TP is the ratio of the number of pixels in the intersection and the total number of pixels in the Ground Truth;
2. FP is the ratio of the number of pixels in the part after the removal of the intersection from the tumor segmentation results and the total number of pixels in the tumor segmentation results;
3. FN is the ratio of the number of pixels in the part after the removal of the intersection from the Ground Truth and the total number of pixels in the Ground Truth;
4. Total error is the sum of FP and FN.

Finally, average all the data from all the slices of different patients in different examination periods and the quantitative comparison on the effectiveness of all the feature selection methods can be obtained. Because the factors that affect the accuracy of the final results are the contour refinement process in addition to the feature selection methods, the comparison and discussion on the effectiveness is divided into cases, with and without contour improvement respectively. The obtained quantitative results are shown in Table 4.1 and Table 4.2.

Table 4.1 Quantitative comparison on the effectiveness of the feature selection methods (without contour improvement)

Feature Selection Methods	TP/%	FP/%	FN/%	Total Error/%
PCA	91.1	8.6	8.9	17.5
Kernel PCA	91.5	9.4	8.5	17.9
KCS	95.0	5.9	5.0	10.9
Kernel F-score	90.7	6.7	9.3	16.0
SVM-RFE	94.2	7.3	5.8	13.1
CFS	88.6	4.7	11.5	16.2
NFS	85.3	10.7	12.7	23.4

Table 4.2 Quantitative comparison on the effectiveness of the feature selection methods (withcontour improvement)

Feature Selection Methods	TP/%	FP/%	FN/%	Total Error/%
PCA	93.9	8.8	6.1	14.9
Kernel PCA	94.3	11.6	5.7	17.3
KCS	98.9	4.5	1.1	5.6
Kernel F-score	91.7	3.6	8.3	11.9
SVM-RFE	98.5	4.2	1.5	5.7
CFS	90.1	2.3	9.9	12.2
NFS	87.5	9.3	12.5	21.8

In Table 4.1 and Table 4.2, NFS refers to the situation of the direct classification on the feature vector matrix after feature extraction without the use of feature selection. From Table 4.1 and Table 4.2 it can be observed that the process of feature selection can indeed significantly improve the accuracy of tumor classification. If there is not feature selection in the process, the feature vector matrix will contain a large number of redundancy, interference and noise, which leads to not only a great amount of computation and training, but also seriously reduction of tumor classification accuracy. In addition, the performances of Kernel PCA, Kernel F-Score, SVM-RFE and KCS are different. The performance of methods with kernel function is significantly higher than other types of algorithms. From the experimental results, it can also explain that feature selection based on Kernel functions can indeed better match the SVM classifier.

In the experiment, just the Gaussian radial basis function is used in this paper. In Kernel PCA and Kernel F-score, the parameters of kernel functions are the experimental data selected by experience; in SVM-RFE and KCS, the parameters of kernel functions are selected based on cross-validation method, which is the the average of 10 better parameters. The introduction of cross-validation is only to prove that the performance of kernel-based methods is significantly relative to the parameters of kernel function. In fact, the parameters obtained from cross validation is not really the optimal parameters, but to some extent, it can improve the accuracy of results.

Evaluation criteria for the effectiveness of the feature selection methods are efficienty in addition to the classification accuracy. Despite the different methods of feature selection, the initial feature vector has the same dimension of 801. After the feature selection only a few of the features elements are retained, of which the number accounts for the similar ratio fro the original feature vectors. It will not have a greater impact on the time performance of the classifier [126], so about the issue of efficiency the dimension of the retained feature elements and the run-time of feature selection

processing are considered in this paper. The quantitative results of the criteria proposed above under the same hardware environment and software environment is shown in Table 4.3.

Table 4.3 Efficiency comparison of feature selection methods

Feature selection methods	Retained dimension	Time/min
PCA	50	1
Kernel PCA	39	2
KCS	26	5
Kernel F-score	55	4
SVM-RFE	36	8
CFS	45	25
NFS	801	23

In Table 4.3, the parameter "time" denotes the required time for one feature selection process on a slice and it is an average value. The time-consuming data after this section is also the same that they are the average time spent value for the relevant operations on one slice. Draw a curve based on the total data of Table 4.2 and Table 4.3, and we can obtain the performance curve of the feature selection methods. Because some data are close to each other, which will make serious overlap in the curve, the FP values are added 20, while FN 40 and total error 60 to improve the visual effects of the curve. The performance curve is shown in Figure 4.10.

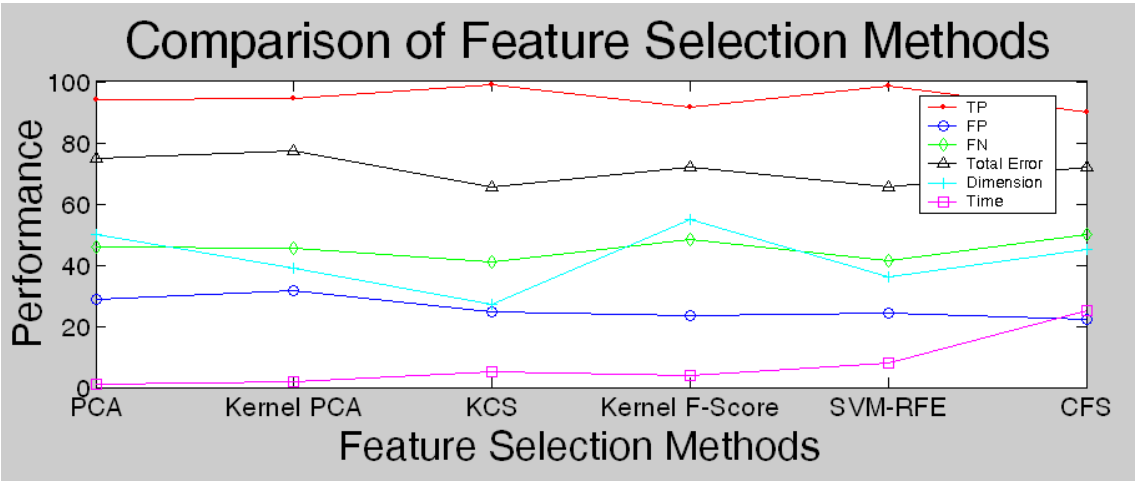


Figure 4.10 Performance curves of feature selection methods.

In the above 6 feature selection methods, comprehensively compare of all the

performance indicators and it can be observed that, PCA is simple and fast, but its accuracy needs to be improved; Kernel PCA has higher accuracy and efficiency, but the results are affected by the parameters of kernel function to great extent; kernel F-score preserves the most feature elements, and its threshold restricts the classification a lot, of which the relevant selection method needs to be improved instead of using the mean value of all the F-Scores of the feature elements as the threshold; CFS has a higher accuracy than NFS, but the efficiency reduces in a large scale. Comprehensively considering the performance of the entire system, including the accuracy, efficiency and the final dimension of the retained feature elements, KCS and SVM-RFE are more suitable for the SVM classifier. The both have the similar effects and their parameters can also be available through cross-validation method, but KCS can ingegrate the parameters selection in both feature selection and SVM classifier design, which reduces a step of parameter selection and some difficulties to some extent.

Because of the above-mentioned advantages of KCS, in the classification of brain tumors in this paper, we choose KCS as the feature selection criterion, and the following parameter selection, feature validity analysis and comparison with other methods are generally discussed with KCS.

4.4 Selection of Some Key Parameters

The key parameters in the experiments discussed here are the side length of the feature extraction window and the threshold in KCS criterion. These two parameters will directly affect the subsequent analysis and processing of other factors.

The side length of the feature extraction window directly affects the dimension of the extracted feature vectors. The longer the side is, the higher the dimension of feature vectors is. Although more useful information is contained as the increasing of the side length, meanwhile the redundancy, interference and noise will inevitably increase to increase the calculation of feature selection and other following treatment. Moreover, it is less likely that all the pixels in the same feature extraction window belong to the same category.

Similarly, if the threshold is chosen too small, the dimension of reatained feature elements is high, and the feature selection does not make any sense. Otherwise, if the threshold value is chosen too high, almost all the feature elements are removed, the classification process can not be properly conducted.

In the experiment, as the threshold increases from 0.60 to 0.90 with a step size of 0.05 and at the same time the size of the feature extraction window from 5×5, 7×7 to 15×15 (the side length of the feature extraction window is odd in order to make the sample points locate at the centers of the feature selection windows), the average dimensions of the retained feature elements from the feature matrix after the feature selection on all patients' cases are shown in Table 4.4.

In Table 4.4, the threshold value of 0 corresponds to the total dimension of all the

feature elements. It can be observed from Table 4.4 that, the dimension of retained feature elements becomes stable with slight differences but not any jump in general when the window size achieves 11×11; after the threshold is superior to 0.85, the remaining feature elements are insufficient to complete the classification process. From the above analysis derived from the table, the threshold of KCS is selected as 0.85, while the size of the feature extraction window is selected as 11×11.

Figure 4.4 Influence to KCS by the size of feature extraction window and threshold

Threshold	0	0.60	0.65	0.70	0.75	0.80	0.85	0.90
5×5	189	84	67	45	33	15	6	3
7×7	345	97	85	63	47	22	9	5
9×9	549	156	117	88	53	39	13	9
11×11	801	203	135	114	85	58	26	15
13×13	1101	376	233	145	93	67	28	18
15×15	1449	482	268	165	107	75	29	24

In order to compare the above verification process, the same experiments are carried out again in the process with PCA as the feature selection method. Because PCA does not require threshold, this experiment just focuses on the selection of window size. The average experimental results in all patients are shown in Table 4.5.

Table 4.5 Influence to PCA by the size of feature extraction window

Window Size	5×5	7×7	9×9	11×11	13×13	15×15
Dimension	19	28	34	37	41	43

From the Experimental Analysis of PCA, it can also be observed that, when the size of the feature extraction window reaches 11×11, the remaining dimension of feature elements has no significant changes. Therefore, the final size of feature extraction window is selected as 11×11.

4.5 Effectiveness of Features

In the feature classification, various features can be proposed and used. Different features will have great impact on the performance of classifiers. Feature selection will also directly affect the final classification results. In this article the effectiveness of 4 kinds of features are discussed, namely two-dimensional wavelet transformation coefficients, DCT transformation coefficients, fractal features and three-dimensional features, respectively. The combination of features from multiple input or multiple sequence are effectively fused by the KCS feature selection criterion to improve the classification accuracy and reduce computation time and the computational complexity to a relative degree.

4.5.1 Wavelet Transformation and DCT Transformation Coefficients

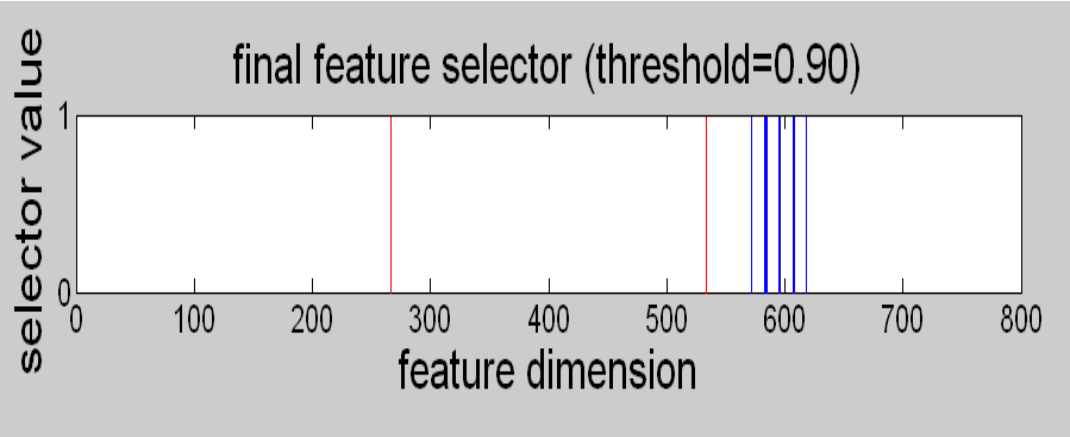
In the previous experiments to obtain the results in Section 4.2, the gray values, the texture features and low frequency coefficients of two-dimensional wavelet transformation are extracted as the total feature elements. The intensities are the most essential features in MRI images, which are directly extracted from the image without any treatment. The texture features Reflect the global or regional statistical properties of a local area, which are from statistics data of the intensities. The low-frequency coefficients of two-dimensional wavelet transformation are derived features, which are obtained from the basic intensities and texture features through a number of mathematical transformation. Due to the mathematical transformation generally has clear physical meaning, so the derived features can also represent the corresponding data characteristics in the transformation domain.

The intensity is the most basic features, thus, its effectiveness is naturally self-evident. This article will discuss here the validity of the derived features, such as the wavelet transformation coefficients and Discrete Cosine Transformation (DCT) coefficients. DCT coefficients are also extracted in the feature extraction window with

the sample points as the centers. Using DCT coefficients as another derivative features, select the elements from low-frequency coefficients of two-dimensional wavelet transformation and DCT transformation coefficients by the feature selection process in the same experimental hardware environment (the same computer environment and the same softwares) and the same experimental conditions (the same input sequences and the same feature selection method KCS), and compare the effectiveness of them.

KCS needs to set the threshold to select the most important features, and the threshold will directly determine the the dimension of the retained feature elements. If the threshold is selected too large, few feature elements are retained and almost all the features are removed, which lead to that the it will be unable to complete the classification process; if the threshold is selected too small, the dimension of the retained feature elements are high, which can not effectively remove the redundancy and noise and increase the computational burden to some extent. In the experiment, 0.60 and 0.90 correspond to the two critical situations of the threshold selection. The discussion on the number of retained features is conducted based on the both critical situations. In actual, considering the compromise of the computational complexity, the dimensions of the retained features and the classification accuracy of the results, the final threshold is determined to be 0.85.

Using the low-frequency coefficients of the two-dimensional wavelet transformation as the derived features, the results after feature selection and thresholding are shown in Figure 4.11.



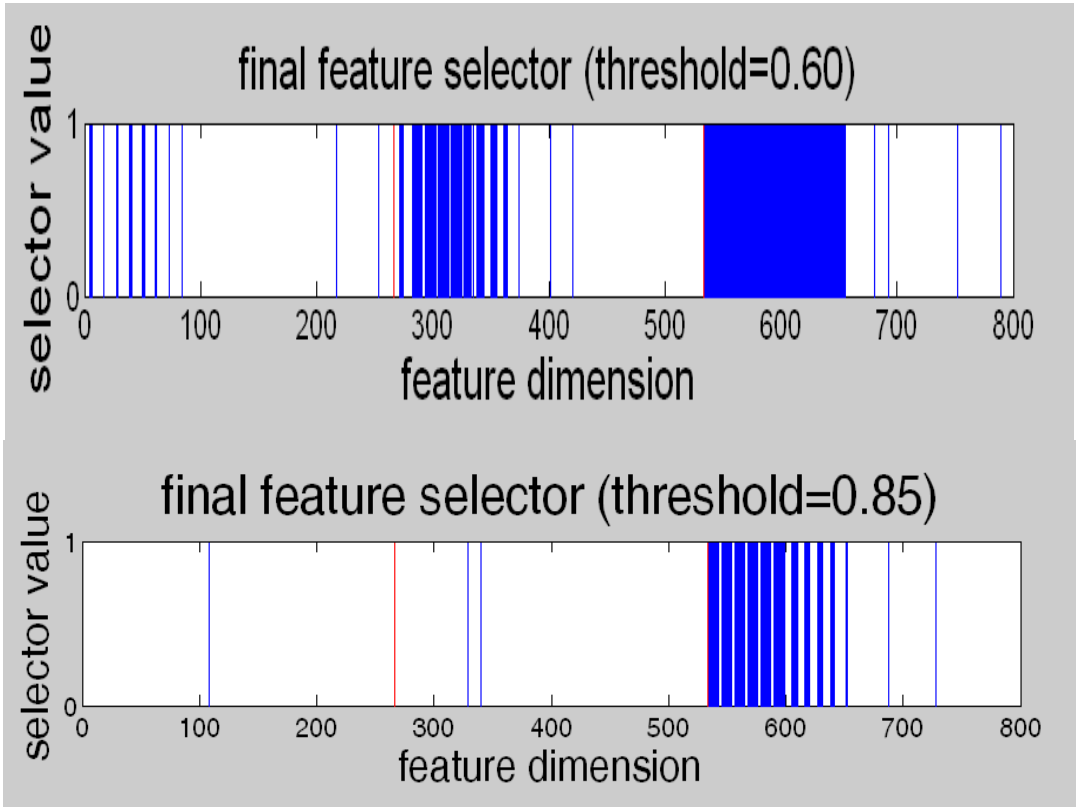


Figure 4.11 Feature selection on 2D wavelet transformation coefficients

In Figure 4.11, the thresholds of 0.60 and 0.90 correspond to the two critical conditions and the threshold value of 0.85 is used for normal classification. The horizontal axis indicates the corresponding position coordinates of the feature elements. The blue lines correspond to the locations of the preserved feature elements after feature selection, and the red line corresponds to the boundary of the T2-weighted, PD-weighted and FLAIR MRI sequences with the corresponding coordinates as 267 and 534.

Using the DCT transformation coefficients as the derived features, the results after the feature selection and thresholding in the experiment are shown in Figure 4.12. The threshold value of 0.60 and 0.90 correspond to the two boundary conditions, while the threshold of 0.85 is used to compare with the situation of low frequency coefficients of the two-dimensional wavelet transformation.

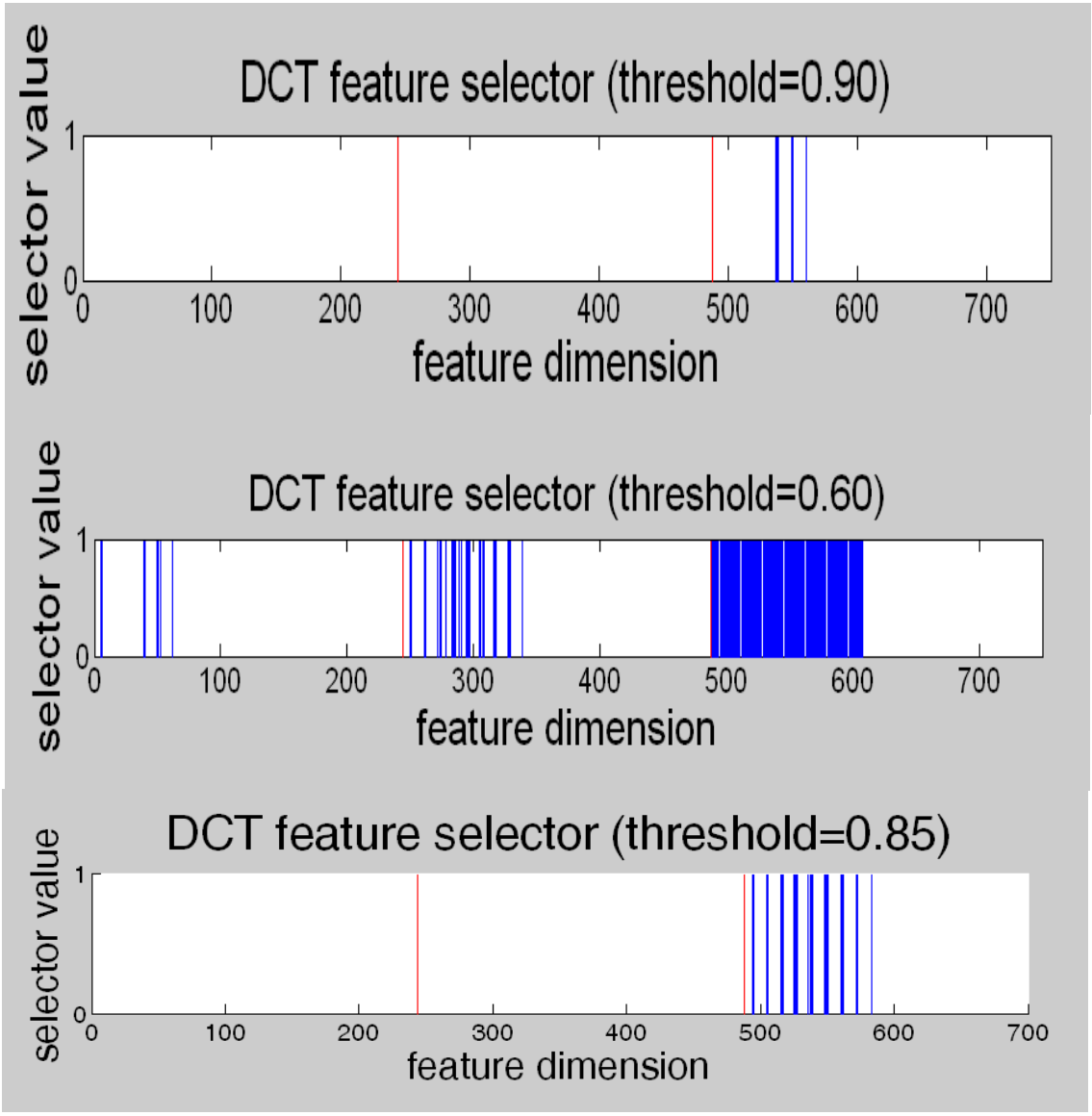


Figure 4.12 Feature selection on DCT coefficients.

Similar to Figure 4.11, the blue lines in Figure 4.12 correspond to the locations of the preserved feature elements after feature selection, while the red line corresponds to the boundary of the T2-weighted, PD-weighted and FLAIR MRI sequences. Different from the situation of using wavelet transformation, the corresponding coordinates are 244 and 488.

From the comparison of Figure 4.11 and Figure 4.12, it can be observed that, when using KCS as the feature selection method, the DCT transformation coefficients will be considered as the second important features and removed no matter how small the threshold is selected in the effective threshold range. On the contrary, low-frequency coefficients of the two-dimensional wavelet transformation are still retained even though the threshold increases to higher values. From the physical sense, the two-dimensional wavelet transformation reflects the multi-scale resolution capabilities

and frequency characteristics of the image. Therefore, compared with a single DCT transformation, it becomes more important.

4.5.2 Effectiveness of Fractal-based Features

Fractal-based features [71], which are used to describe the complex geometric properties of the natural objects, are also discussed in this paper. The fractal characteristics of objects is measured by the Fractal Dimension (FD) [127]. Different from the traditional dimensions in Euler geometry, fractal dimension is not a positive integer, but a real number, and different fractal dimensions represent different texture structures in the images. Generally, the more complex the structure of the object is, the the higher the corresponding value of the fractal dimension is. Fractal-based features has been used in many medical areas, such as the complexity evaluation of the cerebral cortex [128] and the detection of small lung tumors [129].

The traditional methods for calculating the fractal dimension are box-counting, the improved box-counting, piecewise triangle prism surface area (PTPSA), etc. These methods have been successfully applied to two-class tumor classification problem in single-modality medical image [130-131]. the improved PTPSA algorithm is used in this paper to calculate the fractal dimension in order to apply to the analysis of multiple-input or multi-modality image.

Based on the improved PSPTA algorithm, the FD is calculated as follows:

1、 The input image is divided into a number of square sub-image areas with the same size and the side length as r. For each sub-image area, the gray values of the four boundary vertices can be directly obtained from the images. These four gray values are recorded as p_1, p_2, p_3, p_4;

2、 Consider the 4 gray values above as the heights in the third dimension and Project the four points into the third direction based on the four heights. Corresponding to the four vertices p_1, p_2, p_3, p_4, the four projectors are recorded as A, B, C, D;

3、 The mean value of p_1, p_2, p_3, p_4 is considered as the gray value of the center in the sun-image area, which is also the height in the third dimension. Repeat the projection process, and the corresponding projection points of the center pixel can be obtained as p_c;

4、 The above five vertices A, B, C, D and p_c form four triangles in the 3-dimensional space, which are recorded as ABE, BCE, CDE, ADE. Calculate the areas of the four triangles;

5、 Calculate the FD value of the small area by Equation (4-1). S refers to the areas of the small triangles.

$$FD = \frac{\log(S_{ABE} + S_{BCE} + S_{CDE} + S_{ADE})}{\log r} \qquad (4\text{-}1)$$

The above procedure can refer to Figure 4.13 [71].

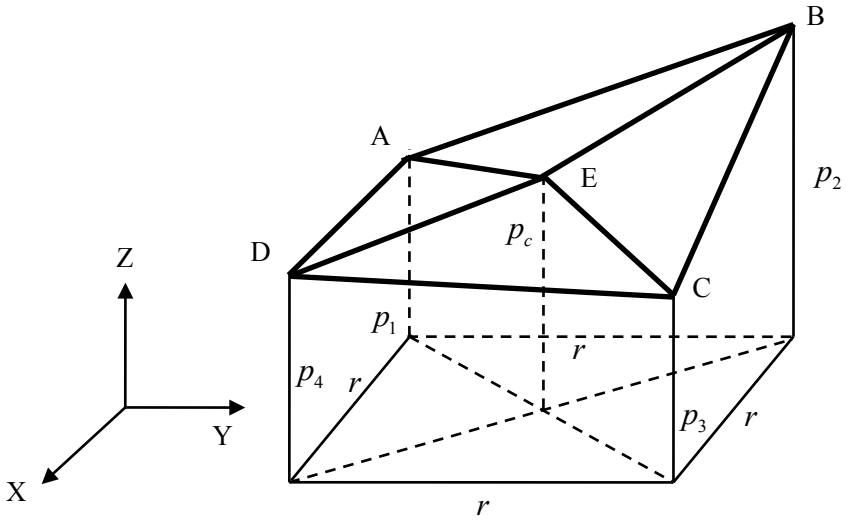


Figure 4.13 Schematic diagram for the calculation of fractal-based features

According to the above process, all the fractal dimensions corresponding to the sub-image regions are calculated and arranged into a vector to complete the extraction and description of fractal-based features.

In the experiments of this paper, the fractal-based features are completely extracted in the windows centered at the sample points, therefore, in one MRI sequence, the fractal–based features just have one dimension for each point. The dimension of total fractal features is the same as the number of the initial sample points. The computation of the fractal-based features involves the acquisition of spatial projection, spatial distance and spatial area, which increases the considerable complexity in the calculation.

Using the same intensities and texture features as the DCT coefficients and two-dimensional wavelet transformation coefficients, add the fractal-based features and select the features by the same KCS criterion. The obtained feature selector is shown in Figure 4.14.

The red and blue lines in Figure 4.14 have the same meaning as those in Figure 4.11 and Figure 4.12. Here the separate coordinates are 124 and 248. It can be observed from Figure 4.14 that, fractal-based features are also removed after the feature selection, which is equal to no contributions for the fractal-based features to the classification. Thus, if the number of extracted fractal-based features is small or these features are just relevant to a limited part of the image, it is unable to extract the data reflecting the geometric characteristics of the object; Similarly, if the selection of the sub-image areas are very random, it is possible to completely cover the areas with complex geometric characteristics. Therefore, the division or selection of the sub-image areas will have a great impact on the validity of the fractal-based features.

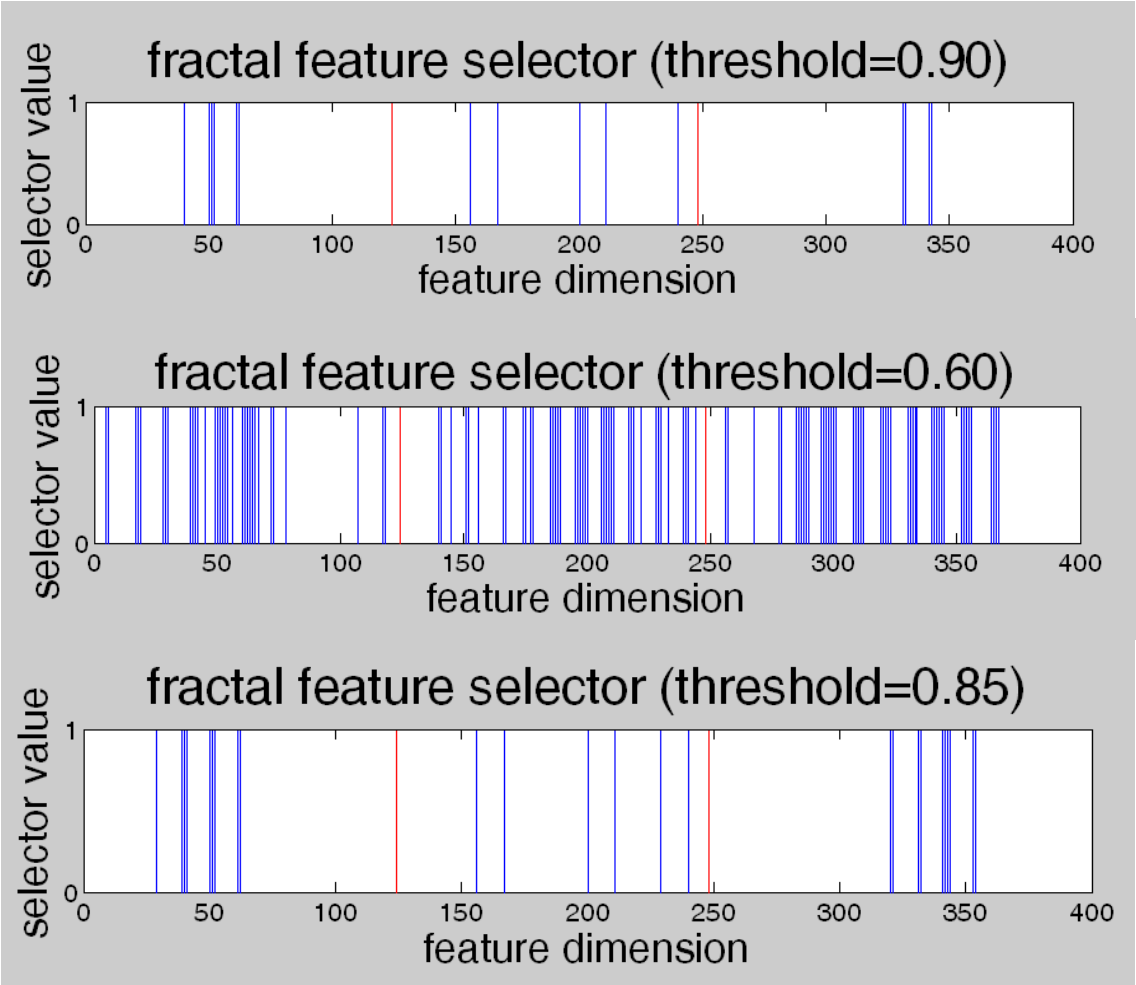


Figure 4.14 The selection of fractal-based features

4.5.3 Effectiveness of 3D Features

MRI images appear to be the three-dimensional structure. As to a single MRI image, the classification is based on the data between the pixels, and to the entire MRI sequence, the classification is based on the relationship between the voxels. Thus, a sequence of MRI images certainly contains the corresponding three-dimensional information implicitly, and it is likely to have some relevance between adjacent slices. In order to explore the relationship between the adjacent slices, a number of researchers have conducted depth analysis of the positive discussion on how to use the three-dimensional features for the classification of tumor and a lot of processing algorithms have been proposed.

In [132], the extraction method in six neighborhood is proposed, that is, to extract the gray values of the points located at the top, bottom, left, right, front and back of sample points as the three-dimensional features. In the classification, use the histogram analysis, morphological operations and symmetry analysis theory to obtain the results of the initial tumor classification; re-use parametric deformation model to optimize and

improve the initial results under the constraints of fuzzy spatial relationships (such as topology relationship and approximate location of the tumor).

In [133], first select the region of interest (ROI) containing the tumor rather than the sample points to extract the relevant information in the region for the calculation of the membership function; and then use a series of pre-defined fuzzy criteria to obtain the certain output for each pixel after the removal of the fuzzyness. This method can track the tumor area. In the tracking process, the previous tumor segmentation results are considered as the ROI for the data in the next examinations. In the whole algorithm, ROI is the only prior knowledge, which the algorithm is highly dependent on. Therefore, ROI are in general described manually before the segmentation and its accuracy will be important for the tumor segmentation.

In [134], a global threshold is used to construct the speed function in the level set algorithm and the threshold is used as the control parameters of convergence. The limitations of this method are the need to pre-set the reference slice which is required to contain the tumor area as large as possible in order to get the intensities of the tumor points as priori knowledge to update the threshold. In addition, the method is appliable to the tumor with convex shapes, thus, the applicability and scalability of the methodology is not strong.

With the great development of computer-aided analysis, there is still no effective three-dimensional features. The six neighborhood extraction method proposed in [132] is extended here, that is to fuse the feature vectors extracted from the same position of the two adjacent slices with the feature vector in the slice to be segmented to increase the amount of information. Specifically, the two feature vectors are extracted based on the same method, thus, each feature vector contains the information of three MRI sequences-T2-weighted, PD-weighted and FLAIR. Finally, the three feature vectors are connected together to form a total feature vector, which contains the three-dimensional information to use.

The dimension of the feature vector constructed in the way above is three times as the dimension of the original feature vector. Set the side length of the feature extraction window is 11 and use the mean value and standard deviation as the texture features while the low frequency coefficients of the two-dimensional wavelet transformation as the derived features, thus, the total dimension of a feature vector from a sample point is $3 \times 3 \times 267 = 2403$.

Using the same feature selection criterion KCS, the retained feature elements is similar to the feature selection carried out on a single slice, therefore, the results are not repeated here. It should be noted that although the dimensions of the original feature vectors increase to three times as that in a single slice, the number of retained feature elements after feature selection is not necessarily three times as that in a single slice, which is determined by the nature of different slices and the implied relationship between the slices.

The above four features are used respectively to complete the tumor detection on

all the patients' data. Compared with the Ground Truth from the manual delineation of the doctors based on medical knowledge and clinical experience, the tumor segmentation results after the contour refinement are quantitatively evaluated. Choosing TP, FP and FN as the criteria, the average values on the entire slices of all the patients are shown in Table 4.6. Similarly, in order to exclude the effects of contour improvement to the final results, the comparison which just takes the effect of selected features into account is also conducted without the contour refinement. The results in this situation are shown in Table 4.7.

Table 4.6 Quantitative evaluation of the effectiveness of features with contour refinment

Feature	TP/%	FP/%	FN/%	Total Error/%
Wavelet Coefficients	98.9	4.5	1.1	5.6
DCT Coefficients	93.1	7.8	6.9	14.7
Fractal Features	94.7	6.6	5.3	11.9
3D Features	98.7	4.9	2.3	7.2

Table 4.6 Quantitative evaluation of the effectiveness of features without contour refinment

Feature	TP/%	FP/%	FN/%	Total Error /%
Wavelet Coefficients	95.0	5.9	5.0	10.9
DCT Coefficients	91.3	8.9	8.7	17.6
Fractal Features	91.1	10.5	8.9	19.4
3D Features	95.3	6.1	4.7	10.8

From the data analysis of Table 4.6 and Table 4.7, it can be observed that, the basic features play an important part in the tumor classification: although the DCT coefficients and fractal characteristics are removed after the feature selection, but the both situations can still achieve a high segmentation accuracy; and the effect of contour refinement in the two cases is not very large (1.8% and 3.6%); synthesize the analysis above, contour refinement step can indeed increase the accuracy of the results to some extent, but the accuracy is mainly determined by the basic features of the intensities and texture features. This also illustrates the important role of the basic features in the classification problems.

Without the contour refinement process, the classification accuracy with the three-dimensional features is slightly higher than that of a single slice, which means that

feature selection can be effective in the removal of the redundancy, interference and noise and the increase of tumor classification accuracy by increasing the amount of information. However, increasing the amount of information will also increase the computational burden to influence the system performance at the same time, which is the efficiency problem to be solved in real applications.

4.6 Discussion and Evaluation of Contour Refinement

4.6.1 Effectiveness of Contour Refinement

For the initial tumor classification results, due to lack of prior knowledge, the differences among individual patients and the impact of feature selection process, there will inevitably be wrong decisions and missing of points near the border. Feature selection can eliminate most of the secondary features, which leads to a smaller the results of the initial tumorclassification, therefore, a region growing-based contour refinement algorithm is designed in this paper to improve the boundary accuracy. The specific details of the algorithm can be found in Chapter 3.3.

Considering the multiple sequence input data of Patient 1 in Figure 4.2 as examples, the effect of the contour refinement is shown in Figure 4.15.

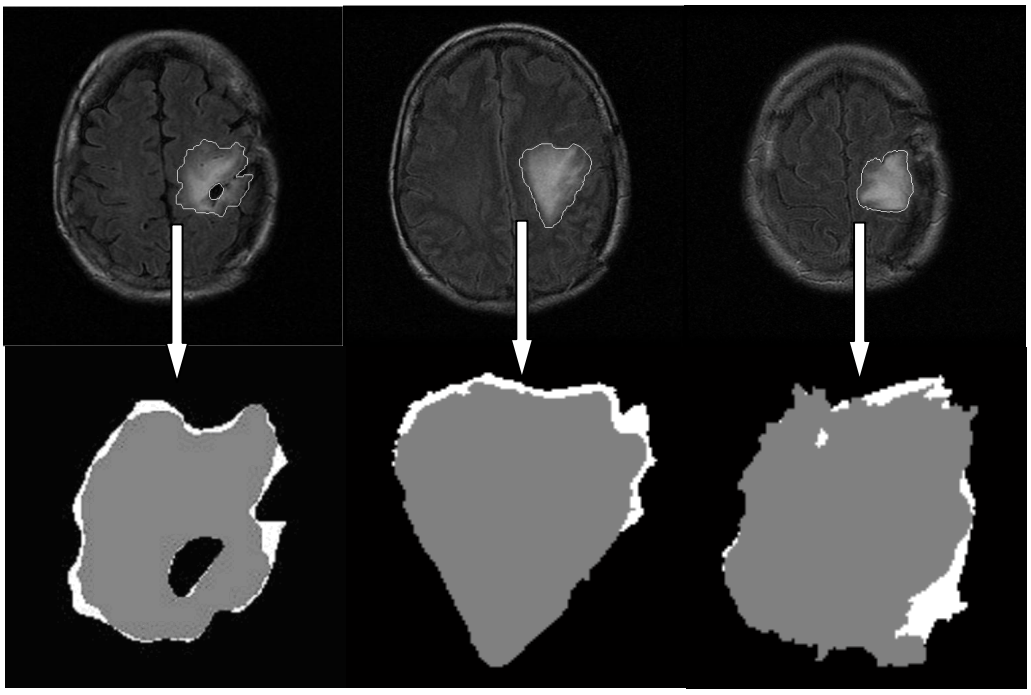


Figure 4.15 Examples of contour refinement on Patient 1 (FLAIR images)

In Figure 4.15, the top row of the images are the final results of tumor segmentation, and the following row is the enlargement of the local areas of the segmented tumor region. The gray parts in the images represnt the initial separated

tumor region, and the white parts are the supplementary region after the region growing-based contour refinement. It can be observed from the above figures, even if the tumor contains necrosis and holes, the proposed conrour refinement algorithm can still further improve the tumor contour.

Using the low-frequency coefficients of the two-dimensional wavelet transformation as the derived features and KCS as the feature selection criterion, the contour refinement experiments are carried out on the MRI image sequences. The evaluation criteria are also selected as TP, FP and FN, and the average evaluation values are shown in Table 4.8.

Table 4.8 Quantitative evaluation of contour refinement

Criteria	TP/%	FP/%	FN/%	Total Error/%
Without contour refinement	95.0	5.9	5.0	10.9
With contour refinement	98.9	4.5	1.1	5.6

It can be seen from Table 4.8 that, after the feature selection-based data fusion and classification, the selection and removal of the secondary information leads to the decrease of the initial region of tumor classification; however, the region growing-based contour refinement algorithm can well improve the missing tumor region to increase the classification precision and accuracy of the system.

4.6.2 Comparison of Contour Refinement Methods

Active Contour Model (ACM), also known as the Snake method, is an important component of deformation models, which has been well applied to medical image segmentation [43-46].

ACM algorithm first defines an initial curve, then define the corresponding energy function of the curve. By optimizing the energy function, the curve evolves, and finally converges to the optimal value which is consedered as the result of the final curve evolution.

The energy defined in ACM algorithm includes the internal energy and the external energy. The former mainly depends on tension and rigidity (corresponding to the first and second derivative of the curve function) in the evolutionary process to ensure the smoothness of the curve profile; the latter is the driving force of the curve evolution, which facilitates the curve to keep evolving in the direction of smaller potential energy. Commonly used the external energy driving force is the negative gradient. The more closer to the edge the area is, the greater the gradient is and the smaller the external energy and the driving force is. It will be more difficult for the curve to evolve.

ACM algorithm just relies on the edge information, thus it is easy to fall into the local optimal value in the evolution; furthermore, the results often require interpolation and internal padding to get a complete segmentation region. ACM is not useful to the area without drastic changes near the edge, because the gradient of the curve in this region are relatively small, resulting in the limited driving force and the external energy, thus it is unable to facilitate the curve to evolve toward the direction of smaller potential energy. These reasons also limit the accuracy and applicability of ACM algorithm.

In the experiment, the primary tumor segmentation contour is considered as the initial curve because the system has obtained the initial contour of the tumor after the SVM classification. Approximately fit the curve function, define the internal energy and external energy function, and improve the curve boundary through the optimization. As to the examples of Patient 1 in Figure 4.2, the visual results about the effect of ACM are shown in Figure 4.16.

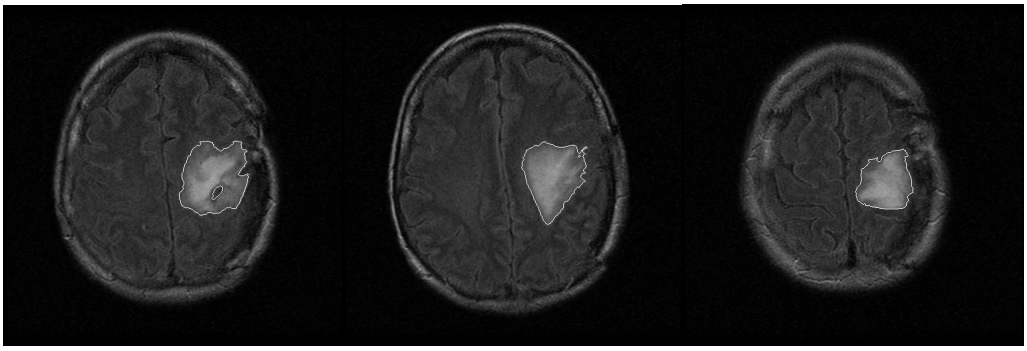


Figure 4.16 Examples of ACM effect on Patient 1 (FLAIR images)

Using TP, FP and FN as the evaluation criteria, compare the final tumor classification results after the contour refinement by ACM and the Ground Truth munally delineated by the doctors, the quantitative evaluation results of the corresponding parameters are shown in Table 4.9.

Table 4.9 Quantitative evaluation of contour refinement by ACM

Criteria	TP/%	FP/%	FN/%	Total Error/%
Without ACM	95.0	5.9	5.0	10.9
With ACM	99.1	7.2	0.9	8.1

It can be seen from Table 4.9 that, using precise tumor region from the initial segmentation as the initial curve of ACM can increase the segmentation accuracy of the tumor boundary contour and reduce FN to a high extent; on the contrary, if the energy function is badly designed and controlled, there are too drastic changes in the boundary

curves of the tumor and FP will increase to a certain extent; and ACM algorithm is not suitable for the aliasing border or boundary contour area with a gentle gradient. Meanwhile, ACM algorithm is influenced by the initial valulargе to large degree and the corresponding complexity and computational burden is much higher than the region growing-based contour improvement algorithm. The increase of computing time will greatly decrease the convergence speed, resulting in a decline in efficiency.

4.7 Effectiveness of Adaptive Training

Adaptive training is mainly used to automatatically follow up the disease progresses and changes in the tumor region of the patients. Through the evaluation on changes in tumor volume between different examination periods of patients, the system can provide a diagnostic advice for the improvement and completement of the medical treatment in time.

Through the training in the first examination, the data characteristics of the patient's tumor have been obtained by the trained classifier. Adaptive training is used to improve and update necessarily the obtained classifier available on last or even earlier stage based on the new MRI data to be completely apply the data changes. Through adaptive training, the disease of the patient is fully capable of tracking automatically, which greatly reduces the workload of doctors, subjective factors in processing, and the difficulties in manual processing.

In accordance with the treatment of adaptive training algorithm described in Section 3.4 of Chapter III, the conditions of Patient 1~Patient 3 who contain multi-period MRI examination data are followed up. The effectiveness of adaptive training method is verified through the comparison of the use of adaptive training on the data from a new examination and the direct classification by the model of last examination without the adaptive training.

Figure 4.17 lists the examples of classification results of Patient 1~Patient 3 with and without adaptive training on the new examination data respectively.

In Figure 4.17, each line corresponds to Patient 1 to Patient 3 respectively. The first colume is the original input images, the second colume is the classification results by the model for last examination, and the third colume is the classification results by the new obtained after the adaptive training on the new iuput data in this examination period.

As shown in Figure 4.17, after the adaptive training, the data characteristics of brain tumors in the new examination period are effectively obtained by the classifier, which leads to a greatly improved tumor classification accuracy. For the case of Patient 1 with a relatively simple tumor, although relatively accurate results can also be obtained without the adaptive training, there are a lot of glitches near the edge and the smoothness decreases greatly. For two more complex tumor cases as Ptient 2 and Patient 3, it is not completely capable to handle them without the adaptive training. In

the case of Patient 2, cerebral spinal fluid is wrong segmented into the tumor, which means the two tissues with a clear distinction between the properties are merged into the same category. Patient 3 has complex contours of the tumor, the same as the new data in the new examination. Using traditional methods to achieve accurate segmentation is difficult, and using the training model in last examination can also not obtain the optimized performance to generate a lot of piecemeal partition areas with obvious errors. The adaptive training can well solve all the above difficulties: the boundary contours of the tumor are smoothed; different tissues with different characteristics are also clearly distinguished from the tumor region to ensure that the region does not contain other normal tissues; complex contours of the tumor can also be clearly and accurately dilineated.

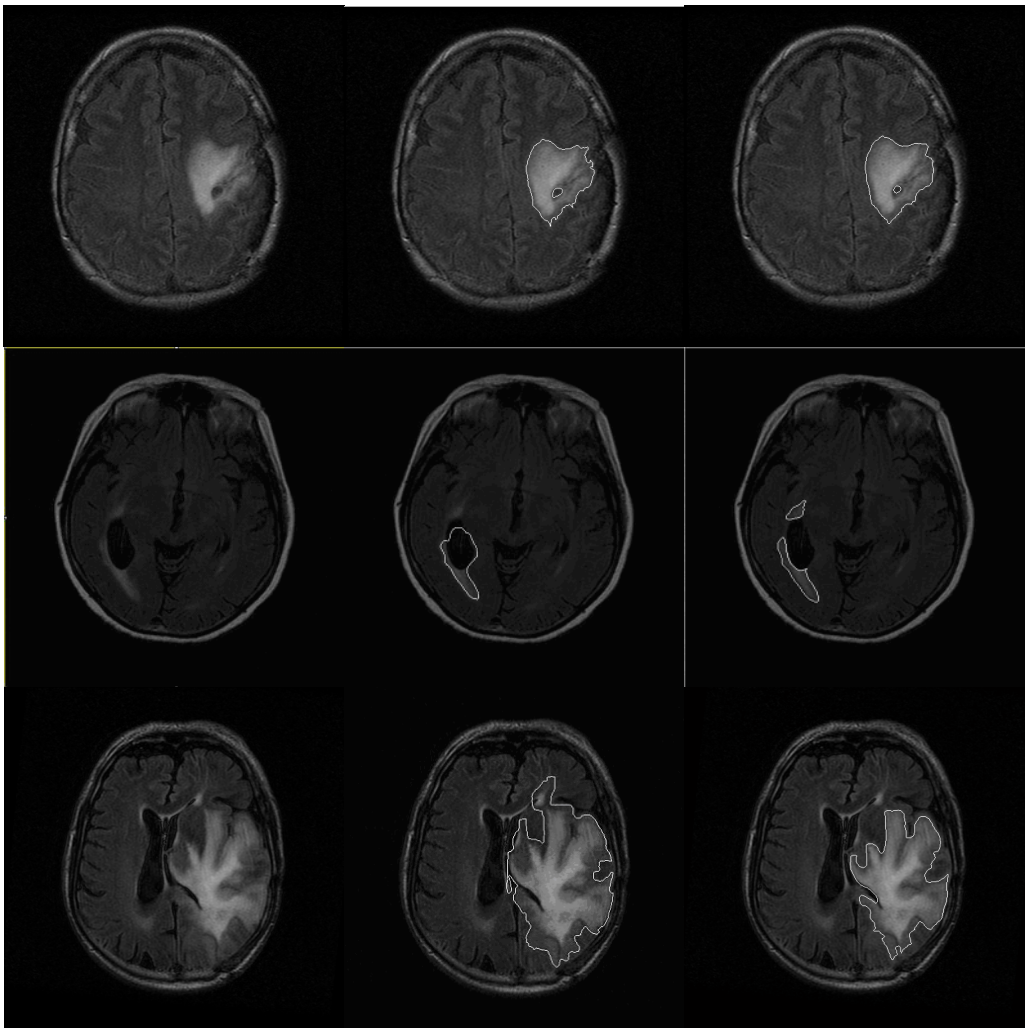


Figure 4.17 Examples of adaptive training for Patient 1 to Patient 3 (FLAIR images)

Comparing all the experimental data of Patient 1 to Patient 3 with the Ground Truth manually dilineated by a team of doctors, the evaluation criteria is selected as TP, FP and FN. The average of all data can be obtained to quantitatively evaluate the

effectiveness of the adaptive training and results are shown in Table 4.10.

By the tracking of patient's tumor based on the adaptive training, the system can automatically separate the tumor area and get the entire tumor volume. Table 4.11 shows the percentages of tumor volume changes of Patient 1 to Patient 3 in the adjacent examination period.

Table 4.10 Effectiveness of adaptive training

Criteria	TP/%	FP/%	FN/%	Total Error/%
Without Adaptive Training	88.1	25.2	11.9	37.1
With Adaptive Training	98.9	4.5	1.1	5.6

Table 4.11 percentages of tumor volume changes of Patient 1 to Patient 3 in multi-examination

Period	$P_1 \sim P_2$ /%	$P_2 \sim P_3$ /%	$P_3 \sim P_4$	$P_4 \sim P_5$ /%
Patient 1	-13.445	-6.417	-3.267	-6.473
Patient 2	17.763	-2.912	9.537	——
Patient 3	-7.831	-6.596	2.335	-3.762

Where P_n refers to the $n-th$ examination period; the negative percentages denote a downward trend of tumor volume; "——" means no data in this examination period. A new MRI examination is usually conducted for the patients at a interval of 4 months, thus all the patients' conditions have been supervised for more than one year.

Figure 4.18 corresponds to Figure 4.1, which lists the examples of the segmentation results automatically tracked by the proposed system in five examination periods. It can be intuitionally observed that the size of the tumor shows a decreasing trend.

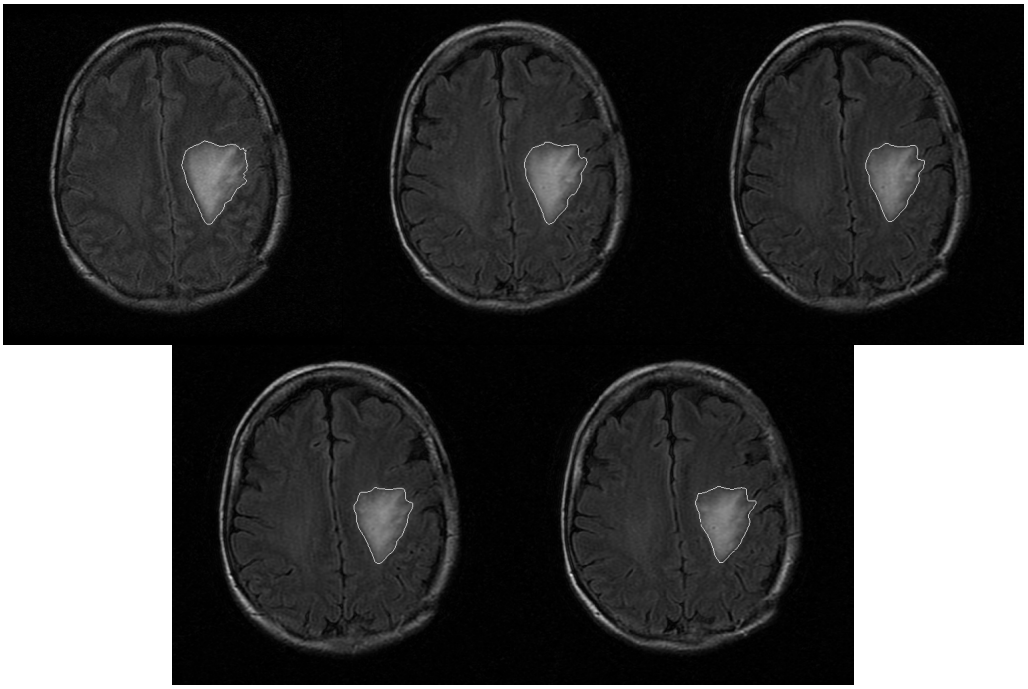


Figure 4.18 Examples of tracking results of Patient 1 in five examination periods (FLAIR images)

Comparing the tumor volume changes in the two adjacent examination periods can describe the patient's condition changes to some extent. Combined with clinical pathology analysis, the doctors can evaluate the the validity and accuracy of the early diagnosis and treatment based on all the experimental data. Then according to their medical knowledge and clinical experience, they will make a new grasp on the development of the disease, conduct the necessary changes and improvement to the treatment plan in next examination and provide more scientific and reasonable treatment recommendations.

For exmmple, the tumor of Patient 1 shows a significantly decreasing trend after the treatment for nearly 2 years. Combined with medical pathology analysis, it proves that the doctor's treatment for this patient is very correct and the condition of the patient can be effectively controlled. The following treatment can be improved based on the therapeutic treatment of the previous period. Although the tumor volume of Patient 2 in the second period decreases slightly, the volume in other periods shows a rapid increasing trend, which means that the doctor is unable to control the rapid development of the disease and the patient's condition has a state of rapid deterioration. Therefore, the doctor can use these reference data to make necessary changes in the next treatment plan. And this is the significance of the proposed system in this paper.

4.8 Comparison with Other Methods

In this paper, the method from tumor classification to tissue classification is proposed, and we hope high accuracy of tumor segmentation can lead to high accuracy of tissue classification. In order to verify the feasibility of this idea, in this section some successful algorithms to get the final tumor regions will be compared and discussed in performance with the proposed method. The comparison is conducted under the same hardware environment and the same software environment with the same input data sequence, in order to ensure that the results are comparable.

4.8.1 Traditional SVM

The traditional SVM method [135] in the comparison directly uses the traditional SVM theory without any feature selection operation. Input data are the fusion of feature sub-vectors directly extracted from the corresponding sample points of T2-weighted images, PD-weighted images and FLAIR images. The dimension of feature vector matrix retains all of the feature dimension. The rest of the operation process is exactly the same as the proposed method.

It has been pointed out in section 4.3 that the feature vector matrix obtained in this case will obviously contain a large amount of interference and noise, not only to increase the computing time considerably, but also to reduce the accuracy. The parameters selection process in traditional SVM will also limit the performance of the entire system. If the parameters is not selected well, it will greatly affect the system performance.

In the experiment, the cross-validation method is used to obtain relatively better classification parameters. The examples of the segmentation results are shown in Figure 4.19.

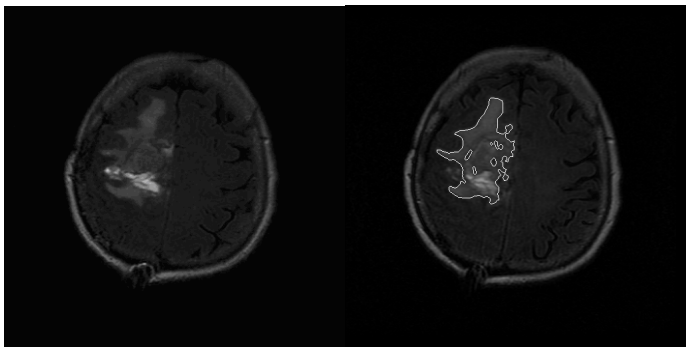


Figure 4.19 Examples of traditional SVM (FLAIR images)

Compared the the Ground Truth, the evaluated performance of the algorithm is shown in Table 4.12.

Table 4.12 Performance evaluation of traditional SVM

Criteria	TP/%	FP/%	FN/%	Total Error/%	Time/Min
With contour refinement	87.5	9.3	12.5	21.8	23
Without contour refinement	85.3	10.7	14.7	25.4	23

It can be observed from the data in Table 4.12 that, the accuracy of the final results in the traditional SVM indeed influenced because of the absence of feature selection and the computing time greatly increases.

4.8.2 Fuzzy C-means

Fuzzy C-means clustering is non-supervised classification algorithm [31]. It combines the traditional C-means clustering algorithm with the fuzzy theory to achieve the classification by dynamic clustering method. In the classification process, all the samples among different categories are constantly revised, until the classification meets the stopping conditions to converge automatically.

In the traditional C-means clustering, through the constant iterations the center of each category is constantly updated and the data to be classified will directly determined to some category. FCM also need to pre-set the number and the initial centers of categories. Different from the hard decision in the traditional C-means clustering, FCM does not directly determine the category the data belongs to, but gives its membership functions relative to the center of each category for a unspecified sample, that is, to indicate the degree of membership to each category for the data. Membership function is introduced to the updating process of the class centers in the traditional C means clustering, and the centers and membership functions are iteratively updated at the same time, which is the iterative update rule of FCM.

In this paper, FCM approach is applied to all the patients' MRI image data. Different from the SVM algorithm, FCM only need to use the basic intensity features and texture features to achieve the pre-set classification. As to the input of multiple sequence data, the results of each sequence can be determined according to the "majority decision" criterion pixel by pixel, to obtain the corresponding final results.

On the other hand, the feature vectors extracted from the fusion of the multiple input sequences can also be used as the data in high-dimensional space in the FCM treatment. However, data is directly classified by FCM without the learning and training

process, therefore, the classification results are worse in general compared with those from supervised learning classification algorithms; and the classification is mainly relied on the intensities, thus the FCM algorithm can not be carried out to effectively achieve more accurate segmentation for data of multiple sources, quite different characteristics and the physical meanings. For these reasons, we use the result fusion method in FCM to deal with the multiple data sequences, that is to say, to cluster each image of each input sequence first based on the intensities, and then the results of each sequence are fused toget the obtain the final classification results of tumors .

Considering the images in Figure 4.2 as examples, use FCM algorithm to deal with the three sequential images. It is obvious that in the actual processing using only two categories can not achieve high precision, so the initial categories are selected as four classes (adding cystic degeneration and edema near the tumor border as categories). After the finish of the clustering, make the appropriate merger of the categories. As edema and cystic are also abnormal tissues, the accuracy of FCM can be effectively improved by the process of first division and the merging on the classification categories. Examples of classification results obtained are shown in Figure 4.20. In order to contrast, data of Patient 2 with the complex boundary contour are also listed, as shown in Figure 4.20.

It is easy to see from Figure 4.20 that, FCM algorithm is quite suitable for the data with obviously different gray values, such as the FLAIR images, to achieve a relatively accurate segmentation; but for the tissues of the images with very similar intensities, such as the PD-weighted sequence, FCM can nearly not conduct the classification. In the easily-confused areas with similar intensities, it is difficult to determine the initial number and the centers of classes, which will seriously affect the clustering results. Therefore, the mechanisms of a majority decision is designed in this paper, that is, for a particular pixel, only when the three sequences all identify it as tumor points, then determine it as the tumor tissues, which can significantly indicate the role of the FLAIR images in the clustering; otherwise T2-weighted images and PD-weighted images will generate a large clustering area because of the similar intensities, leading to obvious mistakes when using the majority decision mechanism.

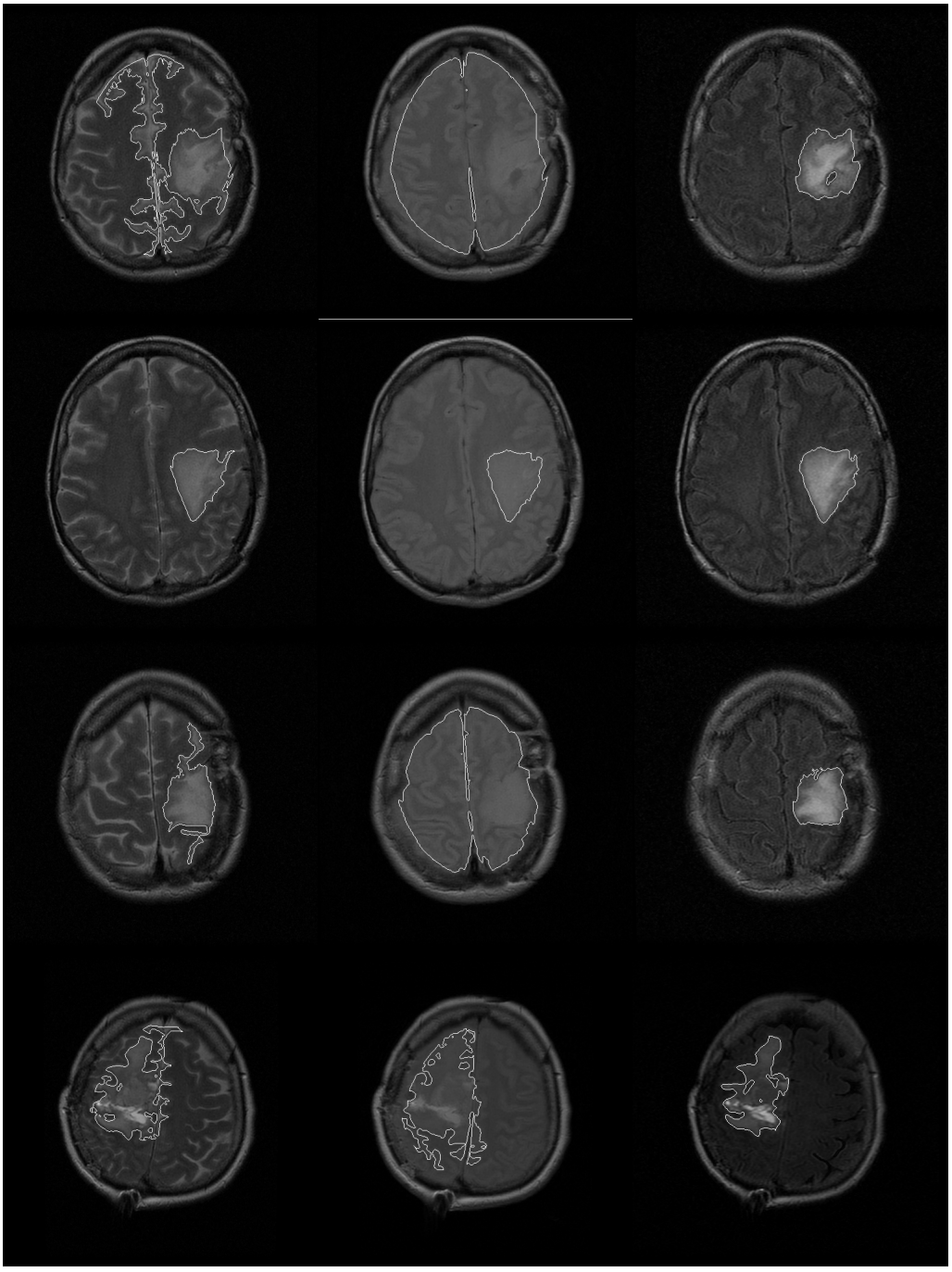


Figure 4.20 Examples of FCM classification results of the three sequences
（the first three rows correspond to data of Patient 1 shown in Figure 4.2, the fourth row corresponds to data of Patient 2. The columes are T2-weighted, PD-weighted and FLAIR images respectively.）

The results are compared with the Ground Truth under the evaluation system of TP, FP and FN. The average values are shown in Table 4.13.

Table 4.13 Performance evaluation of FCM

Criteria	TP/%	FP/%	FN/%	Total Error/%	Time/Min
With Contour Refinement	80.2	17.7	19.8	37.5	3
Without Contour Refinement	75.8	21.9	24.2	46.1	3

From Table 4.13, it can be observed that FCM treatment does not require prior knowledge and training process and does not extract complex features. The algorithm is fast and simple to operate. But its accuracy is limited for the gliomas with overlapping boundaries and complex properties compared to the normal tissues in MRI images. If the tumor appeared to be separeted in the same slice or there are indeed multiple tumors in the patients themselves at the same time, FCM can not achieve the correct clustering. Although the contour refinement can improve the accuracy to some extent and reduce the occurrence of boundary errors in the classification, but FCM can still not reach a high accuracy.

4.8.3 Neural Networks

Tneural network is a widely used pattern classification method [136], also in medical image processing [137]. Generally speaking, the complexity of neural network is proportional to the complexity of the problem. The more the number of neurons is, the number of hidden layers is, the higher accuracy the classification algorithm can be and the more correctly the algorithm converges. However, the corresponding computational complexity and time-consuming of the process will dramatically increase, while the algorithm is not easy to convergence, which leads to a manually set stopping condition or falling into the local optimum.

The neural network used in our experiments is based on the BP algorithm to complete selection a two-category classification. The initial values of weights are randomly generated by the smaller data; the neural network contains three neurons and three layers network using back propagation algorithm. The propagation errors are calculated from the output layer backwardthe front layer to revise and update the weights of neuron in different layers and obtain the trained classification model.

Similar to SVM classifiers, the neural network also requires that the input sample data have the same format with the data to be classified. That is to say, training data and test data needs to have the same composition. From this point of view, neural networks, like support vector machines, can be conveniently extended to other applications from the essential image processing of pixels, such as the field of image retrieval with the whole images as input (the same as matrix elements) and so on.

Also the same as SVM, the feature vectors matrix extracted from the selected sample points is selected as the training data. KCS criterion is also used for the feature selection to train and obtain the classification model. Test data must also be made of the same feature extraction method to obtain a feature vector composed of the same feature elements. Re-use classifier to make a classification pixel by pixel. After the contour refinement , the final classification is completed.

The data of all patients are tested using the above method and the examples are shown in Figure 4.21.

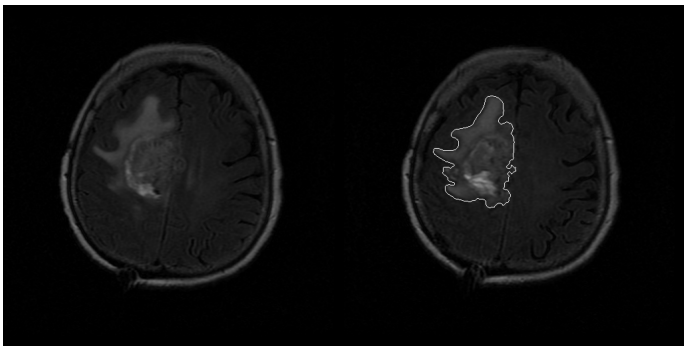


Figure 4.21 Classification results of Patient 2 by neural networks (FLAIR images)

Compared with the Ground Truth and using the same evaluation criteria, the results are shown in Table 4.14.

Table 4.14 Performance evaluation of neural networks

Criteria	TP/%	FP/%	FN/%	Total Error/%	Time/min
With Contour Refinement	96.0	5.1	4.0	9.1	8
Without Contour Refinement	91.8	4.9	8.2	13.1	8

From Figure 4.21 and Table 4.14 it can be seen that, if the numbers of the hidden layers and neurons are well set, the neural networks can also achieve optimum classification results by iteratively updating the weights. Meanwhile, the KCS feature selection algorithm and region growing-based contour refinement algorithms are also be effective for neural networks. One disadvantage of neural network is its obvious increase of time-consumption in the training rather than SVM, which leads to a slight lack of efficiency; the other is that the convergence is generally need to be controlled and the obtained results are not the global optimal solution.

4.8.4 Improved Level Set

Level set algorithm is the classification algorithm based on the geometric

deformable model [47-50]. Assuming that the tumor image to be segmented corresponds to a section of a high-dimensional surface, the spatial analytic functions corresponding to the high-dimensional surface are generally the partial differential equations, to which the curve of the implicit solution is the corresponding boundaries of the tumor when the equations are equal to 0. The solution set is also called the zero level set. Level set method transforms the evolution of the curve into the surface evolution and the surface deformation is conducted based on the theory of surface evolution. The surface evolution is controlled by the information in the images to be segmented, of which the speed function determines the evolution speed of each point on the curve.

Different from active contour model, the surface equation given by level set method is not a parametric equation dependent on some variables, but expressed as the implicit function equation. Therefore, the algorithm is not sensitive to the initial values of the curve and especially suitable for the evolution and iteration of the topology structures. Also based on this difference with ACM, level set method can be better used for the segmentation of medical images with the tumor, not just for the contour improvements after obtaining the initial segmentation results.

The performance limitations of level set method are mainly the selection of the speed control function. The evolution of surfaces need to optimize the corresponding function, which is quite time-consuming. Thus, the experiments are just carried out on only 2 patients with complex properties (Patient 2 and Patient 3). The examples of the obtained tumor segmentation results are shown in Figure 4.22, and the evaluation results compared with the Ground truth are shown in Table 4.15.

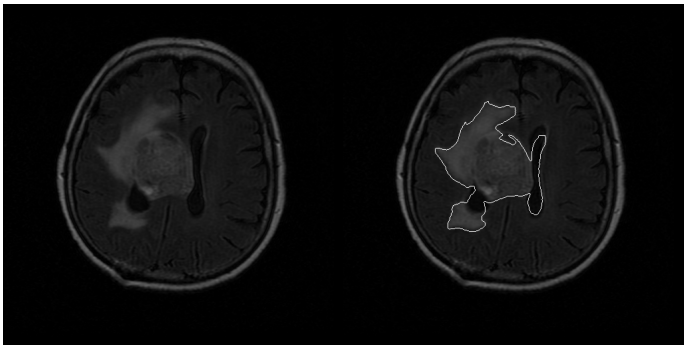


Figure 4.22 Examples of classification results on Patient 2 by level set method (FLAIR images)

Table 4.15 Performance evaluation of level set method

Criteria	TP/%	FP/%	FN/%	Total Error/%	Time/min
With Contour Refinement	96.7	7.4	3.3	10.7	38
Without Contour Refinement	92.6	5.3	7.4	12.7	38

From Figure 4.22 and Table 4.15, it can be observed that, level set algorithm is more applicable to the situation with only one curve need to evolute. If the tumor in the image is divided into several parts, the level set algorithm will connect them together although the topology structure changes to form a complete area. For the segmentation of a single tumor, level set algorithm can achieve high precision, but it is quite time-consuming, which is not only relative to the selection of the speed control function, but also due to the reason that the core of the algorithm is to evalute the surface corresponding to the entire image. Therefore, the calculation amount is very large and the efficiency is relatively low.

4.9 Tissure Classification

Tissue classification is a multi-class decision problem. A novel and successful framework for the solution to this problem is the optimal decision rule with class-selective rejection and performance constraints. It contains three kinds of criteria: the first is the average expected loss, which is to cost to be minimized; the second is the label sets, which is the decision options, determined by the admissible assignment classes or subsets of classes; and the last is the constrains, which define the performance the classifier can achieve [138].

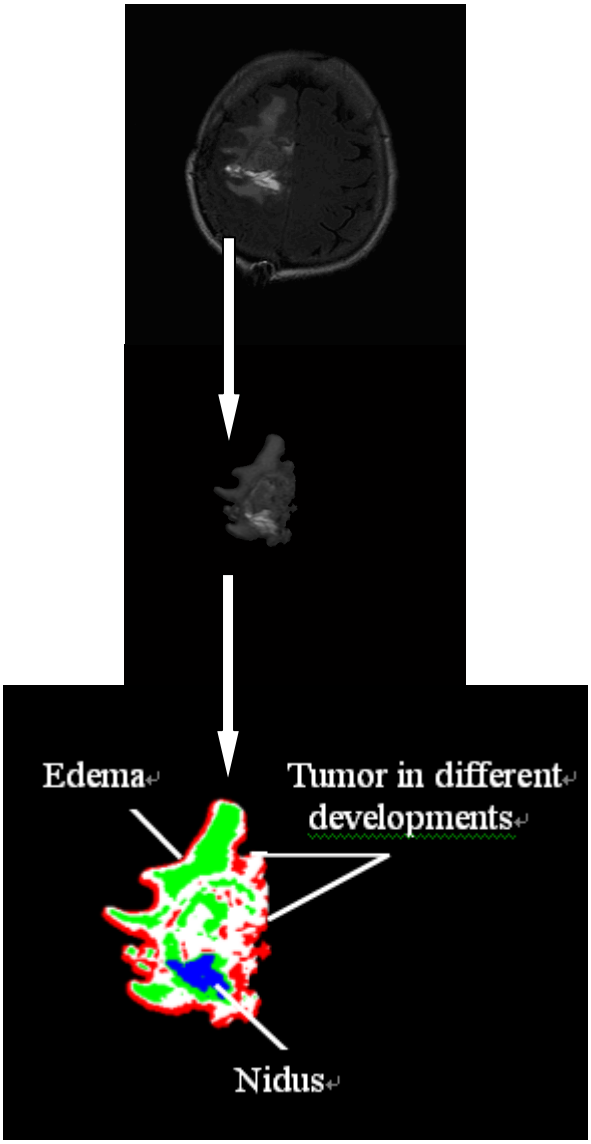


Figure 4.23 Tissue classification in Patient 2 (FLAIR images)

In our problem, based on the segmentation above and considering the initial separated tumor region as the image to be segmented, the abnormal tissues in the tumor region are further classified by the clustering methods. Although it can not reach the same high accuracy as the algorithm in [138], as a sub-problem in tumor segmentation, its efficiency and simplicity can be acceptable.

The procedure of clustering method in tissue classification is as follows:

First determine the number of sample categories to be classified, and then use the unsupervised FCM clustering algorithm as mentioned before. The examples of the results are shown in Figure 4.23.

In Figure 4.23, the top image is the original input image; the middle image is the tumor segmentation results; the bottom image is the amplified tissue classification results. The tumor region is divided into 3 different parts by the proposed algorithm.

The area with the highest intensities in the middle is the nidus of the tumor, that is the source of tumor, thus its state of disease is the most serious region with the highest stage of development. And the red part outside near the contour is the cystic and edema, both of which are often around the tumor, although they do not contain any tumor cells. As the development of the disease and the continuous erosion and expansion of the tumor, cystic degeneration and edema may deteriorate rapidly as the tumor. Other green and white parts of the image are different tumor lesions at the development stages.

The results in Figure 4.23 can assist the doctors for the evaluation of patients' conditions (tumor development periods). After the surgical operations, the doctors can get some cell samples with the probes in the area of the brain given by the computer and analyze the amount and proportion of cancer cells in the sample to determine whether the operation has completely removed the tumor and the following treatment program.

4.10 Summary

This chapter outlines in detail the data used in this experiment, the experimental results and related quantitative comparison of the algorithm. The key operations involved in the data processing is analyzed, and the advantages and disadvantages of the algorithm as well as possible problems are described by the comparison with other algorithms.

Through the actual comparison, it can be seen that the proposed algorithm has high precision and efficiency, and it can achieve the tracking of the patients' conditions at the same time. The proposed algorithm has strong reliability and applicability.

CHAPTER 5 CONCLUSION

5.1 Conclusion of Contributions

Magnetic Resonance Imaging (MRI) is an important examination and diagnosis method for brain tumors in medical imaging. With a sound mechanism and clear imaging of soft tissues, the doctor on the patient's diagnosis can be scientific and rational, to grasp the exact progression of the disease state, which would set out the appropriate treatment, surgery and following-up to a series of disease control measures. Computer-aided analysis is to reduce the workload of doctors, to improve the diagnostic accuracy of the para-medical analysis, and meanwhile to improve the automatic degree in practice.

This paper presents a semi-automatic and Support Vector Machine (SVM) based tumor detection system, which can deal with multiple input MRI sequences and track on the patient's condition in the whole therapeutic treatment. The system only requires interactive participation once the sample points in the analysis of data of the first examination are randomly selected, and all the processings in next examinations can be carried out automatically by the system itself.

The segmentation algorithm works in an ordered manner, from tumor to tissues. That is, a first classification consists of segmenting the area of the entire tumor, and then separating the abnormal tissues based on the obtained tumor area. The classification algorithm uses multi-kernel SVM as its classifier and fusion of T2 weighted, PD-weighted and FLAIR images as its input data to extract the corresponding feature vector matrix. This matrix is dealt with by a feature selection method to choose the most important features using some criterion. Then the obtained matrix is utilized to train the optimal classifier and accomplish the initial tumor classification.

In addition to the SVM-based classification subsystem, the system proposed in this thesis also contains other two important components: the region growing-based contour refinement subsystem and the adaptive training-based following-up subsystem. The former aims to improve the border quality of the initial tumor segmentation, to reduce incorrect and missing points in the contour and to improve the classification accuracy; the latter is to grasp the characteristics of tumors in the patients, then to automatically segment and track tumor changes in a series of examiantions and finally to estimate the volume variation between successive treatments to assist the doctors in assessing the effects of treatment with clinical pathology analysis and improving the therapeutic treatment in the next pathological periods.

The proposed system is tested on 13 patients with 24 examinations, 72 image sequences and totally 1728 MRI images. Compared with some traditional methods, the proposed method reaches the average accuracy rate of 98.9%. In the experiments, the gray values, texture features and other mathematical transformation-based derivative features, such as DCT transformation coefficients, wavelet transformation coefficients,

fractal features and so on, are also applied and compared in the system.

Integrated with the fusion strategies, the system achieves a multi-kernel SVM to realize the data-level fusion and decision-level fusion respectively. Meanwhile, the system also puts forward the method to integrate the feature selection and the SVM classifier design, and some relatively novel feature selection methods are applied to the system. In practical experimentations, the feature selection methods above are tested to verify the validities and their applicabilities to the SVM classifier. Compared with some traditional feature selection methods, the system can be more effective to simplify the process of parameter selection in the feature selection in SVM design, which can integrate the two parameter selection processes together to reduce the complexity and improve the robustness and adaptability of the system.

The innovations of this thesis are as follows:

1. we propose a segmentation and estimation system which allows classifying the tumor region first and then abnormal tissues. The accuracy in tumor segmentation maintains the accuracy in tissue classification.
2. the system tries to apply some common and derivative features, and the effectiveness of these features are quantitatively analyzed and compared.
3. the system achieves some feature selection and verifies the validity and applicability of the latter to the SVM classifier. A method, with which the feature selection criteria are integrated with the design of the SVM classifier, is proposed to overcome the difficulty of parameter selection in the design and improve the applicability of the system.
4. combined with feature selection-based data fusion and decision fusion strategies, our system constructs the multi-kernel SVM to accomplish the multi-class classification of different tissues, the accuracy of which is influenced by the accuracy of tumor segmentation.
5. We propose simple and effective algorithms about contour refinement and tumor tracking to improve the accuracy, robustness and applicability of the proposed system.

5.2 Future Work

Since SVM is very suitable for high dimensional data, especially for processing the data from different inputs, the proposed method in this thesis would allow a better analysis of three-dimensional medical images corresponding to multiple sequences. With the proposed method, new MRI imaging sequences can also be readily added as input data. Futhermore, the developped algorithm can be easily extended to the analysis of other medical images, such as CT, PET images, and so on.

As present there is not a public test database of high capacity of MRI sequences internationally, and common evaluation criteria are still relatively simple, so in the future work, we will attempt to contact other laboratories carrying out similar research

to share MRI data and work together in order to establish a large-capacity, open test database, and continue to test the proposed method with the aid of the new database to validate its effectiveness and further improvement.

Combined with the homologous MRS spectral analysis, our research will be expanded to the fusion of two-dimensional images and one-dimensional signal, in which the accuracy of the algorithm can be further improved through the two types of signals matching each other.

Finally we hope that in the entire study, existing computer-aided analysis and evaluation systems on medical images can be improved and new evaluation criteria can be proposed and applied in the future works.

Publications de l'auteur

[1]Nan Zhang, Su Ruan, Stéphane Lebonvallet, Qingmin Liao and Yuemin Zhu. Kernel Feature Selection to Fuse Multi-spectral MRI Images for Brain Tumor Segmentation. Computer Vision and Image Understanding, 2011, 115(2):256-269.

[2]张楠, 廖庆敏, Su Ruan, Stéphane Lebonvallet, Yuemin Zhu. 一种新的基于 3D 信息的脑肿瘤分割和评估系统. 清华大学学报: 自然科学版, 2011, 51(1): 96-100.

Nan Zhang, Qingmin Liao, Su Ruan, Stéphane Lebonvallet and Yuemin Zhu. A Novel Tumor Segmentation and Evaluation System based on 3D Information. Tsinghua Science and Technology, 2011, 51(1): 96-100.

[3]Nan Zhang, Su Ruan, Stéphane Lebonvallet, Qingmin Liao and Yuemin Zhu. Multi-kernel SVM based Classification for Brain Tumor Segmentation of MRI Multi-Sequence. IEEE International Conference on Image Processing, Cairo, Egypt: IEEE, 2009: 3373-3376.

[4]Nan Zhang, Qingmin Liao, Su Ruan, Stéphane Lebonvallet and Yuemin Zhu. Multi-kernel SVM based Classification for Tumor Segmentation from Fusion of MRI Images. IEEE International Workshop on Imaging Systems and Techniques. Shenzhen, China: IEEE, 2009: 71 -75.

[5]Nan Zhang, Su Ruan, Stéphane Lebonvallet, Qingmin Liao and Yuemin Zhu. SVM Based Follow-up System for Brain Tumor Evolution from Magnetic Resonance Images. Modeling and Control in Biomedical Systems, 2009, 7(1).

[6]Su Ruan, Nan Zhang, Stéphane Lebonvallet, Qingmin Liao and Yuemin Zhu. Fusion and Classification of Multi-Source Images by SVM with Selected Features in a Kernel Space. IEEE Workshops on Image Processing Theory, Tools and Applications. Paris: IEEE, 2010: 17-20.

[7]Su Ruan, Nan Zhang, Stéphane Lebonvallet, Qingmin Liao and Yuemin Zhu. Fusion et Classification d'images Multi-Sources par SVM avec Selection des Charactacristiques dans l'espace. GRETSI Symposium on Signal and Image Processing. Dijon, France: 2009.

[8]Su Ruan, Nan Zhang, Qingmin Liao and Yuemin Zhu. Image Fusion for Following up Brain Tumor Evaluation. IEEE International Symposium on Biomedical Imaging. Chicago: 2011. Accepted.

[9]张楠, 杨文明, 廖庆敏. MRI 图像脑肿瘤分割算法概述. 中国生物医学工程学会成立 30 周年纪念大会暨 2010 中国生物医学工程学会学术大会. 北京: 中国生物医学工程学会, 2010: 75-81.

Nan Zhang, Wenming Yang and Qingmin Liao. MRI Tumor Segmentation Algorithms: a Survey. Chinese Conference on Biomedical Engineering. Beijing, 2010: 75-81.

REFENENCE

[1] Y. Gao. One-third of Cancers Can Be Cured [N]. Wen Wei Po, 20110205.

[2] X. Xuan. MRI Brain Tumor Segmentation based on Statistical Analysis [D]. Master Thesis of Tsinghua University, 2007.

[3] C.Z. Wang. Neurosurgery [M]. Beijing: People's Medical Publishing House, 2002.

[4] S.K. Qin, J. Ma, W.C. Pan, et al. Advances in Clinical Oncology 2010 [M]. Beijing: People's Medical Publishing House, 2010.

[5] C.X. Ji and C.M. Yu. Oncology [M]. Beijing: Science Press, 2008.

[6] J. Tian, S.L. Bao and M.Q. Zhou. Medical Image Processing and Analysis [M]. Beijing: Electronic Industry Press, 2003. 35-77.

[7] E.H. Wu, G.S. Feng, R.J. Bai, et al. Medical Imaging (the 6th Edition) [M]. Beijing: People's Medical Publishing House, 2008.

[8] http://www.chinabaike.com/2011/0121/201166.html [J/OL].

[9] L.Y. Huang. The Basic Principles of Medical Imaging [M]. Beijing: Electronic Industry Press, 2009.

[10] Y. Zhan and G.X. Tan. Medical Imaging Diagnostic Lab Manual [M]. Wuhan: Hubei Science and Technology Press, 2009.

[11] Z.H. Yang, F. Feng, X.Y. Wang, et al. MRI Guide-Check Specification, Clinical Strategies and Application of New Technologies [M]. Beijing: People's Medical Publishing House, 2010.

[12] J.Y. Huang and X.Y. Liang. Principles of Magnetic Resonance Imaging [M]. Shaanxi: Shaanxi Science and Technology Press, 1998.

[13] http://jinshungroup.com/ProductShow.asp?ID=92 [J/OL].

[14] http://euroscientist.com/ [J/OL].

[15] M. Sezgin, B. Sankur**.** Survey over Image Thresholding Techniques and Quantitative Performance Evaluation [J]. Journal of Electronic Imaging, 2004, 13 (1): 146–165.

[16] Y.J. Zhang. Image Processing and Analysis [M]. Beijing: Tsinghua University Press, 1999.

[17] N. Otsu. A Threshold Selection Method from Gray-Level Histograms [J]. IEEE Transactions on System, Man and Cybernetics, 1979, 9 (1): 62-66.

[18] R.C. Gonzalez. Digital Image Processing [M]. Peking: Publishing House of Electronics Industry, 2003.

[19] M.R. Kaus, S.K. Warfield, P.M. Black, et al. Automated Segmentation of MR Images of Brain Tumors [J]. Radiology, 2001, 218: 586-591.

[20] M. Prastawa, E. Bullitt, S. Ho, et al. A Brain Tumor Segmentation Work based on Outlier Detection [J]. Medical Image Analysis, 2004, 8: 275-283.

[21] M. Prastawa, E. Bullitt, N. Moon, et al. Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors [J]. Medical Image Computing, 2003, 10: 1341-1348.

[22] M.B. Cuadra, C. Pollo, A. Bardera, et al. Atlas-based Segmentation of Pathological MR Brain Images Using a Model of Lesion Growth [J]. IEEE Transactions on Medical Imaging, 2004, 10: 1301-1313.

[23] P.Y. Lau, F.C.T. Voon, and S. Ozawa. The Detection and Visualization of Brain Tumors on T2-weighted MRI Images Using Multiparameter Feature Blocks [C]. 27th IEEE Annual Conference of the Engineering in Medicine and Biology (EMBS), 2005: 5104-5107.

[24] R.A. Lerski, K. Straughan, L.R. Schad, et al. MR Image Texture Analysis: An Approach to Tissue Characterization [J]. Magnetic Resonance Imaging, 1993, 11: 873-887.

[25] A. Zizzari, U. Seiffert and B. Michaelis. Detection of Tumor in Digital Images of the Brain [C]. Poceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications (ICSPPRA), 2001: 132-137.

[26] R. Nezafat and H. Soltanian-Zadeh. Multiwavelet-based Feature Extraction for MRI Segmentation [C]. Processing of SPIE Wavelet Applications in Signal and Imaging Processing, 1998, 3458: 182-191.

[27] K.M. Ifterkharuddin, M.A. Islam, J. Shaik, et al. Automatic Brain Tumor Detection in MRI: Methodology and Statistical Validation [C]. Proceedings of Medical Imaging, 2005, 5747: 2012-2022.

[28] S. Herlidow-Meme, J.M. Constans, B. Carsin, et al. MRI Texture Analysis on Texture Test Objects, Normal Brain and Intracranial Tumors [J]. Magnetic Resonance Imaging, 2001, 21: 989-993.

[29] Z.Q. Bian and X.G. Zhang. Pattern Recognition [M]. Beijing: Tsinghua University Press, 2000.

[30] R.O. Duda, P.E. Hart and D.G. Stork. Pattern Classification (second edition) [M]. London: Springer, 2001.

[31] M.C. Clark, L.O. Hall, D.B. Goldgof, et al. Automatic Tumor Segmentation Using Knowledge-based Techniques [J]. IEEE Transactions on Medical Imaging, 1998, 17: 238-251.

[32] J.K. Udupa and S. Samarasekera. Fuzzy Connectedness and Object Definition: Theory, Algorithm, and Applications in Image Segmentation [J]. Graphical Model an Image Processing, 1995, 58: 246-261.

[33] J.G. Liu, J. K. Udupa, D. Odhner, et al. A System for Brain Tumor Volume Estimation via MR Imaging and Fuzzy Connectedness [J]. Computerized Medical Imaging and Graphics, 2005, 29: 21–34.

[34] J.K. Udupa, L. Wei, S. Samarasekera, et al. Multiple Sclerosis Lesion Quantification Using Fuzzy-Connectedness Principles [J]. IEEE Transactions on Medical Imaging, 1997, 16 (5): 598-609.

[35] W.M. Wells, W.E.L. Grimson, R. Kikinis, et al. Adaptive Segmentation of MRI Data [J]. IEEE Transactions on Medical Imaging, 1996, 15: 429-442.

[36] A.S. Capelle, O. Alata, C. Fernandez, et al. Unsupervised Segmentation for Automatic Detection of Brain Tumors in MRI [C]. Proceedings of International Conference on Image Processing (ICIP), 2000, 1: 613-616.

[37] D.T. Gering, W.E.L. Grimson and R.Kikinis. Recognizing Deviations from Normalcy for Brain Tumor Segmentation [C]. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2002, 2488: 388-395.

[38] S. Ruan, B. Moretti, J. Fadili, et al. Fuzzy Markovian Segmentation in Application of Magnetic Resonance Images [J]. Computer Vision and Image Understanding, 2002, 85: 54–69.

[39] W.B. Dou, S. Ruan, Q.M. Liao, et al. Fuzzy Information Fusion Scheme Used to Segment Brain Tumor from MR Images [J]. Fuzzy Logic and Applications, 2006, 2955: 208-215.

[40] W.B. Dou, S. Ruan, Q.M. Liao, et al. Fuzzy Information Fusion Scheme Used to Segment Brain Tumor from MR Images [C]. International Workshop on Fuzzy Logic and Application (WILF), 2003: 331-334.

[41] J.T. Jian. Deformable Model and Its Application in Medical Image Segmentation [D]. Doctor Thesis of Chinese University of Science and Technology, 2006.

[42] T. McInerney and D. Terzopoulos. Deformable Models in Medical Image Analysis: a Survey [J]. Medical Image Analysis, 1996, 1 (2): 91-98.

[43] M.Q. Wang, G.Q. Han and Q.Q. Tu. Deformable Model-based Image Segmentation: a Survey [J]. Medical Equipment, 2009, 2: 37-39.

[44] J. Luo, J.W. Zhang and D.S. Xia. FCM and the Deformable Model based on MRI Automatic Segmentation of Left Ventricle [J]. Computer Engineering and Applications, 2006, 9: 198-202.

[45] J.Y. Yan. Deformable Model for Image Segmentation and Its Low Contrast in the Application of Medical Ultrasound Images [D]. Doctor Thesis of Shanghai Jiaotong University, 2004.

[46] Y. Chen and T.G. Zhuang. Medical Image Analysis of Deformable Models [J]. Biomedical Engineering, 1999, 16(4): 497-501.

[47] D. Terzopoulos, J. Platt, A. Barr, et al. Elastically Deformal Models [J].

Computer Graphics, 1987, 21 (4): 205-214.

[48] X.L. Huang, Z. Qian, R. Huang, et al. Deformable-Models based Texture Object Segmentation [J]. 5th International Workshop of Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2005, 3757: 119-135.

[49] U. Meier, O. Lopez, C. Moncerrat, et al. Real-Time Deformable Models for Surgery Simulation: a Survey [J]. Computer Methods and Programs in Biomedicine, 2005, 77: 183-197.

[50] J. Feng and H.S. Horace. A Multi-Resolution Statistical Deformable Model (MISTO) for Soft Tissue Organ Reconstruction [J]. Pattern Recognition, 2009, 43 (7): 1543-1558.

[51] C.Y. Xu. Deformable Models with Application to Human Cerebral Cortex Reconstruction from Magnetic Resonance Images [D]. PhD Thesis, John Hopkins University, 2000.

[52] A.K.W. Law, F.K. Lam and F.H.Y. Chan. A Fast Deformable Region Model for Brain Tumor Boundary Extraction [C]. Proceedings of the Second Joint Engineering in Medicine and Biology Society / Biomedical Engineering Society (EMBS/BMES) Conference, 2002: 1055-1056.

[53] T.F. Chan, S. Esedoglu and M. Nikolova. Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models [J]. SIAM Journal on Applied Mathematics, 2006, 66 (5): 1632-1648.

[54] H.D. Mao, H.F. Liu and P.C. Shi. A Convex Neighbor-Constrained Active Contour Model for Image Segmentation [C]. IEEE 17th International Conference on Image Processing (ICIP), 2010: 793-396.

[55] X. Bresson, S. Esedoglu, O. Vandergheynst, et al. Fast Global Minimization of the Active Contour/Snake Model [J]. Journal of Mathematical Imaging and Vision, 2007, 28 (2): 151-167.

[56] X. Xuan and Q.M. Liao. Statistical structure analysis in MRI brain tumor segmentation [C]. IEEE 4th International Conference on Image and Graphics (ICIG), 2007: 421-426.

[57] D.N. Jin. Level Set Based Segmentation of Medical Images [D]. Master Thesis of First Military Medical University, 2005.

[58] Y. Gao. Level Set based Image Segmentation [D]. Master Thesis of East China Normal University, 2010.

[59] X.F. Wang. Level Set Method and Its Application in Image Segmentation [D]. Doctor Thesis of University of Science and Technology, 2009.

[60] F.P. Zhu, J. Tian, Y. Lin, et al. Level Set Method Based Medical Image Segmentation [J]. Software, 2002, 13 (9): 1866-1871.

[61] N. Parogios. A Level Set Approach for Shape-driven Segmentation and Tracking of the Left Ventricle [J]. IEEE Transactions on Medical Imaging, 2002, 22 (6): 773-776.

[62] T. Brox and J. Weickert. Level Set Based Image Segmentation with Multiple Regions [J]. Pattern recognition, 2004, 3175: 415-423.

[63] T.F. Chen. Medical Image Segmentation using Level Sets [R]. Technical Report. Canada, University of Waterloo, 2008.

[64] D.N. Metaxas, Z. Qian, X.L. Huang, et al. Hybrid deformable models for medical segmentation and registration [C]. IEEE 9th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2006: 1-6.

[65] K.H. Seo, W. Kim, C. Oh, et al. Face detection and facial feature extraction using color snake [C]. Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE), 2002: 457-462.

[66] H.K. Zhao, S. Osher and R. Fedkiw. Fast Surface Reconstruction Using the Level Set Methods [C]. Proceedings of IEEE Workshop on Variational and Level Set Methods in Computer Vision, 2001: 194-201.

[67] W.B. Dou, S. Ruan, Y.P. Chen, et al. A Framework of Fuzzy Information Fusion for the Segmentation of Brain Tumor Tissues on MR Images [J]. Image and Vision Computing, 2007, 25 (2): 164-171.

[68] S. Ruan, S. Lebonvallet, A. Merabet, et al. Tumor Segmentation from a Multi-Spectral MRI Images by Using Support Vector Machine Classification [C]. 4th IEEE International Symposium on Biomedical Imaging (ISBI), 2007: 1236-1239.

[69] J.L. Marroquin, B.C. Vemuri, S. Botello, et al. An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI [J]. IEEE Transactions on Medical Imaging, 2002, 21 (8): 934-945.

[70] J.Y. Yeh and J.C. Fu. A Hierarchical Genetic Algorithm for Segmentation of Multi-Spectral Human-Brain MRI [J]. Expert Systems with Applications, 2008, 34: 1285-1295.

[71] K.M. Iftekharuddin, J. Zheng, M.A. IsIam, et al. Fractal-based Brain Tumor Detection in Multimodal MRI", Applied Mathematics and Computation, 2009, 207 (1): 23-41.

[72] J.R. Jimenez-Alaniz, V. Medina-Bañuelos and O. Yáñez-Suárez. Data-Driven Brain MRI Segmentation Supported on Edge Confidence and a Priori Tissue Information [J]. IEEE Transactions on Medical Imaging, 2006, 25 (1): 74-83.

[73] H. Zhu. Data Fusion for Pattern Classification via the Dempster-Shafer Evidence Theory [C]. IEEE International Conference on Systems, Man and Cybernetics (ICSMC), 2002: 109-110.

[74] A. Taleb-Ahmed and L. Gautier. On Information Fusion to Improve Segmentation of MRI Sequences [J]. Information Fusion, 2002, 3 (2): 103-117.

[75] Q. Tao and R. Veldhuis. Threshold-optimized Decision Level Fusion and Its Application to Biometrics [J]. Pattern Recognition, 2008, 42 (5): 823-836.

[76] D. Maurer and P. Baker. Fusing Multimodal Biometrics with Quality Estimates

via a Bayesian Belief Network [J]. Pattern Recognition, 2008, 41 (5): 821-832.

[77] K. Nandakumar, Y. Chen, S. Dass, et al. Likelihood Rratio-based Biometric Score Fusion [J] IEEE Transactions on Pattern Analysis and Machine Intellegence, 2008, 30 (2): 342-347.

[78] K. Toh, J. Kim, and S. Lee. Biometric Scores Fusion based on Total Error Rate Minimization [J]. Pattern Recognition, 2008, 41 (5): 1066-1082.

[79] F.R. Bach, G.R.G. Lanckriet and M.I. Jordan. Multiple Kernel Learning, Conic Duality, and the SMO Algorithm [C]. Proceedings of the 21th International Conference on Machine Learning (ICML), 2004, 69.

[80] K. Popuri, D. Cobzas, M. Jagersand, et al. 3D Variational Brain Tumor Segmentation on a Clustered Feature Set [C]. Proceedings of SPIE, 2009, 7258: 72591N-72591N-10.

[81] D. Cobzas, N. Birkbeck, M. Schmidt, et al. 3D Variational Brain Tumor Segmentation using a High Dimensional Feature Set [C]. IEEE 11th International Conference on Digital Object Identifier (ICCV), 2008: 1-8.

[82] N. Moon, E. Bullitt, K.V. Leemput, et al. Automatic Brain and Tumor Segmentation [J]. Medical Image Computing and Computer-Assisted Intervention, 2002, 2488: 372-379.

[83] H. Khotanlou, O. Colliot, J. Atif, et al. 3D Brain Tumor Segmentation in MRI using Fuzzy Classification, Symmetry Analysis and Spatially Constrained Deformable Models [J]. Fuzzy Sets and Systems, 2009, 160 (10): 1457-1473.

[84] P.Y. Bondiau, G. Malandain, S. Chanalet, et al. Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context [J]. International Journal of Radiation Oncology Biological Physics, 2005, 61 (1): 289-298.

[85] T. Wang, I. Cheng and A. Basu. Fluid Vector Flow and Applications in Brain Tumor Segmentation [J]. IEEE Transactions on Biomedical Engineering, 2009, 56 (3): 781 – 789.

[86] D.T. Gering. Diagonalized Nearest Neighbor Pattern Matching for Brain Tumor Segmentation [C]. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2003, 2879: 670-677.

[87] D. Selvathi. Performance Evaluation of Kernel Based Techniques for Brain MRI Data Classification [C]. International Conference on Computational Intelligence and Multimedia Applications (ICCIMA), 2007: 456-460.

[88] V. Vapnik. The Nature of Statistical Learning Theory [M]. London: Springer, 1995.

[89] J. Zhou, K.L. Chan, V.F.H. Chong, et al. Extraction of Brain Tumor from MR Images Using One-Class Support Vector Machine [C.] 27th Annual International Conference on Engineering Medicine and Biology Society (EMBS), 2005: 6411-6414.

[90] Z. Wang, S.C. Chen and T.K. Sun. MultiK-MHKS: A Novel Multiple Kernel

Learning Algorithm [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008 (2), 30: 348-353.

[91] A. Rakotomamonjy, F. Bach, S. Canu, et al. More Efficiency in Multiple Kernel Learning [C]. Proceedings of 24th International Conference on Machine learning (ICML), 2007: 775-782.

[92] F. Bach. Exploring Large Feature Spaces with Hierarchical Multiple Kernel Learning [C]. Advances in Neural Information Processing Systems (NIPS), 2008: 105-112.

[93] X. Chen, T. Fang, H. Huo, et al. Graph-based Feature Selection for Object-Oriented Classification in VHR Airborne Imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49 (1): 353-365.

[94] Y.J. Li, D.F. Hsu and S.M. Chung. Combining Multiple Feature Selection Methods for Text Categorization by Using Rank-Score Characteristics [C]. 21st International Conference on Tools with Artificial Intelligence (ICTAI). 2009: 508-517.

[95] Q.W. Wang, B.Y. Li and J.L. Hu. Feature Selection for Human Resource Selection based on Affinity Propagation and SVM Sensitivity Analysis [C]. World Congress on Nature & Biologically Inspired Computing (NaBIC), 2009: 31-36.

[96] K.C. Khor, C.Y. Ting and S.P. Amnuaisuk. Forming an Optimal Feature Set for Classifying Network Intrusions Involving Multiple Feature Selection Methods [C]. 2010 International Conference on Information Retrieval & Knowledge Management (CAMP), 2010: 179-183.

[97] H. Nguyen, K. Franke and S. Petrovic. Improving Effectiveness of Intrusion Detection by Correlation Feature Selection [C]. International Conference on Availability, Reliability, and Security (ARES), 2010: 17-24.

[98] G.R. Li, X.H. Hu, X.J. Shen, et al. A Novel Unsupervised Feature Selection Method for Bioinformatics Data Sets through Feature Clustering [C]. IEEE International Conference on Granular Computing (GrC), 2008: 41-47.

[99] Y.L. Li and L. Song. Threshold Determining Method for Feature Selection [C]. Second International Symposium on Electronic Commerce and Security (ISECS), 2009: 273-277.

[100] H.L. Bu, S.Z. Zheng and J. Xia. Genetic Algorithm based Semi-Feature Selection Method [C]. International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing (IJCBS), 2009: 521-524.

[101] P.Y. Xia, X.Q. Ding and B.N. Jiang. A GA-based Feature Selection and Ensemble Learning for High-Dimensional Datasets [C]. International Conference on Machine Learning and Cybernetics (ICMLC), 2010: 7-12.

[102] E.H. Zhong, S.H. Xie, W. Fan, et al. Graph-based Iterative Hybrid Feature Selection [C]. 8th IEEE International Conference on Data Mining (ICDM), 2008: 1133-1138.

[103] Z.L. Xu, I.K. King, M.R. Lyu, et al. Discriminative Semi-Supervised Feature Selection via Manifold Regularization [J]. IEEE Transactions on Neural networks, 2010,

21(7): 1033-1047.

[104] T. Jolliffe. Principal Component Analysis [M]. New York: Springer, 1986: 151-156.

[105] B. Scholkopf, E. Smola, K.R. Muller, et al. Nonlinear Component Analysis as a Kernel Eigenvalue Problem [J]. Neural Computation, 1998, 10 (5): 1299-1319.

[106] C.J. Liu. Gabor-based Kernel PCA with Fractional Power Polynomial Models for Face Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26 (5): 572-581.

[107] M.S. Kim, I.H. Yang and H.J. Yu. Robust Speaker Identification Using Greedy Kernel PCA [C]. 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2008, 2: 143-146.

[108] T. Ogawa and M. Haseyama. Kernel PCA-based Resolution Enhancement Approach of Still Images using Different Levels of Pyramid Structure [C]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008: 1293-1296.

[109] L. Wang. Feature Selection with Kernel Class Separability [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30 (9): 1534-1546.

[110] Y.W. Chen and C.J. Lin. Combining SVMs with Various Feature Selection Strategies [C]. Advances in Neural Information Processing Systems (NIPS) on Feature Selection Challenge, 2003: 1-10.

[111] K. Polat and S. Güneş. A New Feature Selection Method on Classification of Medical Datasets: Kernel F-score Feature Selection [J]. Expert Systems with Applications, 2009, 36(7): 10367-10373.

[112] M.D. Shieh and C.C. Yang. Multiclass SVM-RFE for Product Form Feature Selection [J]. Expert Systems with Applications: An International Journal, 2008, 35 (1-2): 531-541.

[113] I. Guyon, J. Weston, S. Barnhill, et al. Gene Selection for Cancer Classification using Support Vector Machines [J]. Machine Learning, 2002, 46 (1-3): 389 – 422.

[114] Y. LeCun, J. Denker, S. Solla, et al. Optimal Brain Damages [J]. Advances in Neural Information Processing Systems, 1990: 598–605.

[115] A. Rakotomamonjy. Variable Selection using SVM-based Criteria [J]. Journal of Machine Learning Research, 2003, 3: 1357–1370.

[116] D.Q. Han, C.Z. Han and Y.yang. Multi-class SVM Classifiers Fusion based on Evidence Combination [C]. International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), 2007, 2:579-584.

[117] B. Calvo, P. Larrañaga and J.A. Lozano. Feature Subset Selection from Positive and Unlabelled Examples [J]. Pattern Recognition Letters, 2009, 30(11): 1027-1036.

[118] T.M. Cover and J.A. Thomas. Elements of Information Theory (second edition) [M]. New York: John Wiley & Sons Inc, 2006.

[119] http://en.wikipedia.org/wiki/Support_vector_machine [J/OL].

[120] G. Wang. Evidence Theory and Its Application of Image Recognition [D]. Master Thesis of Northwestern Polytechnical University, 2002.

[121] Y.R. Li. Information Fusion and Intelligent Processing [D]. Doctor Thesis of Zhejiang University, 2001.

[122] http://www.fil.ion.ucl.ac.uk/spm/ [DB/OL].

[123] J. Ashburner, G. Flandin, R. Hensen, et al. SPM Manual [M]. 2008.

[124] A.P. Holmes. Statistical Issues in Functional Brain Mapping [D]. PhD Thesis, University of Glasgow, 1994.

[125] S. Chaplot and N.R. Jagannathan. Classification of Magnetic Resonance Brain Images using Wavelets as Input to Support Vector Machine and Neural Network [J]. Biomedical Signal Processing and Control, 2006, 1 (1): 86-92.

[126] www.csie.ntu.edu.tw/~cjlin/libsvm/ [DB/OL].

[127] K.M. Iftekharuddin, J. Zheng, M.A. IsIam, et al. Fractal-based Brain Tumor Detection in Multimodal MRI [J]. Applied Mathematics and Computation, 2009, 207 (1): 23-41.

[128] P.M. Thompson, A.D. Lee, R.A. Dutton, et al. Abnormal Cortical Complexity and Thickness Profiles Mapped in Williams Syndrome [J]. The Journal of Neuroscience, 2005, 25 (16): 4146-4158.

[129] S. Kido, K. Kuriyama, M. Higashiyama, et al. Fractal Analysis of Internal and Peripheral Textures of Small Peripheral Bronchogenic Carcinomas in Thin-Section Computed Tomography: Comparison of Bronchioloalveolar Cell Carcinomas with Nonbronchioloalveolar Cell Carcinomas [J]. The Journal of Computer Assisted Tomography, 2003, 27 (1): 56 – 61.

[130] J.M. Zook and K.M. Iftekharuddin. Statistical Analysis of Fractal-based Brain Tumor Detection Algorithms [J]. The Journal of Magnetic Resonance Imaging, 2005, 23: 671 – 678.

[131] K.M. Iftekharuddin, M.A. Islam, J. Shaik, et al. Automatic Brain Tumor Detection in MRI Methodology and Statistical Validation [C]. Proceedings of the Progress in Biomedical Optics and Imaging, Medical Imaging, Image Processing, 2005:

[132] H. Khotanlou, O. Colliot, J. Atif, et al. 3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models [J]. Fuzzy Sets and Systems, 2009, 160: 1457-1473.

[133] J.L. Fooa, G. Miyanob, T. Lobec, et al. Three-Dimensional Segmentation of Tumors from CT Image Data using an Adaptive Fuzzy System [J]. Computers in Biology and Medicine, 2009, 62: 869-878.

[134] S. Taheri, S.H. Ong and V.F.H. Chong. Level-set Segmentation of Brain Tumors using a Threshold-based Speed Function [J]. Image and Vision Computing, 2010, 28 (1): 26-37.

[135] S. Ruan, S. Lebonvallet, A. Merabet, et al. Tumor Segmentation from a Multi-Spectral MRI Images by using Support Vector Machine Classification [C]. 4th IEEE International Symposium on Biomedical Imaging (ISBI), 2007: 1236-1239.

[136] I. Aleksander and H. Morton. An introduction to neural computing (Second edition) [M]. Santa Clara: Intel Thomson Computer, 1995

[137] X. Li, S. Lebonvallet, T. Qiu et al. An Improved Level Set Method for Automatically Volume Measure: Application in Tumor Tracking from MRI Image [C]. 29th Annual International Conference on Engineering in Medicine and Biology Society (EMBS), 2007: 808-811.

[138] E. Grall-Maes and P. Beauseroy. Optimal Decision Rule with Class-Selective Rejection and Performance Constraints [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(11): 2073-2082.

FOLIO ADMINISTRATIF

THESE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

NOM : ZHANG **DATE de SOUTENANCE** : 12 Septembre 2011
Prénoms : Nan

TITRE : Feature selection based segmentation of multi-source images: application to tumor segmentation in multi-sequence MRI

NATURE : Doctorat **Numéro d'ordre** :

Ecole doctorale : Ecole Doctorale Electronique, Electrotechnique, Automatique

Spécialité : Image et System

Cote B.I.U. - Lyon : **CLASSE** :

Résumé : Magnetic resonance imaging (MRI) is widely applied to the examination and assistant diagnosis of brain tumors due to its advantages of high resolution to soft tissues and none of radioactive damages to human bodies. Brain tumors have a large diversity in shape and appearance with intensities. Multi-spectral images have the advantage of providing complementary information to resolve some ambiguities. However, they may also bring along a lot of redundant information, increasing data processing time and segmentation errors. The challenge is how to make use of the multi-spectral images effectively. In this thesis, our study focuses on the fusion of multi-spectral images by extracting the most useful features to obtain the best segmentation with the least cost in time. The Support Vector Machine (SVM) classification integrated with a selection of the features in a kernel space is proposed. The selection criterion is defined by the kernel class separability. Based on this SVM classification, a framework to follow up brain tumor evolution is proposed, which consists of the following steps: to learn the brain tumors and select the features from the first MRI examination of the patients; to automatically segment the tumor in new data using a multi-kernel SVM based classification; to refine the tumor contour by a region growing technique; and to possibly carry out an adaptive training.

The features used in our system contain low-frequency coefficients of 2-dimensional wavelets transformation, fractal features and 3-dimentional features. The idea on integration of both feature selection and classifier design in a unified framework is proposed to decrease the difficulty of system design. That is to say, the parameters used by SVM classifier are designed to be relative with the parameters of feature selection by using the kernel-based theory and other relevant theories, in which both parameters can be chosen simultaneously or transformed to each other in order to reduce the difficulty of parameter selection and improve the applicability of our system.

The proposed system is test on 13 patients with 24 examinations, including 72 MRI sequences and 1728 images. Compared with the manual traces of the doctors as the ground truth, the average classification accuracy reaches 98.9%. The system utilizes several novel feature selection methods to test the integration of feature selection and SVM classifiers. Also compared with the traditional SVM, Fuzzy C-means, the neural network and an improved level set method, the segmentation results and quantitative data analysis demonstrate the effectiveness of our proposed system.

MOTS-CLES: brain tumor detection; Support Vector Machine (SVM); feature selection; contour refinement; following-up

Laboratoire (s) de recherche :
CREATIS, CNRS UMR 5220, INSERME U1044, INSA de Lyon, Université de Lyon

Directeur de thèse: Yue-Min ZHU, Su RUAN, Qingmin LIAO

Président de jury :

Composition du jury : Yue-Min ZHU, Su RUAN, Qingmin LIAO, Pierre BEAUSEROY, Da-Tian YE, Yue-Sheng ZHU.