2012

Automatic Identification of Ultrasound Liver Cancer Tumor Using Support Vector Machine

textural 纹理 co-occurrence 共现，共生 speckle 斑点

Abstract-Ultrasound liver tumor image are naturally having more spackle noise. Automatic identification of ultrasound liver tumor image is a challenging task. In this proposed system, we approach fully automatic machine learning system for identifying the liver cancer tumor from ultrasound images. First, we segment the liver image by calculating the textural features from co-occurrence matrix and run length method. This is the best method for segmentation of ultrasound liver cancer tumor images because it is not affected speckle noise and also preserves spatial information. For classification Support Vector machine are a general algorithm based on the risk bounds of statistical learning theory. They have found numerous applications, such as in optical character recognition, object detection, face verification, text categorization and so on. The textural features for different features methods are given as input to the SVM individually. Performance analysis train and test datasets carried out separately using SVM Model. Whenever an ultrasonic liver cancer tumor image is given to the SVM classifier system, the features are calculated, classified, as normal, benign and malignant liver cancer tumor. We hope the result will be helpful to the physician to identify the liver cancer in non-invasive method.

Keywords-Segmentation, Support Vector Machine, Ultrasound Liver Cancer Tumor

使用支持向量机来超声自动识别肝癌肿瘤

摘要

超声肝肿瘤图像具有更多的散斑噪声。超声自动识别肝肿瘤图像是一项具有挑战性的任务。在提出的系统中，我们以接近全自动机器学习系统的方法从超声图像识别肝癌肿瘤。首先，我们通过计算共生矩阵的纹理特征分割肝脏图像，然后运行长度方法。这是用于超声肝癌肿瘤图像分割的最佳方法，因为它不受斑点噪声的影响并且还保留空间信息。分类支持向量机是基于统计学习理论风险边界的一般算法。他们已经发现了许多应用，例如在光学字符识别，对象检测，面部验证，文本分类等。不同特征方法的纹理特征作为单独的SVM的输入。使用SVM模型单独在训练和测试数据集进行性能分析。每当超声肝癌肿瘤图像被给予SVM分类器系统时，计算特征，分类为正常，良性和恶性肝癌肿瘤。我们希望结果将有助于医生以非侵入性方法鉴定肝癌。

关键字:分割 支持向量机 超声肝癌肿瘤

I. INTRODUCTION

In the medical field computer are now being used virtually in every aspect of modern medicine. Computers are used widely in medical research, where there is a vital need for better microelectronic sensors for data acquisition. Imaging modalities like Ultrasound, MRI (Magnetic Resonance Imaging), CT (Computed Tomography) and PET (Positron Emission Tomography) are widely used techniques for liver cancer tumor diagnosis [1]. Liver cancer tumor is sixth dangerous diseases in the world. Liver diseases are considered seriously because of the liver's vital importance to human beings. There are two classes of liver tumors: benign and malignant [2]. Ultrasound image is a powerful tool for characterizing the state of soft tissues for medical diagnostic purposes. Ultrasound has been extremely valuable in differentiating a simple liver cancer tumor from other liver masses.

An approach has been made in this research to design a diagnostic classifier system for the identification of liver cancer tumor in ultrasound images using image texture features in non-invasive manner. Image processing modifies pictures to improve them (enhancement, restoration), extract information (analysis, recognition), and change their structure (composition, image editing). Images can be processed by optical, photographic, and electronic means, but image processing using digital computers is the most common method because digital methods are fast, flexible, and precise [3]. So, the proposed system we applied the co-occurrence matrix features and gray level run-length features for identifying the seed point for given ultrasound liver images. After the detection of automated seed point we have to segment the liver image applying the region growing algorithm using gray space map and Otsu algorithm. After segmentation of the image we analyzed calculated texture features parameters to classified, as normal, benign and malignant liver cancer tumor. We explain the image processing procedures, segmentation of image in the section 2, Computation of different image texture features of different feature extraction methods namely first order statistics, run length statistics, wavelet based texture features in section 3, machine learning classifier SVM used for classification of Ultrasound Liver cancer tumor image using texture are describes in the section 4, Trialing result in section 5 and conclusion in the section 6.

I.引言

在医疗领域中，计算机现在被用于现代医学的各个方面。计算机广泛用于医学研究，其中存在对用于数据采集的更好的微电子传感器的重要需求。成像模式如超声，MRI（磁共振成像），CT（计算机断层扫描）和PET（正电子发射断层扫描）是广泛使用的肝癌肿瘤诊断技术[1]。肝癌肿瘤是世界上第六种危险的疾病。肝脏疾病被认为是严重的，因为肝脏对人类至关重要。有两类肝肿瘤：良性和恶性[2]。超声图像是用于表征用于医学诊断目的的软组织的状态的强有力的工具。超声对于区分简单的肝癌肿瘤与其他肝脏肿瘤非常有价值。

 在该研究中已经做出了一种方法来设计用于以非侵入方式使用图像纹理特征来识别超声图像中的肝癌肿瘤的诊断分类器系统。图像处理修改图片以改进它们（增强，恢复），提取信息（分析，识别）以及改变它们的结构（构图，图像编辑）。图像可以通过光学，照相和电子手段进行处理，但是使用数字计算机的图像处理是最常见的方法，因为数字方法是快速，灵活和精确的[3]。因此，所提出的系统我们应用共生矩阵特征和灰度行程长度特征来识别给定超声肝图像的种子点。在检测到自动种子点后，我们必须使用灰色空间图和Otsu算法来分割肝脏图像，应用区域生长算法。分割图像后，我们分析计算出的纹理特征参数进行分类，作为正常，良性和恶性肝癌肿瘤。我们解释图像处理程序，图像分割在第2节，不同特征提取方法的不同图像纹理特征的计算，即第一阶统计，运行长度统计，第3节中的小波基纹理特征，用于分类的机器学习分类器的超声肝癌肿瘤图像使用纹理描述在第4节，试验结果在第5节和结论在第6节。

II. MATERIAL AND METHODOLOGY

2.1. Image Preprocessing

An Ultrasound liver cancer tumor images has been taken for this study. The preprocessing step typically is used for reduce the noise and to prepare the ultrasound liver image for further processing such as segmentation and classification. To get a high-pass filter, the general procedure is to apply a low-pass filter to the original image and then subtract this low-frequency image from the original image. The result is then an image containing only high frequencies. Sometimes it is desired to enhance the high frequencies without removing the low frequencies. This is called giving the image a high-frequency boost. The preprocessing work could be done for removing the noise of the images. After the removal of noise from the image we applied the histogram to identify the maximum of the intensity value. Then we applied the techniques for segmentation of the ultrasound liver cancer tumor.

II 材料和方法

2.1图像预处理

已经为该研究采用了超声肝癌肿瘤图像。 预处理步骤通常用于减少噪声并且准备超声肝图像以用于诸如分割和分类的进一步处理。为了获得高通滤波器，通常的过程是对原始图像应用低通滤波器，然后从原始图像中减去该低频图像。结果是仅包含高频的图像。 有时期望增强高频而不去除低频。 这被称为给予图像高频提升。 可以进行预处理工作以去除图像的噪声。 在从图像中去除噪声之后，我们应用直方图来识别强度值的最大值。 然后我们应用超声肝癌肿瘤的分割技术。

2.2. Segmentation

Segmentation is played an important role in the image processing. Normally, Segmentation of Ultrasound images are very difficult because it contains more speckle noise. Segmentation of medical images involves three main image related problems. Images contain noise that can alter the intensity of a pixel such that its classification becomes uncertain, images exhibit intensity no uniformity where the intensity level of a single tissue class varies gradually over the extent of the image, and images have finite pixel size and are subject to partial volume averaging where individual pixel volumes contain a mixture of tissue image may not be consistent with any one class [4].Segmentation of ultrasound liver cancer tumor is more critical because it contains more speckle noise and artifacts. The proposed system we planned to apply the co-occurrence matrix features and gray level run-length features for identifying the seed point for given ultrasound liver images. After the detection of automated seed point we have to segment the liver image applying the region growing algorithm using gray space map and Otsu algorithm for segmenting the ultrasound liver image. These Co-occurrence matrix features and the run length also used for the classification of the ultrasound liver cancer tumor images.

2.2分割

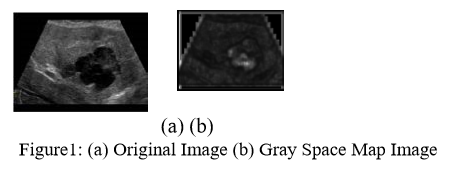
分割在图像处理中起重要作用。通常，超声图像的分割是非常困难的，因为它包含更多的斑点噪声。医学图像的分割涉及三个主要的图像相关问题。图像包含可以改变像素的强度使得其分类变得不确定的噪声，图像表现的强度没有均匀性，其中单个组织类别的强度水平在图像的范围上逐渐变化，并且图像具有有限的像素尺寸，其中单个像素体积包含组织图像的混合物可能与任何一个类别不一致[4]。超声肝癌肿瘤的分割更关键，因为它包含更多的斑点噪声和伪像。所提出的系统，我们计划应用共生矩阵特征和灰度行程长度特征，用于识别给定超声肝图像的种子点。在检测到自动种子点之后，我们必须使用用于分割超声肝图像的灰度空间图和Otsu算法来应用区域生长算法来分割肝图像。这些共生矩阵特征和游程长度也用于超声肝癌肿瘤图像的分类。

2.3.Gray Space Map

The algorithm of region growing is very simple. We compute the seed gray level: U, then look for structures which have the same gray level than the seed overlapping the seed position. At the second iteration, we look for structures having a small gray level difference from the seed. In other words we define a set of gray levels from U-D to U+D. Then we keep those structures which overlap the seed position. At each iteration we increase the difference D by 1. In this way structures which are closed from a spatial AND intensity point of view to the seed are highlighted with higher values [6]. In new image if we far spatially and from an intensity the point of view from the seed, the lower intensity is labeled. The resulted image is Gray Space map of image.

2.3.Gray空间地图

区域生长的算法非常简单。 我们计算种子灰度级：U，然后查找具有与种子位置重叠的种子相同的灰度级的结构。 在第二次迭代时，我们寻找与种子具有小的灰度差的结构。 换句话说，我们定义一组从U-D到U + D的灰度级。 然后我们保留那些与种子位置重叠的结构。 在每次迭代中，我们将差值D增加1.以这种方式，从空间和强度观点到种子关闭的结构以更高的值突出[6]。 在新图像中，如果我们在空间上远离强度从种子的角度，则标记较低的强度。 所得到的图像是图像的灰色空间图。

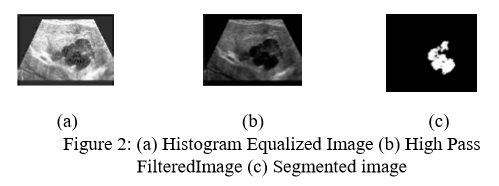


2.4. Region Growing Segmentation

First we find the maximum area variation in which means that from this intensity to 0 we are sure that this is not the ROI. Second we cut the histogram from MAX to 0. Then, we have to find the threshold from MAX to the highest intensity which separates the uncertainty area from the ROI. This is simply done using the well-known Otsu thresholding method [7]. This is a parameter free thresholding technique which maximizes the inter-class variance. It is interesting to observe that the Otsu method is more accurate in cutting into two classes. Otsu also takes care to get compact clusters using the inter-class variance. In Fig.2 we can see the segmented image.

2.4。 区域生长分割

首先，我们发现最大面积变化意味着从这个强度到0，我们确信这不是ROI。 其次，我们从切MAX直方图为0。然后，我们必须找到MAX以分隔从投资回报率的不确定性领域的最高强度的门槛。 这是使用着名的Otsu阈值法[7]。 这是最大化类间方差的无参数阈值技术。 有趣的是观察到Otsu方法在切割成两个类时更准确。 Otsu也注意使用类间方差获得紧凑的集群。 在图2中，我们可以看到分割的图像。



III. TEXTURE FEATURE EXTRACTION METHODS

Texture feature extraction is the procedure of generating descriptions of a textured surface in terms of measurable parameters. The extracted features represent the relevant properties of the surface, and may be used with a classifier. The following textural features groups are used in the proposed system, First order statistics (Histogram), Second order statistics, Run - length matrices and Wavelet features. 3.1 First Order Statistics in this method, the features are derived from the gray level histogram. The digital image can be represented as a two-dimensional array in the computer. For the digital images, 8 bits are sufficient and the gray-level values range from 0 to 255.Lower values are attributed to darker pixels, and higher values to brighter pixels. Therefore 0 represents the black and white represents the 255. 3.2. Co-occurrence Matrix Feature A Co-Occurrence Matrix (COM) is square matrices of relative frequencies P (i, j, d, q) with which two neighboring pixels separated by distance d at orientation q occur in the image, one with gray level i and the other with gray level j[4]. Therefore, a square matrix that has the size of the largest pixel value in the image and presents the relative frequency distributions of gray levels and describe how often one gray level will appear in a specified spatial. In our project 2 textural features were calculated from the COM for direction h values of 0° and a distance d of 1. In this work the co-occurrence features energy and entropy which can easily differentiate non-homogeneous region from homogeneous region are considered. Energy is called Angular Second Moment. It is a measure the homogeneousness of the image and can be calculated from the normalized COM. Energy is expected to be high if the occurrence of repeated pixel pairs is high. It denotes the normalized co-occurrence matrix by total number of the occurrence of two neighboring pixels between I gray-intensity at vertical direction and angle Ө. Entropy gives a measure of complexity of the image. Complex textures tend to have higher entropy if the gray levels are distributed randomly through out of the image. These two parameters can identify seed pixel from the abnormal region of the ultrasound liver cancer tumor images. Some times for some cases the normal liver region also can appear be a homogeneous. So to avoid that situation by calculating the run length features.

III。纹理特征提取方法

 纹理特征提取是根据可测量参数生成纹理表面的描述的过程。提取的特征表示表面的相关属性，并且可以与分类器一起使用。在所提出的系统中使用以下纹理特征组，一阶统计（直方图），二阶统计，运行长度矩阵和小波特征。

3.1一阶统计

在该方法中，从灰度直方图导出特征。数字图像可以表示为计算机中的二维阵列。对于数字图像，8位是足够的，并且灰度值的范围从0到255.较低的值归因于较暗的像素，较高的值归因于较亮的像素。因此0表示黑白表示255。

3.2 共生矩阵特征

共生矩阵（COM）是相对频率P（i，j，d，q）的方阵，利用该矩阵，在方位q处以距离d分离的两个相邻像素出现在图像中，一个具有灰度级i，另一个是灰度j [4]。因此，具有图像中最大像素值的大小的方矩阵，并呈现灰度级的相对频率分布，并描述一个灰度级将在指定空间中出现的频率。在我们的项目2中，从COM的方向h值为0°和距离d为1计算了纹理特征。

在这项工作中，共生特征是能量和熵，可以容易地区分非均匀区域和均匀区域。能量称为角二次矩。它是图像的均匀性的度量，并且可以从归一化的COM计算。如果重复的像素对的出现高，则期望高。

它表示归一化的共生矩阵与垂直方向上的I灰度与角度θ之间的两个相邻像素的出现总数。熵给出图像的复杂性的度量。如果灰度级随机分布到图像中，则复杂纹理倾向于具有较高的熵。这两个参数可以从超声肝癌肿瘤图像的异常区域识别种子像素。有时，对于一些情况，正常肝脏区域也可以是均质的。所以通过计算运行长度特征来避免这种情况。

3.3. Gray Level Run-Length Features

In ultrasound liver images, there are run-length features calculated from run–length matrix that are capable of capturing the texture primitives’ properties for different structures in 2D image data, such as the homogeneous texture structure of the image. It denotes the number of runs of length and gray level occurring in the image region. LongRunEmphasis (LRE) feature measures distribution of long runs. The LRE is highly dependent on the occurrence of ling runs and is expected large for coarse structural textures. RunLengthNon-uniformity (RLN)measures the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning. Using K-means clustering classified the ultrasound liver cancer image as Normal, Benign, and Malignant.

3.3灰度级运行长度特性

在超声肝图像中，存在从游程长度矩阵计算的游程长度特征，其能够捕获2D图像数据中的不同结构的纹理图元的属性，例如图像的均匀纹理结构。它表示在图像区域中发生的长度和灰度级的游程的数量。 LongRunEmphasis（LRE）功能测量长运行的分布。 LRE高度依赖于ling运行的发生，并且预期对于粗糙结构纹理是大的.RUNLength非一致性（RLN）测量将一组对象分配到组（称为集群）的任务，使得同一集群中的对象彼此比其他群集中的那些更相似。聚类是探索性数据挖掘的主要任务，并且是用于许多领域（包括机器学习）中的统计数据分析的常用技术。使用K均值聚类将超声肝癌图像分类为正常，良性和恶性。

V. RESULTS

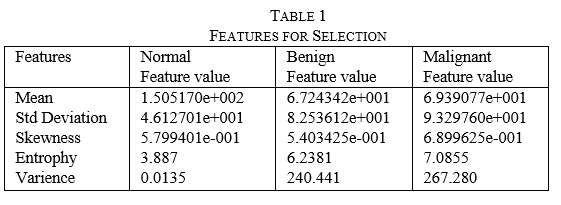
We describe the performance of the SVM classifier to identification of liver cancer tumor from different textural features method used for segmentation and classification. The texture features are extracted from the high intensity value using Otsu’s thresholding method. SVM used as a classifier.

This classifier offers the better results in identifying the malignant liver tumor from the normal. In some cases, the benign tumors are misclassified as a normal liver. The intensity value of the normal liver and begin may be same. The overall accuracy of classification value 96.72% approximately for the ultrasound liver images.

V.结果

我们描述了SVM分类器的性能，以从用于分割和分类的不同结构特征方法鉴定肝癌肿瘤。 使用Otsu的阈值法从高强度值提取纹理特征。 SVM用作分类器。

这种分类器在从正常识别恶性肝肿瘤中提供更好的结果。 在一些情况下，良性肿瘤被错误分类为正常肝脏。 正常肝脏的强度值和开始可以相同。 对于超声肝图像，分类值的总精度大约为96.72％。



VI. CONCLUSION

This paper proposes the automatic identification of ultrasound liver cancer tumor using SVM classifier. First we detect the seed point for the given ultrasound liver image automatically using features of co-occurrence matrix and run length method. Second, we segment the ultrasound liver images using of gray space map and Otsu method. Finally using SVM classifier we classified the ultrasound liver cancer tumor image as normal, benign and malignant. We hope this similarity of the length of the runs throughout the image. The RLE is expected small if the run lengths are alike throughout the image. These run length features will check the selected seed point of the image which is calculated from the co-occurrence matrix is belongs to affected region of the liver image or not.

IV. SVM CLASSIFIER

Support Vector Machine (SVM) performs the robust nonlinear classification with kernel trick. SVM is independent of the dimensionality of the feature space and that the results obtained are very accurate. It outperforms other classifiers even with small numbers of available training samples. SVM is a supervised learning method and is used for one class and n class classification problems [5]. Cluster analysis or clustering system will help the physician to diagnose the liver cancer with non-invasive method

VI结论

本文提出使用SVM分类器自动识别超声肝癌。首先，我们使用共生矩阵和运行长度方法的特征自动检测给定超声肝图像的种子点。第二，我们使用灰色空间图和Otsu方法分割超声肝图像。最后使用SVM分类器，我们将超声肝癌肿瘤图像分类为正常，良性和恶性。我们希望这种相似的运行的长度在整个图像。如果运行长度在整个图像中是相似的，则RLE预期很小。这些游程长度特征将检查从共生矩阵计算的图像的所选种子点是否属于肝脏图像的受影响区域。

IV SVM分类器

 支持向量机（SVM）用内核技巧执行鲁棒的非线性分类。 SVM独立于特征空间的维度，并且所获得的结果是非常准确的。它优于其他分类器，即使有少量的可用训练样本。 SVM是一种有监督的学习方法，用于一类和n类分类问题[5]。聚类分析或聚类系统将帮助医生用非侵入性方法诊断肝癌

Support Vector Machine based Liver Cancer Early Detection using Magnetic Resonance Images

Abstract

Magnetic Resonance Imaging (MRI) has become an important tool for doctors to diagnose liver cancer for decays. The survival rate of liver cancer patients can be significantly improved by an early diagnosis. In this paper, we present a computer aided kernel based support vector machine (SVM) algorithm for diagnosing liver cancer in early stage by applying our proposed method to the patients’ magnetic resonance (MR) images. We apply the histogram-based feature extraction method to extract feature information from each raw MR image acquired. And 100 confirmed liver cancer and 100 confirmed benign type liver tumor (BLT) patients’ feature information are used to form our training data set to train or SVM classification engine. The model is tested with a set of 30 confirmed early stage liver cancer and 30 BLT samples. Our trained SVM achieves an accuracy of 86.67% in classifying early stage liver cancer and 80.00% in classifying BLT.

Keywords—Classification, Histogram-based feature, Kernel, Machine learning, Diagnosis assistance, MR images

磁共振成像（MRI）已经成为医生诊断肝癌的衰变的重要工具。肝癌患者的存活率可以通过早期诊断显着改善。在本文中，我们提出了一种基于计算机辅助内核的支持向量机（SVM）算法，用于早期诊断肝癌，通过将我们提出的方法应用于患者的磁共振（MR）图像。我们应用基于直方图的特征提取方法从所获取的每个原始MR图像提取特征信息。 100例确诊的肝癌和100例确诊的良性肝癌（BLT）患者的特征信息用于形成训练数据集或SVM分类引擎。该模型用一组30个确认的早期肝癌和30个BLT样品进行测试。我们训练的SVM在分类早期肝癌中达到86.67％的准确度，在分类BLT中达到80.00％的准确度。

关键词 - 分类，基于直方图的特征，内核，机器学习，诊断帮助，MR图像

I. INTRODUCTION

Liver is the largest internal organ in human body and the very important part for numerous metabolic, regulatory, transport, and immune functions to maintain human lives. Liver cancer (also known as hepatocellular carcinoma) is one of the most lethal diseases in the world. In Pacific Rim and Southeast Asia area, liver cancer is responsible for at least 400,000 people’s death every year [1]. It is still very difficult to eradicate liver cancer in the late stage, but with numerous possible treatments have been developed, the survival rate of liver cancer has been increased significantly if patient can be diagnosed in early stage. Thus, the importance and benefit of a method of diagnosing liver cancer in early stage are obvious. Based on the physical principles of MR scanners [2], MR scans have been used by doctors in diagnosing lesions in brain, nervous system and solid organs manually for more than 30 years. Many researchers in computer vision and machine learning field have done a lot of work in developing MR image based automatic classification systems. For example, classification of tumours in brain [3][4] and prostate [5] by machine learning scheme with MRI images have been proved with high accuracy rate. Classification of liver diseases have also been done by researchers. Detection of liver metastases and liver fibrosis from MRI images under machine learning ([6], [7] and [8]) have all get a high classification accuracy rate.

In this work, we present an automatic detection method of early stage liver cancer by machine learning approach. First, we construct a machine learning model trained by our acquired 200 confirmed liver cancer and BLT patients’ MR image samples. Then, new testing data set obtained additionally which contains 30 confirmed early staged liver cancer and 30 BLT samples are used to test the performance of our model. We expect our proposed method can help doctors and radiologists in improving the diagnose rate of early stage liver cancer.

I.引言

肝是人体中最大的内脏器官，是众多代谢，调节，运输和免疫功能维持人类生命的重要部分。肝癌（也称为肝细胞癌）是世界上最致命的疾病之一。在太平洋沿岸和东南亚地区，肝癌每年至少造成40万人死亡[1]。在后期仍然很难根除肝癌，但是已经开发了许多可能的治疗方法，存活率的肝癌已显着增加，如果病人可以在早期诊断。因此，早期诊断肝癌的方法的重要性和益处是显而易见的。基于MR扫描仪的物理原理[2]，医生在手术中使用MR扫描诊断大脑，神经系统和实体器官的病变已有30多年。许多计算机视觉和机器学习领域的研究人员在开发基于MR图像的自动分类系统方面做了大量工作。例如，通过具有MRI图像的机器学习方案对脑中的肿瘤[3] [4]和前列腺[5]的分类已经证明具有高准确率。肝脏疾病的分类也由研究人员完成。在机器学习下从MRI图像检测肝转移和肝纤维化（[6]，[7]和[8]）都具有高的分类准确率。

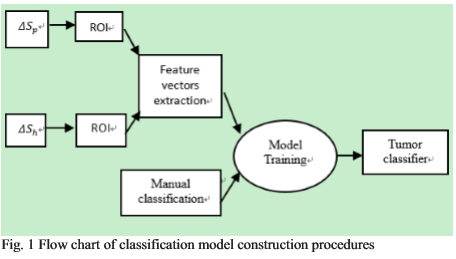
在这项工作中，我们提出一种通过机器学习方法的早期肝癌的自动检测方法。首先，我们构建一个机器学习模型训练由我们收购200确诊的肝癌和BLT患者的MR图像样本。然后，另外获得的包含30个确认的早期分期肝癌和30个BLT样品的新测试数据集用于测试我们的模型的性能。我们期望我们提出的方法可以帮助医生和放射科医生提高早期肝癌的诊断率。

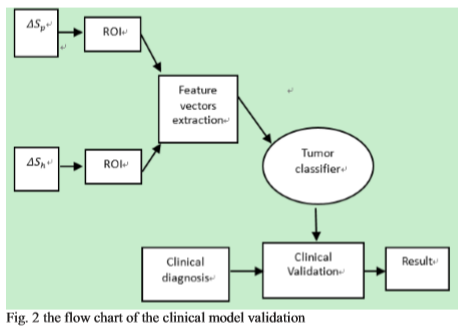
II. DESIGN AND IMPLEMENTATION METHODOLOGY

Our method of designing the classification system is introduced by the flowchart in Fig 1 and 2. The first step of constructing the classification system is forming a training data set contains the information of confirmed liver cancer and BLT MR images we gained. Subsequently, after finish the model construction, we test our model’s performance by using another testing data set that contains information of 30 confirmed early staged liver cancer and 30 BLT MR images. We will introduce each component of our system in this section

II。 设计和实现方法

我们的分类系统的设计方法由图1和图2中的流程图引入。构建分类系统的第一步是形成包含确认的肝癌和我们获得的BLT MR图像的信息的训练数据集。 随后，完成模型构建后，我们使用另一个包含30个确诊的早期肝癌和30个BLT MR图像的测试数据集来测试我们的模型的性能。 我们将在本节中介绍我们系统的每个组件



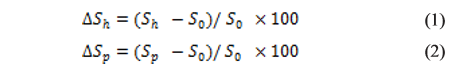


A. Input Data and Pre-processing

In our experiments, we acquire 3 MR images for each patient. They are all transverse relaxation time (T2) weighted MR images. The first image, S0, is MR image without contrast agent. The other two images are captured under the effect of contrast agent. Sh, is MR image captured when the peak value of contrast agent appears in the hepatic artery. Sp is MR image captured when the peak value of the contrast agent is in the Portal vein. According to the experience from expert doctors and other researchers’ work [6], maps of the MR signal intensity difference between MR images with contrast agent ( , and ) and MR images without contrast agent ( ) are used as the input data of our system. The changes of the MR signal intensities and are calculated as follows:

A.输入数据和预处理

在我们的实验中，我们为每个患者采集3张MR图像。 它们都是横向弛豫时间（T2）加权的MR图像。 第一图像S0是没有造影剂的MR图像。 其他两个图像是在造影剂的影响下捕获的。 Sh是当造影剂的峰值出现在肝动脉中时拍摄的MR图像。 Sp是当造影剂的峰值在门静脉中时捕获的MR图像。 根据专家医生的经验和其他研究者的工作[6]，使用造影剂（和）和没有造影剂（MR）的MR图像的MR信号强度差图作为我们系统的输入数据 。 MR信号强度的变化计算如下：



B. Region of interest (ROI) selection and feature extraction

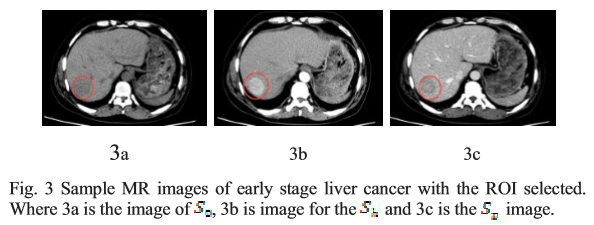
Our classification system will not accept the entire piece of MR image as input data. Thus, for each Sh or Sp computed from (1) or (2), one or more ROIs that actually contain the liver tumor need to be selected out. In our project, all the ROIs are manually identified by experienced doctors and radiologists according to the observations of their anatomical MR images and the pathology results from the surgical operations.

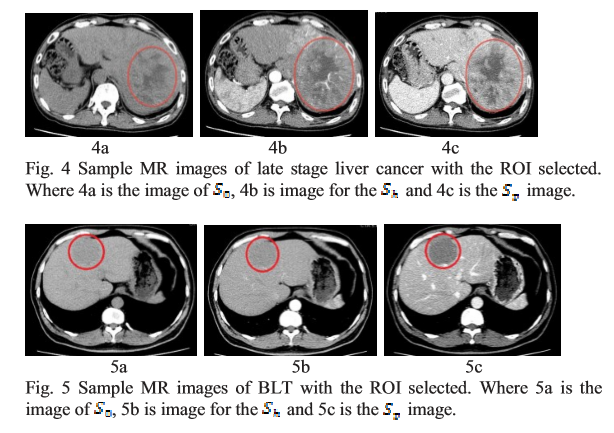
Fig. 3, 4 and 5 show the selected ROIs for early stage liver cancer, late stage liver cancer and BLT image samples respectively.

B.感兴趣区域（ROI）选择和特征提取

我们的分类系统将不接受整张MR图像作为输入数据。 因此，对于从（1）或（2）计算的每个Sh或Sp，需要选择实际包含肝肿瘤的一个或多个ROI。 在我们的项目中，所有的ROI都由经验丰富的医生和放射科医生根据其解剖MR图像的观察结果和外科手术的病理结果手动识别。

图。 图3,4和5分别显示了早期肝癌，晚期肝癌和BLT图像样品的选择的ROI。





C. Model construction

Many recent research results in the medical image classification field, for example [6], show the result that Support vector machine (SVM) [9] method can give the most favorable performance compared with other well-used machine learning methods such as Linear discriminant analysis method (LDA) [10] and K-nearest neighbors method (k-NN) [11]. Thus, in our project, we apply SVM classification algorithm to perform the model construction.

According to many liver MR image based classification results such as [6], [7] and [8], and biomedical image classification works in other organs such as [3] and [12], Radial Basis Function (RBF) gives the best performance when works together with SVM algorithm. In this case, we choose to use RBF as the kernel function of our SVM engine. The RBF kernel function in the SVM engine is described by

C.模型构建

在医学图像分类领域中的许多最近的研究结果，例如[6]，示出支持向量机（SVM）[9]方法可以给出与其他常用的机器学习方法相比最有利的性能的结果，例如线性判别 分析方法（LDA）[10]和K最近邻法（k-NN）[11]。 因此，在我们的项目中，我们应用SVM分类算法来执行模型构建。

  根据许多基于肝MR图像的分类结果如[6]，[7]和[8]，以及其他器官的生物医学图像分类工作[3]和[12]，径向基函数 性能与SVM算法配合使用。 在这种情况下，我们选择使用RBF作为我们的SVM引擎的内核函数。 SVM引擎中的RBF核函数由



where xi and xj are the feature vectors for different sample data, a preset parameter γ is a preset parameter to form our kernel function. Since we only perform two-class classification operations in our project, a C-support vector classification (C-SVC) engine [13] [14] [15] is applied to learn and perform the classification work with our data samples.

其中xi和xj是不同样本数据的特征向量，预设参数γ是形成我们的核函数的预设参数。 由于我们只在我们的项目中执行两类分类操作，所以应用C支持向量分类（C-SVC）引擎[13] [14] [15]来学习和执行我们的数据样本的分类工作。

III. EXPERIMENTAL RESULTS

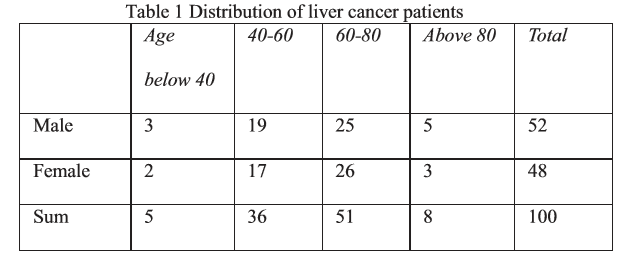
A. MRI data acquisition

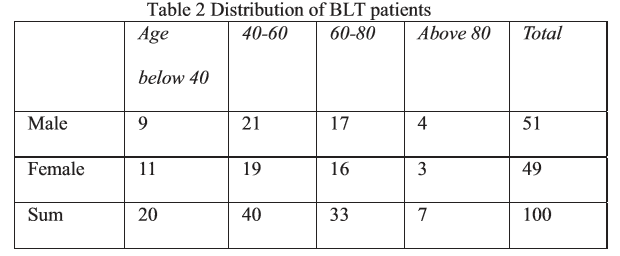
In our project, we acquired 100 liver cancer and 100 BLT samples to form our training data set. The distribution of our sample patients are shown in Tables 1 and 2, respectively.

III 实验结果

A.MRI数据采集

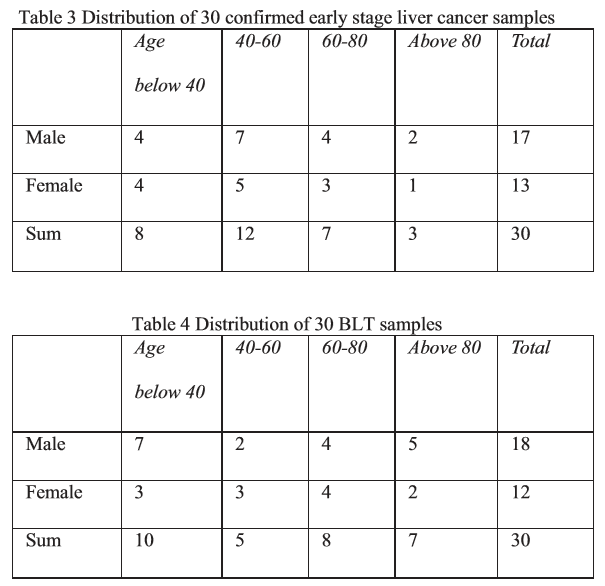
在我们的项目中，我们收集了100个肝癌和100个BLT样品，形成我们的训练数据集。 我们的样本患者的分布分别示于表1和2。





To test and verify the trained machine, we conducted the testing data set by 30 confirmed early stage liver cancer samples and another 30 BLT samples shown in Table 3 and 4, respectively.

为了测试和验证受训的机器，我们分别由30个确认的早期肝癌样品和另外30个BLT样品进行了测试数据集，如表3和4所示。



B. Data normalization, model training and testing

We perform the data normalization to eliminate the effects of signal intensity differences between image samples and narrow down the samples’ variance for selecting parameter γ. The normalization is carried out as follows:

B.数据规范化，模型训练和测试

我们执行数据归一化以消除图像样本之间的信号强度差异的影响并且缩小样本的方差以选择参数γ。 归一化如下进行：



where fk\* is the normalized feature vector,fk is the histogram feature vector of each ROI sample, ufk and afk are the mean and standard deviation values of each element in the feature vector fk.

After normalized all the training and testing data samples, we use the training data set contains 100 liver cancer and 100 BLT samples to train our SVM engine. When the SVM model is constructed, the 30 early stage liver cancer and 30 BLT samples in the testing data set are used to test the performance of our classification engine. In our experiment, our SVM engine correctly classifies 26 early stage liver cancer samples and 24 BLT samples out of 30.

其中fk \*是归一化特征向量，fk是每个ROI样本的直方图特征向量，ufk和afk是特征向量fk中每个元素的平均值和标准偏差值。

在归一化了所有训练和测试数据样本后，我们使用包含100个肝癌和100个BLT样本的训练数据集来训练我们的SVM引擎。 当构建SVM模型时，使用测试数据集中的30个早期肝癌和30个BLT样品来测试我们的分类引擎的性能。 在我们的实验中，我们的SVM引擎正确地分类了26个早期肝癌样品和24个BLT样品中的30个。

IV. CONCLUSION

In this paper, we present a computer aided classification method of early stage liver cancer diagnosis based on liver MR images. By applying histogram based feature vectors extracted from substantial clinical samples of liver cancer and BLT, a kernel based SVM tumor classifier is trained. The effectiveness of the method is also validated by experimental tests with clinical testing data. From these experimental results, the trained SVM achieves an accuracy of 86.67% in classifying early stage liver cancer and 80.00% in BLT. According to experience of expert doctors and radiologists, the classification results of our model are much better than the accuracy of diagnosis early stage liver cancer by the nakedeye observations. Therefore, our proposed method is solid in theory and can be used in the practice. In the future, we plan to practice more with methods of higherlevel texture analysis features and other advanced classification techniques to improve the classification results of our system.

IV。结论

在本文中，我们提出了基于肝MR图像的早期肝癌诊断的计算机辅助分类方法。通过应用从肝癌和BLT的基本临床样品提取的基于直方图的特征向量，训练基于核的SVM肿瘤分类器。该方法的有效性也通过临床测试数据的实验测试验证。从这些实验结果，训练的SVM在分类早期肝癌中达到86.67％的准确度，在BLT中达到80.00％的准确度。根据专家医生和放射科医师的经验，我们的模型的分类结果比通过裸眼观察的早期肝癌的诊断的准确性好得多。因此，我们提出的方法在理论上是固定的，并且可以在实践中使用。在未来，我们计划更多地使用更高层次的纹理分析特征和其他高级分类技术的方法来改进我们的系统的分类结果。