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（co-occurrence）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

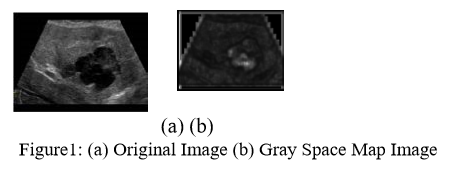
分割在图像处理中起重要作用。通常，超声图像的分割是非常困难的，因为它包含更多的斑点噪声。医学图像的分割涉及三个主要的图像相关问题。图像包含可以改变像素的强度使得其分类变得不确定的噪声，图像表现的强度没有均匀性，其中单个组织类别的强度水平在图像的范围上逐渐变化，并且图像具有有限的像素尺寸，其中单个像素体积包含组织图像的混合物可能与任何一个类别不一致[4]。超声肝癌肿瘤的分割更关键，因为它包含更多的斑点噪声和伪像。所提出的系统，我们计划应用共生矩阵特征和灰度行程长度特征，用于识别给定超声肝图像的种子点。在检测到自动种子点之后，我们必须使用用于分割超声肝图像的灰度空间图和Otsu算法来应用区域生长算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

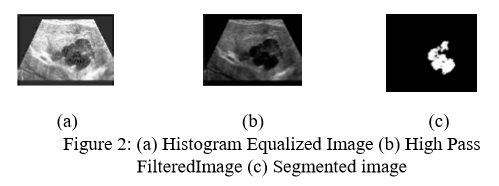


2.4. Region Growing Segmentation

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

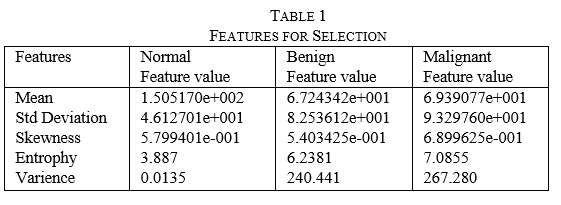
We describe the performance of the SVM classifier to identification of liver cancer tumor from different textural features method used for segmentation and classification. The texture features are extracted from the high intensity value using Otsu’s thresholding method. SVM used as a classifier.

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。

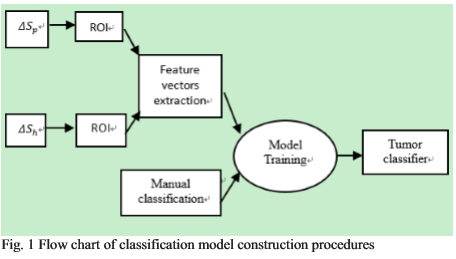
在这项工作中，我们提出一种通过机器学习方法的早期肝癌的自动检测方法。首先，我们构建一个机器学习模型训练由我们收购200确诊的肝癌和BLT患者的MR图像样本。然后，另外获得的包含30个确认的早期分期肝癌和30个BLT样品的新测试数据集用于测试我们的模型的性能。我们期望我们提出的方法可以帮助医生和放射科医生提高早期肝癌的诊断率。

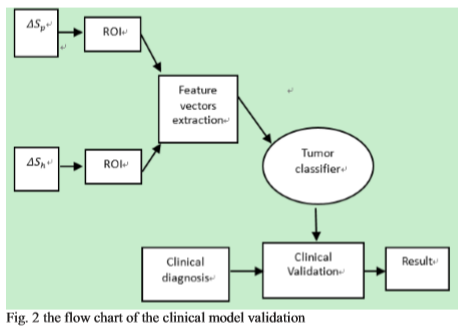
II. DESIGN AND IMPLEMENTATION METHODOLOGY

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



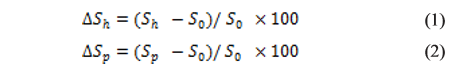


A. Input Data and Pre-processing

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

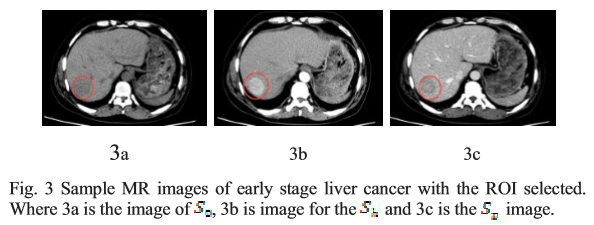
Our classification system will not accept the entire piece of MR image as input data. Thus, for each Sh or Sp computed from (1) or (2), one or more ROIs that actually contain the liver tumor need to be selected out. In our project, all the ROIs are manually identified by experienced doctors and radiologists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

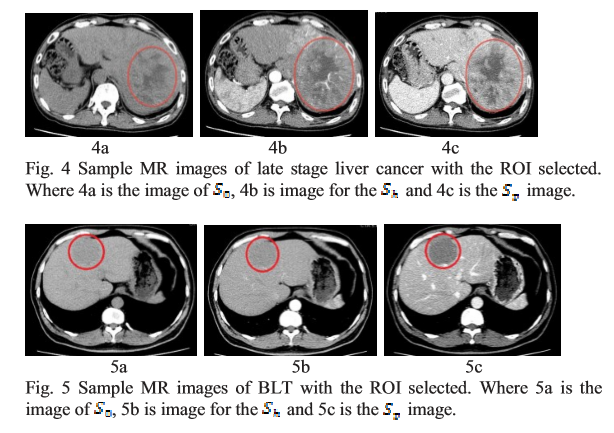
Fig. 3, 4 and 5 show the selected ROIs for early stage liver cancer, late stage liver cancer and BLT image samples respectively.

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

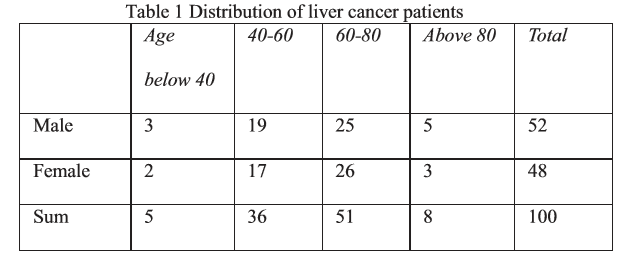
A. MRI data acquisition

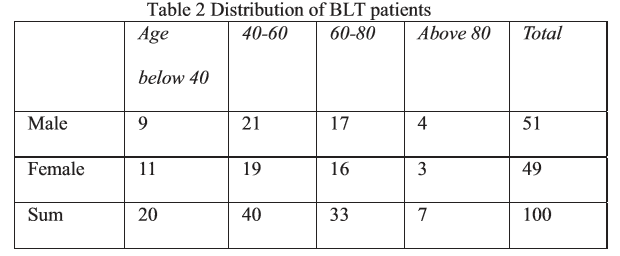
In our project, we acquired 100 liver cancer and 100 BLT samples to form our training data set. The distribution of our sample patients are shown in Tables 1 and 2, respectively.

III 实验结果

A.MRI数据采集

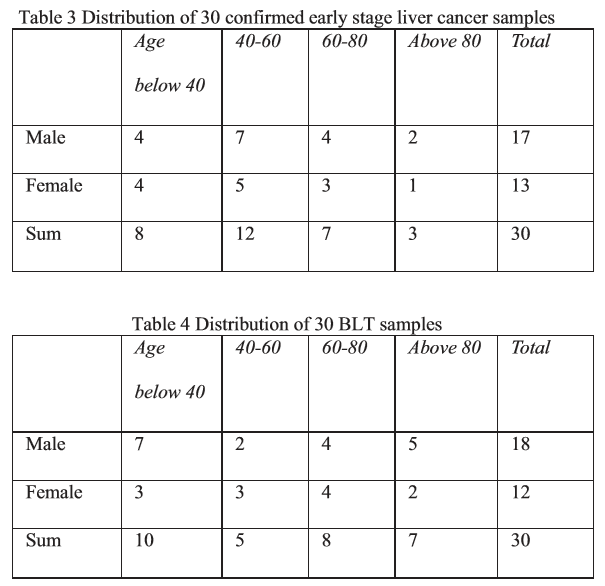
在我们的项目中，我们收集了100个肝癌和100个BLT样品，形成我们的训练数据集。 我们的样本患者的分布分别示于表1和2。





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。

2014 A Study of Support Vector Machine Algorithm for Liver Disease Diagnosis

Abstract

Patients with liver disease have been continuously increasing because of excessive consumption of alcohol, inhale of harmful gases, intake of contaminated food, pickles and drugs. The liver has many essential functions, and liver disease presents a number of concerns for the delivery of medical care. Chronic liver disease (CLD) is common long-term conditions in the developed and developing world. Classification techniques are very popular in various automatic medical diagnosis tools. Early identification of the cancer has been often vital for the survival of the patients. Support vector machine (SVM) is supervised learning model with associated learning algorithms that analyze data and recognize patterns. In this work, Support vector machine is used for classifying liver disease using two liver patients datasets with different features combinations such as SGOT, SGPT and Alkaline Phosphates, evaluating a support vector machine classifier by measuring its performance based on: accuracy, error rate, sensitivity, prevalence and specificity. Results show that the accuracy, error rate, sensitivity and prevalence at first 6ordered features are the best for ILPD dataset compared to BUPA dataset. This can be attributed to a number of useful attributes like Total bilirubin, direct bilirubin, Albumin, Gender, Age and Total proteins are available in the ILPD liver dataset compared to the BUPA dataset which can help in diagnosis of liver disease.

Keywords Chronic Liver Disease (CLD) and Support Vector Machine (SVM)

随着酒精摄入过量，有害气体的吸入，受污染的食物，腌菜和药物的摄入，肝病患者不断增加。肝脏具有许多基本功能，肝脏疾病对于提供医疗护理存在一些问题。慢性肝病（CLD）是发达国家和发展中国家的常见长期疾病。分类技术在各种自动医疗诊断工具中非常流行。癌症的早期鉴定对于患者的存活通常是至关重要的。支持向量机（SVM）是具有相关联的学习算法的监督学习模型，其分析数据并识别模式。在这项工作中，支持向量机用于使用具有不同特征组合（例如SGOT，SGPT和碱性磷酸盐）的两个肝脏患者数据集来分类肝脏疾病，通过基于以下来评估支持向量机分类器的性能：准确度，错误率，灵敏度，流行和特异性。结果表明，与BUPA数据集相比，ILPD数据集的准确性，错误率，敏感性和流行率是最好的。这可以归因于许多有用的属性，如总胆红素，直接胆红素，白蛋白，性别，年龄和总蛋白可用于ILPD肝数据集与可以帮助诊断肝脏疾病的BUPA数据集相比。

关键词慢性肝病（CLD）和支持向量机（SVM）

1. Introduction

The liver is an essential body organ that forms an important barrier between the gastrointestinal blood, which contains large amounts of toxins and antigens in the body. The liver produces a large amount of hormones, enzymes, and performs several functions essential to life. It is also the organ responsible for cleansing of toxins from the bloodstream, by turning them into removable substances. Liver disease refers to many diseases and disorders that may cause impaired liver function that can make liver decrease of its functions. The dysfunction may be primary, but the liver is often secondarily affected by disorders of other organ systems, since it is involved in many metabolic and detoxifying processes.

Hepatic ﬁbrosis and its end stage cirrhosis are an increasing worldwide concern. Cirrhosis is the irreversible end result of ﬁbrosis scarring and normal hepatic architecture is replaced by interconnecting bands of ﬁbrosis tissue. The most common etiological factors resulting in cirrhosis are hepatitis B, hepatitis C, and excessive alcohol consumption [1].

Chronic HCV infection is normally a slow, progressive disease that may produce few or no symptoms for many years after infection. Some patients develop chronic infection and suffer no signiﬁcant liver damage, while others progress quickly to liver cirrhosis and may develop hepatocellular carcinoma[2]. Patients with chronic liver diseases belong to a high-risk group for hepatocellular carcinoma and should be followed up regularly for early diagnosis.

Chronic HCV infection is the major cause of cirrhosis and hepatocellular carcinoma (HCC). In this condition, alpha fetoprotein levels may be elevated. The incidence of hepatocellular carcinoma is rising, and this trend is expected to continue for years[3]. Figure 1, show that liver cancer is the most cause of death in Egypt among other types of cancer.

According to the current studies, the majority of HCC patients contracted the disease from the accumulation of genetic abnormalities, probably induced by exterior etiological factors especially HBV and HCV infections[4]. These risk factors can induce mutations and damage in DNA sequences, such as p53 mutation induced by aflatoxin and DNA damage induced by the intrusion of the HBV genome[5]. The important thing in preventing liver cancer is to prevent hepatitis virus infection and eliminate hepatitis virus in chronic hepatitis patients.

1.介绍

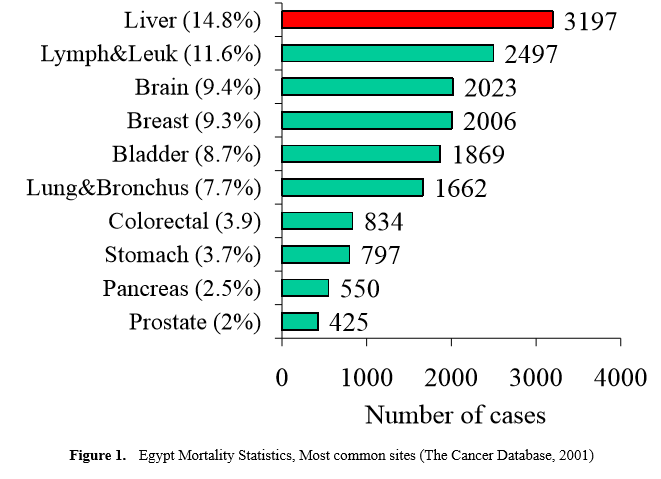
肝脏是在胃肠血液之间形成重要屏障的必需身体器官，胃肠血液在体内含有大量的毒素和抗原。肝产生大量的激素，酶，并且执行生命所必需的几种功能。它也是负责从血流中清除毒素，通过将其转化为可移除物质的器官。肝病是指可能引起肝功能受损的许多疾病和病症，其可以使肝功能降低。功能障碍可能是原发性的，但肝脏常常受其他器官系统的疾病的二次影响，因为它涉及许多代谢和解毒过程。

肝纤维化及其终末期肝硬化是全世界日益关注的问题。肝硬化是纤维化瘢痕形成的不可逆的最终结果，正常肝脏结构被纤维组织的互连带替代。导致肝硬化的最常见的病因是乙型肝炎，丙型肝炎和过度饮酒[1]。

慢性HCV感染通常是一种缓慢的进行性疾病，在感染后许多年可能产生很少或没有症状。一些患者发展为慢性感染并且没有显着的肝损伤，而另一些患者很快进展到肝硬化并可能发展成肝细胞癌[2]。慢性肝病患者属于高危组肝细胞癌，应定期随访以进行早期诊断。

慢性HCV感染是肝硬化和肝细胞癌（HCC）的主要原因。在这种情况下，甲胎蛋白水平可能升高。肝细胞癌的发病率正在上升，这种趋势预计将持续多年[3]。图1显示，在其他类型的癌症中，肝癌是埃及死亡的最主要原因。

根据目前的研究，大多数HCC患者从遗传异常的累积中收缩该疾病，可能由外部病因，特别是HBV和HCV感染诱导[4]。这些危险因素可能在DNA序列中诱导突变和损伤，如由黄曲霉毒素诱导的p53突变和HBV基因组侵入诱导的DNA损伤[5]。在预防肝癌中重要的是预防肝炎病毒感染和消除慢性肝炎患者中的肝炎病毒。



Automatic classification tools may reduce the burden on doctors. Data classification is a two phase process in which first step is the training phase where the classifier algorithm builds a classifier with the training set of dataset the second phase is classification phase where the model is used for classification and its performance is analyzed with the testing set of datasets[6].

Existing feature selection methods broadly fall into two categories, filter methods and wrapper methods. Filter methods select features based on some discriminate criteria that rely on the characteristics of data and are independent of any classification algorithms[7]. Wrapper methods use the predictive accuracy of predetermined classification algorithms as the criteria to determine the goodness of a subset of features[8, 9].

Most wrapper methods adopt sophisticated multivariate machine learning tools such as SVMs that take the combinatorial effects of features into account. These have been shown in many experiments to be more powerful in terms of classification accuracy than the filter methods[10].

Support Vector Machines proved to be effective for a lot of classifications problems. For binary-class classification, SVM constructs an optimal separating hyper plane between the positive and negative classes with the maximal margin. It can be formulated as a quadratic programming problem involving inequality constraints[11, 12].

SVMs are one of the most promising machine learning algorithms and there are many examples, where SVMs are used successfully, e.g. text classification, face recognition, and Bioinformatics. On these data sets SVMs perform very well and often outperform other traditional techniques[13]. SVMs have gained an enormous popularity in statistics, learning theory, and engineering[14, 15], and the many references therein. With a few exceptions most support vector learning algorithms have been designed for binary problems. A few attempts have been made to generalize SVM to multiclass problems[16].

In this work, a support vector machine is used as a classification algorithm in order to compare its performance with different features combinations such as aspartate aminotransferase (SGOT), glutamic pyruvic transaminase (SGPT), andalkaline phosphatase (Alkphos) using two datasets. The first one is BUPA Liver Disorders datasets taken from the University of California at Irvine (UCI) Machine Learning Repository, and the second one is from ILPD (Indian Liver Patient Dataset), it was collected from north east of Andhra Pradesh, India.

自动分类工具可以减轻医生的负担。数据分类是两阶段过程，其中第一步是训练阶段，其中分类器算法利用数据集的训练集构建分类器，第二阶段是分类阶段，其中模型用于分类，并且其性能用测试集分析的数据集[6]。

现有的特征选择方法大致分为两类，过滤方法和包装方法。过滤方法基于一些依赖于数据特征的区分标准来选择特征，并且独立于任何分类算法[7]。包装方法使用预定分类算法的预测精度作为确定特征子集的良好性的标准[8,9]。

大多数包装器方法采用复杂的多变量机器学习工具，例如考虑特征的组合效应的SVM。这些已经在许多实验中显示出在分类精度方面比滤波方法更强大[10]。

支持向量机证明对许多分类问题有效。对于二分类分类，SVM在最大边界的正和负类之间构建最佳分离超平面。它可以被制定为涉及不等式约束的二次规划问题[11,12]。

SVM是最有前途的机器学习算法之一，并且存在许多示例，其中SVM被成功地使用，例如。文本分类，面部识别和生物信息学。对这些数据集，SVM执行得非常好，并且经常胜过其他传统技术[13]。 SVM在统计学习，学习理论和工程[14,15]以及其中的许多参考文献中获得了巨大的人气。除了几个例外，大多数支持向量学习算法已经被设计用于二进制问题。已经做了一些尝试将SVM推广到多类问题[16]。

在这项工作中，支持向量机用作分类算法，以便使用两个数据集比较其性能与不同特征组合，如天冬氨酸转氨酶（SGOT），谷氨酸丙酮酸转氨酶（SGPT）和碱性磷酸酶（Alkphos）。第一个是从加利福尼亚大学尔湾（UCI）机器学习库获取的BUPA肝脏疾病数据集，第二个是来自ILPD（印度肝病患者数据集），它是从印度安得拉邦东北部收集的。

2. Materials and Methods

The liver is one of the major targets for insulin and its count regulatory hormones, such as glucagon. HCC patients who abuse alcohol are more likely to develop cirrhosis than those who do not. The most common cause of liver disease is non-alcoholic fatty liver disease. Cirrhosis is the end-result of many liver conditions and involves severe scarring of the liver. It is associated with a progressive decline in liver function resulting in liver failure. Hepatocellular carcinoma is the most common primary cancer of the liver. There are factors that may impact progression include age, gender, chronic alcohol abuse, and quantity of virus of exposure. The disease appears to be more aggressive in patients that acquire HCV after age 40 and may be more progressive in men than women[2]. In this paper SVM classification algorithm has been applied to: BUPA liver disorders dataset and Indian Liver Patient Dataset for evaluating SVM performance with different features.

2。材料和方法

肝脏是胰岛素及其计数调节激素（如胰高血糖素）的主要靶标之一。 酗酒的HCC患者比没有酗酒的HCC患者更容易发展成肝硬化。 肝病最常见的原因是非酒精性脂肪性肝病。 肝硬化是许多肝脏病症的结果，并且包括肝脏的严重瘢痕形成。 它与肝功能的逐渐下降导致肝衰竭相关。 肝细胞癌是最常见的肝原发性癌症。 有可能影响进展的因素包括年龄，性别，慢性酒精滥用和暴露病毒的数量。 该疾病在40岁后获得HCV的患者似乎更具侵袭性，并且在男性中可能比女性更加进展[2]。 在本文中，SVM分类算法已应用于：BUPA肝脏病症数据集和印度肝病人数据集，用于评估具有不同特征的SVM性能。

3. Dataset

3.1. BUPA Liver Disorders

BUPA liver disordershas 6 numeric Attributes, 345 Instances. Relevant information: The first 5 variables are all blood tests which are thought to be sensitive to liver disorders that might arise fromexcessive alcohol consumption, each line in the BUPA. Data file constitutes the record of a single male individual. It appears that drinks>5 is some sort of a selector on this dataset. University of California at Irvine (UCI) machine learning repository (WWW.UCI.Com).

3.2. Indian Liver Patient Dataset (ILPD)

Indian Liver Patient Dataset (ILPD) has 9 attribute, 483 Instances. The data set was collected from north east of Andhra Pradesh, India. Selector is a class label used to divide into groups (liver patient or not). This data set contains 441 male patient records and 142 female patient records .this data downloaded from ([WWW.UCI.Com](http://WWW.UCI.Com)).

3.数据集

3.1。 BUPA肝病

BUPA肝脏疾病6个数字属性，345个实例。 相关信息：前5个变量是所有血液检查，被认为对肝脏疾病敏感，可能由于过度饮酒，BUPA中的每一行。 数据文件构成单个男性个体的记录。 看来饮酒> 5是这个数据集上的某种选择器。 加州大学尔湾分校（UCI）机器学习库（WWW.UCI.Com）。

3.2。 印度肝病患者数据集（ILPD）

印度肝病患者数据集（ILPD）有9个属性，483个实例。 数据集是从印度安得拉邦东北部收集的。 选择器是用于分成组（肝病人或不是肝病人）的类标签。 该数据集包含441个男性患者记录和142个女性患者记录。从（WWW.UCI.Com）下载的该数据。

4. Support Vector Machine (SVM)

Classification algorithms are widely used in various medical applications. Classification aims to build an effective model for predicting class labels of unknown data. The model is built on the training data, which consists of data points chosen from input data space and their class labels. A Support Vector Machine (SVM) separates the data into two categories of performing classification and constructing an N-dimensional hyper plane. These models are closely related to classical multilayer perceptron neural networks.

A support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space. A good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

There are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, the unconstrained minimization problem as in standard neural network training[17], There are many possible kernel functions and the most common kernel are: Linear, polynomial, sigmoid and radial basis function (RBF). In this paper we use linear kernel function shows in equation .1:

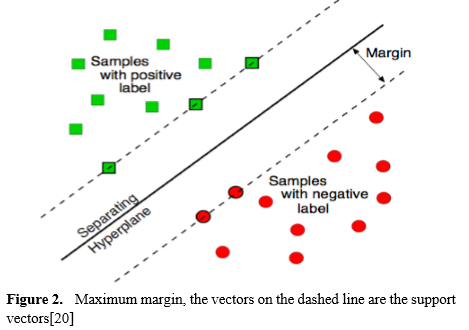


Depending on the kernel type we choose the kernel parameters have to be set. Which kernel type performs best, depends on the application and can be determined by using cross-validation.

In the SVM literature, a predictor variable which is called an attribute and a transformed attribute that is used to define the hyper plane is called a feature[18]. Here, choosing the most suitable representation can be taken as feature selection. A set of features that describes one case is called a vector. The goal of this modeling is to find the optimal hyperplane which separates clusters of vector in such a way those cases with one category of the targetVariable are on one side of the plane and cases with the other category are on the other side of the plane. The vectors near to the hyper plane are the support vectors[19] as in figure 2.

根据内核类型，我们选择内核参数必须设置。 哪种内核类型执行得最好，取决于应用程序，可以通过使用交叉验证来确定。

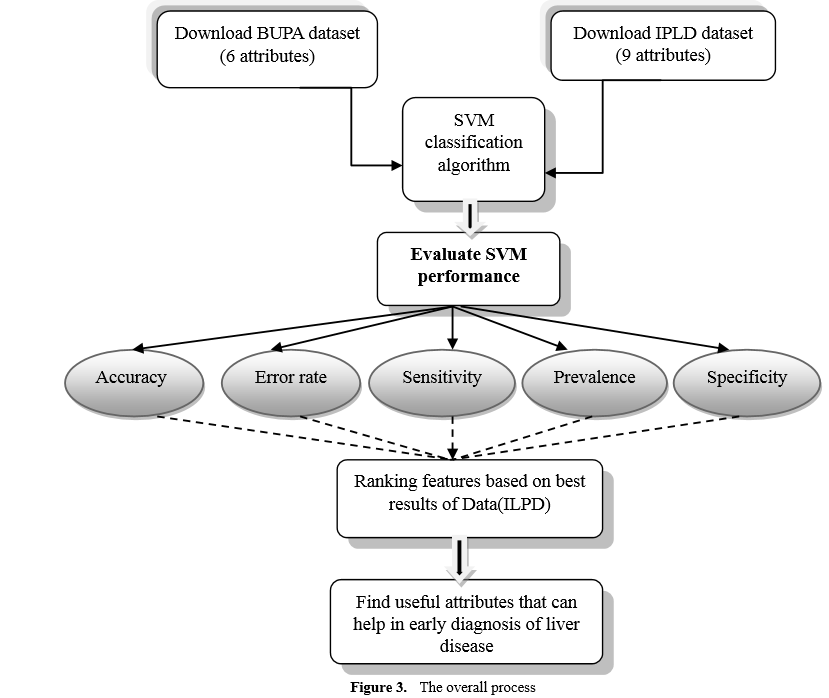
在SVM文献中，称为属性的预测变量和用于定义超平面的变换属性称为特征[18]。 在这里，选择最适合的表示可以作为特征选择。 描述一种情况的一组特征称为矢量。 这种建模的目的是找到以这样的方式分离矢量的聚类的最佳超平面，即具有targetVariable的一个类别的情况在平面的一侧，而另一类别的情况在平面的另一侧。 靠近超平面的向量是支持向量[19]，如图2所示。



5. Evaluation

To evaluate performance of SVM, accuracy, error rate, sensitivity, prevalence and specificity are calculated for each dataset. They are defined as follows:

Feature ranking is useful to gain knowledge of data and identify relevant features, also it helps reduce the number of features a learner has to examine and reduce errors from irrelevant features. Figure 3 summarize overall the process of this work.

6. Implementation

SVM is a new promising non-linear, non-parametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields. The SVM algorithm is written and implemented using MATLAB and it is also tested on the WINDOWS platform using MATLAB version 7.0 with its powerful Bioinformatics and statistics for machine learning Toolbox. The Accuracy, Prevalence, Sensitivity, Error rate and Specificity are calculated for classifying BUPA liver patient’s dataset and ILPD Liver dataset using support vector machine classification algorithm. The features are ranked based on priority using the ranking algorithm available in MATLAB for each dataset

6.实施

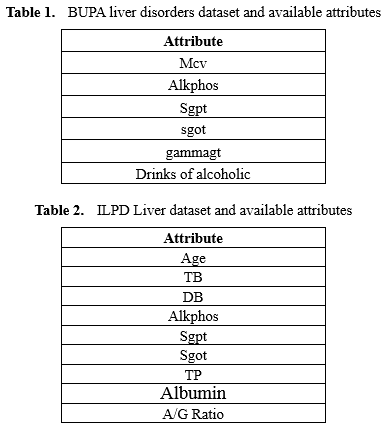
SVM是一种新的有前途的非线性，非参数分类技术，在医学诊断，光学字符识别，电负荷预测等领域已经显示出良好的效果。 SVM算法是使用MATLAB编写和实现的，它还在WINDOWS平台上使用MATLAB版本7.0及其强大的生物信息学和统计学机器学习工具箱进行测试。 使用支持向量机分类算法对BUPA肝病人数据集和ILPD肝数据集进行分类，计算准确度，流行率，灵敏度，错误率和特异性。使用MATLAB中可用的排序算法为每个数据集对特征进行优先级排序

7. Result and Discussion

Liver tumors are divided into two categories: benign and malignant. HCC is a malignant tumor derived from hepatocytes that belong to primary malignant epithelial tumors of the liver. An early diagnosis of liver problems will increase the patient’s survival rate. Performance of Support vector machine classification algorithm is tested and evaluated using two datasets (BUPA liver disorders, ILPD Indian Liver Dataset) as shown in table 1, and table 2. Table 1show the attributes of BUPA liver disordersand table 2 show the attributes of ILPD IndianLiver Patient datasets.

7.结果与讨论

肝肿瘤分为两类：良性和恶性。 HCC是衍生自属于肝脏原发性恶性上皮肿瘤的肝细胞的恶性肿瘤。 早期诊断肝脏问题会增加患者的生存率。 支持向量机分类算法的性能使用如表1和表2所示的两个数据集（BUPA肝脏疾病，ILPD印度肝数据集）进行测试和评估。表1显示了BUPA肝脏疾病的属性，表2显示了ILPD印度肝 患者数据集。



Previous experimentations motivates us to use cross validation in this analysis with SVM by randomly divide data to 50/50 training set equal to testing set. The error rate, specificity, accuracy, prevalence and sensitivity are calculated for BUPA Liver dataset as shown in table3 and the ILPD dataset as shown in table 4.

The performance of SVM Classification Algorithm is analyzed with BUPA and ILPD datasets, the Specificity at first 6 ordered features are best for BUPA dataset compared to other dataset. The Sensitivity, Error rate, Accuracy and Prevalence at first 6 ordered features are best for ILPD Liver dataset compared to BUPA dataset.

The features are ranked based on priority for each dataset. Table 5 shows ordering of the attributes of BUPA liver disorders which are: aspartate aminotransferase (SGOT), gamma-glutamyl transpeptidase (gammagt), mean corpuscular volume (mcv), alkaline phosphotase (alkphos), alamine aminotransferase (SGPT) and drinks. Ordering of the attributes of ILPD datasets as: A/G Ratio Albumin and Globulin Ratio, Direct Bilirubin (DB), Total Bilirubin (TB), Alkphos, SGPT, SGOT, Albumin, Age and Total Proteins (TP) given in table.6

Poor results with BUPA dataset can be attributed to the limited number of samples compared to ILPD dataset, However, We have taken common attributes (SGOT, SGPT, Alkphos) of both BUPA and ILPD datasets and implemented the experimentation.

SGOT aspartate aminotransferase or (AST) test is part of an initial screening for liver disease. AST is normally found in red blood cells, liver, heart, muscle tissue, pancreas, and kidneys. AST formerly was called serum glutamic oxaloacetictransaminase (SGOT). this test is done to check for liver damage and Help identify liver disease, especially hepatitis and cirrhosis.[21]

SGPT isan alanine aminotransferase (ALT) test measures the amount of this enzyme in theblood[21]. ALT is found mainly in the liver, ALT was formerly called serum glutamic pyruvic transaminase (SGPT). ALT is measured to see if the liver is damaged or diseased.

Alkphos is An alkaline phosphatase (ALP) test measures the amount of the enzyme ALP in the blood. ALP is made mostly in the live and used to help detect liver disease or bone disorders.[22]

Table 4 shows the observations with ILPD dataset, observed parameters were very good which indicates that these three common features are important for (1) detect the presence of liver disease, (2) distinguish among different types of liver disorders, (3) gauge the extent of known liver damage, and (4) follow the response to treatment

先前的实验激励我们在该分析中使用与SVM的交叉验证，通过将数据随机划分为等于测试集合的50/50训练集合。计算BUPA肝脏数据集的错误率，特异性，准确性，流行率和灵敏度，如表3所示，ILPD数据集如表4所示。

使用BUPA和ILPD数据集分析SVM分类算法的性能，与其他数据集相比，前6个有序特征的特异性最适合BUPA数据集。与BUPA数据集相比，ILPD肝数据集的最初6个有序特征的灵敏度，误差率，准确度和流行性最好。

基于每个数据集的优先级对功能进行排名。表5显示了BUPA肝脏疾病的属性的排序：天冬氨酸转氨酶（SGOT），γ-谷氨酰转肽酶（gammagt），平均红细胞体积（mcv），碱性磷酸酶（alkphos），胺酶转氨酶（SGPT）和饮料。 ILPD数据集的属性排序为：表6中给出了A / G比白蛋白和球蛋白比，直接胆红素（DB），总胆红素（TB），Alkphos，SGPT，SGOT，白蛋白，年龄和总蛋白

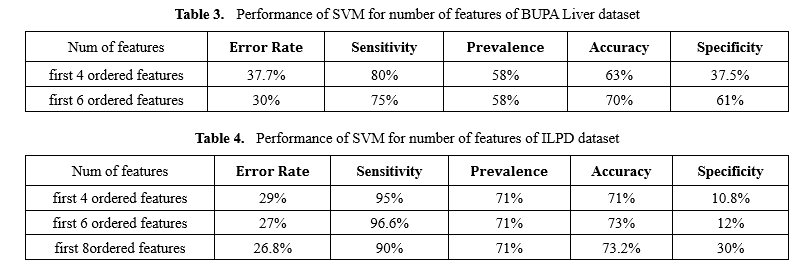
与ILUP数据集相比，BUPA数据集的差的结果可归因于有限数量的样本。然而，我们已经采用了BUPA和ILPD数据集的共同属性（SGOT，SGPT，Alkphos）并实施了实验。

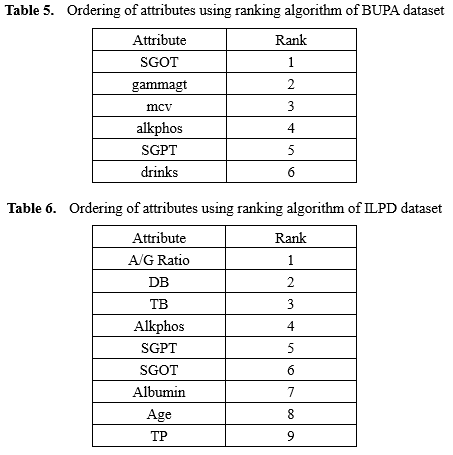
SGOT天冬氨酸氨基转移酶或（AST）测试是肝脏疾病的初始筛选的一部分。 AST通常在红细胞，肝，心脏，肌肉组织，胰腺和肾中发现。 AST以前称为血清谷草转氨酶（SGOT）。这个测试是为了检查肝脏损伤和帮助识别肝脏疾病，特别是肝炎和肝硬化。[21]

SGPT isan丙氨酸氨基转移酶（ALT）测试测量血液中这种酶的量[21]。 ALT主要在肝脏中发现，ALT以前称为血清谷氨酸丙酮酸转氨酶（SGPT）。测量ALT以观察肝脏是否损伤或患病。

Alkphos是一种碱性磷酸酶（ALP）测试，测量血液中ALP酶的量。 ALP主要在活体，并用于帮助检测肝脏疾病或骨疾病。[22]

表4显示了使用ILPD数据集的观察结果，观察到的参数非常好，这表明这三个共同特征对于（1）检测肝脏疾病的存在，（2）区分不同类型的肝脏疾病，（3）已知肝损伤的程度，以及（4）跟随对治疗的反应





8. Conclusions

Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. An important task in cancer research is to separate healthy patients from cancer patients and to distinguish patients of different cancer subtypes, based on their cytogenetic profiles. This is known as the classification problem. These tasks help successful cancer diagnosis and treatment. Machine learning is about designing algorithms that allow a computer to learn. Support vector machine has proved to be a powerful approach for classifier design. It has become an increasingly popular tool for machine learning tasks involving classiﬁcation, regression or novelty detection. The key idea of SVM is to find a hyper plane that maximizes the margin separating the two classes of instances. In this study, the SVM classification algorithm were considered in evaluating its classification performance in terms of Accuracy, Prevalence, Sensitivity, Error rate and Specificity in classifying BUPA liver patient dataset and ILPDIndian Liver dataset.

The Specificity at first 6 ordered features are best for BUPA dataset compared to ILPD dataset. The Sensitivity, Error rate, Accuracy and Prevalence at first 6 ordered features are best for ILPD Liver dataset compared to BUPA dataset, so the three common attributes (SGOT, SGPT, Alkphos) are important in diagnosis of liver diseases. This conclusion can be attributed to more number of useful attributes like Total bilirubin, direct bilirubin, Albumin, Gender, Age and Total proteins are available in the ILPD liver dataset compared to the BUPA dataset. So as to increase the number of features it improves the performance in classification algorithm that can help in early diagnosis and treatment of liver cancer.

8.结论

肝脏患者的问题不容易在早期阶段发现，因为它将正常运行，即使它被部分损坏。癌症研究中的一个重要任务是将健康患者与癌症患者分离，并基于其细胞遗传学概况区分不同癌症亚型的患者。这被称为分类问题。这些任务有助于成功的癌症诊断和治疗。机器学习是关于设计允许计算机学习的算法。支持向量机已经证明是分类器设计的强大方法。它已经成为一种越来越受欢迎的工具，用于机器学习任务涉及分类，回归或新奇检测。 SVM的关键思想是找到一个超平面，最大化分隔两类实例的边距。在本研究中，在分类BUPA肝脏患者数据集和ILPDIndian肝数据集时，在准确度，流行率，灵敏度，错误率和特异性方面评估其分类性能时考虑SVM分类算法。

特定性在前6个有序特征是最好的BUPA数据集相比，ILPD数据集。与BUPA数据集相比，ILPD肝数据集的前6个有序特征的灵敏度，误差率，准确度和流行性最好，因此三个常见属性（SGOT，SGPT，Alkphos）在肝脏疾病诊断中很重要。这个结论可归因于更多数量的有用属性，如总胆红素，直接胆红素，白蛋白，性别，年龄和总蛋白可用于ILPD肝数据集与BUPA数据集相比。为了增加特征的数量，它提高了分类算法的性能，可以帮助肝癌的早期诊断和治疗。

2015 Liver Disease Prediction using SVM and Naïve Bayes Algorithms

Abstract— In recent years in healthcare sectors, data mining became an ease of use for disease prediction. Data mining is the process of dredge up information from the massive datasets or warehouse or other repositories. It is a very challenging task to the researchers to predict the diseases from the voluminous medical databases. To overcome this issue the researchers use data mining techniques such as classification, clustering, association rules and so on. The main objective of this research work is to predict liver diseases using classification algorithms. The algorithms used in this work are Naïve Bayes and support vector machine (SVM). These classifier algorithms are compared based on the performance factors i.e. classification accuracy and execution time. From the experimental results it is observed that the SVM is a better classifier for predict the liver diseases.

Keywords - Classification, Liver function test, Naïve bayes, SVM

摘要 - 近年来在医疗保健领域，数据挖掘变得易于用于疾病预测。 数据挖掘是从海量数据集或仓库或其他存储库挖掘信息的过程。 研究人员从庞大的医学数据库中预测疾病是一项非常具有挑战性的任务。 为了克服这个问题，研究人员使用数据挖掘技术，如分类，聚类，关联规则等。 这项研究工作的主要目的是使用分类算法预测肝脏疾病。 在这项工作中使用的算法是朴素贝叶斯和支持向量机（SVM）。 基于性能因素（即分类精度和执行时间）来比较这些分类器算法。 从实验结果，观察到SVM是一个更好的分类器，用于预测肝脏疾病。

关键词：分类，肝功能测试，Naïvebayes，SVM

I. INTRODUCTION

Researchers faces more challenging task in healthcare sectors to predict the diseases from the voluminous medical databases. Nowadays data mining became more essential in healthcare sectors. Data mining techniques which includes classification, clustering, association rule mining for finding frequent patterns are applied to medical data for disease prediction. In data mining, classification techniques are much popular in medical diagnosis and predicting diseases [1]. In this research work, Naïve Bayes and Support Vector Machine (SVM) classifier algorithms are used for liver disease prediction. There are several numbers of liver disorders that required clinical care of the physician [3]. The main objective of this research work is to predict liver diseases such as Cirrhosis, Bile Duct, Chronic Hepatitis, Liver Cancer and Acute Hepatitis from Liver Function Test (LFT) dataset using above classification algorithms.

The liver is the second largest internal organ in the human body, playing a major role in metabolism and serving several vital functions, e.g. Decomposition of red blood cells, etc.,. [7] Its weight comes around three pounds. The liver performs many essential functions related to digestion, metabolism, immunity, and the storage of nutrients within the body. These functions make the liver as an important organ, without this, body tissues would quickly die from lack of energy and nutrients. There are number of factors which increase the risk of liver disease. Some of them are listed below:

• Family history of liver disease

• Smoking

• Consumption of alcohol

• Intake of contaminated food

• Obesity

• Diabetes

The remaining portion of the paper is organized as follows. Related works are discussed in Section 2. The proposed methodology is given in Section 3. Section 4 analyzes the experimental results. Section 5 gives conclusion..

I.引言

研究人员在医疗保健行业面临更多的挑战性任务，以从庞大的医疗数据库预测疾病。如今，数据挖掘在医疗保健领域变得更加重要。包括分类，聚类，用于发现频繁模式的关联规则挖掘的数据挖掘技术被应用于用于疾病预测的医学数据。在数据挖掘中，分类技术在医学诊断和预测疾病中非常流行[1]。在这项研究工作中，朴素贝叶斯和支持向量机（SVM）分类器算法用于肝脏疾病预测。有几种肝脏疾病需要医生的临床护理[3]。这项研究的主要目的是使用上述分类算法从肝功能测试（LFT）数据集预测肝脏疾病，如肝硬化，胆管，慢性肝炎，肝癌和急性肝炎。

肝是人体中第二大的内部器官，在新陈代谢中起主要作用并且服务于几种重要的功能，例如，红细胞分解等。 [7]它的重量约三磅。肝脏执行许多与消化，代谢，免疫和体内营养物质的储存相关的基本功能。这些功能使肝脏作为重要的器官，没有这个，身体组织会迅速死于缺乏能量和营养。有许多因素增加肝脏疾病的风险。其中一些列举如下：

•肝病家族史

• 抽烟

•饮酒

•摄入受污染的食物

•肥胖

•糖尿病

本文的剩余部分组织如下。相关工作在第2节中讨论。所提出的方法在第3节中给出。第4节分析实验结果。第5节给出结论。

II. LITERATURE REVIEW

Dhamodharan et.al [3] has predicted three major liver diseases such as Liver cancer, Cirrhosis and Hepatitis with the help of distinct symptoms. They used Naïve Bayes and FT Tree algorithms for disease prediction. Comparison of these two algorithms has been done based on their classification accuracy measure. From the experimental results they concluded the Naïve bayes as the better algorithm which predicted diseases with maximum classification accuracy than the other algorithm.

Rosalina et al [13] predicted a hepatitis prognosis disease using Support Vector machine (SVM) and Wrapper Method. Before classification process they used wrapper methods to remove the noise features. Firstly SVM carried out feature selection to get better accuracy. Features selection were implemented to minimize noisy or irrelevance data. From the experimental results they observed the increased accuracy rate in the clinical lab test cost with minimum execution time. They have achieved the target by combining Wrappers Method and SVM techniques. Omar S.

Soliman et al [10] has proposed a hybrid classification system for HCV diagnosis, using Modified Particle Swarm Optimization algorithm and Least Squares Support Vector Machine (LS-SVM). Feature vectors are extracted using Principle Component Analysis algorithm. As LS-SVM algorithm is sensitive to the changes of values of its parameters, Modified-PSO Algorithm was used to search for the optimal values of LS-SVM parameters in less number of iterations. The proposed system was implemented and evaluated on the benchmark HCV data set from UCI repository of machine learning databases. It was compared with another classification system, which utilized PCA and LS-SVM. From the experimental results the proposed system obtained maximum classification accuracy than the other systems.

Karthik et.al [7] were applied a soft computing technique for intelligent diagnosis of liver disease. They have implemented classification and its type detection in three phases. In the first phase, they classified liver disease using Artificial Neural Network (ANN) classification algorithm In the second phase, they generated the classification rules with rough set rule induction using Learn by Example (LEM) algorithm. In the third phase fuzzy rules were applied to identify the types of the liver disease.

Chaitrali S. Dangare et.al [2] has analyzed prediction systems for Heart disease using more number of input attributes. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database. The performances of these techniques are compared, based on accuracy. Authors’ analysis shows that out of these three classification models Neural Networks has predicted the heart disease with highest accuracy.

II 文献评论

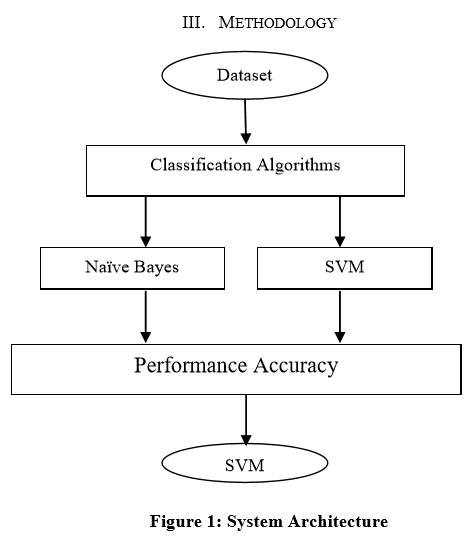
Dhamodharan等人[3]已经预测了三种主要的肝脏疾病，如肝癌，肝硬化和肝炎在不同的症状的帮助。他们使用NaïveBayes和FT Tree算法进行疾病预测。这两种算法的比较是基于它们的分类精度测量。从实验结果，他们认为Naive Bayes是更好的算法，预测疾病与最大分类精度比其他算法。

Rosalina等[13]使用支持向量机（SVM）和包装方法预测了肝炎预后疾病。在分类过程之前，他们使用包装方法来消除噪声特征。首先SVM进行特征选择以获得更好的精度。实施特征选择以最小化噪声或不相关数据。从实验结果，他们观察到在临床实验室测试成本增加的准确率增加最小执行时间。他们通过结合包装器方法和SVM技术实现了目标。奥马尔

Soliman等人[10]提出了一种用于HCV诊断的混合分类系统，其使用改进的粒子群优化算法和最小二乘支持向量机（LS-SVM）。使用主成分分析算法提取特征向量。因为LS-SVM算法对其参数值的变化敏感，所以使用Modified-PSO算法在较少次数的迭代中搜索LS-SVM参数的最佳值。对来自UCI机器学习数据库的基准HCV数据集进行了所提出的系统的实施和评估。将其与另一个使用PCA和LS-SVM的分类系统进行比较。从实验结果提出的系统获得最大分类精度比其他系统。

Karthik等[7]应用软件计算技术进行肝脏疾病的智能诊断。他们分三个阶段实施分类及其类型检测。在第一阶段，他们使用人工神经网络（ANN）分类算法分类肝脏疾病。在第二阶段，他们使用粗略集规则诱导使用学习示例（LEM）算法生成分类规则。在第三阶段，应用模糊规则来识别肝脏疾病的类型。

 Chaitrali S. Dangare等[2]使用更多的输入属性分析了心脏疾病的预测系统。在心脏病数据库上分析数据挖掘分类技术，即决策树，朴素贝叶斯和神经网络。基于准确性比较这些技术的性能。作者的分析表明，在这三个分类模型中，神经网络以最高的准确度预测了心脏病。



Dataset

Indian Liver Patient Dataset (ILPD) has been taken from the UCI Repository. This dataset has five hundred and seventy six instances and ten attributes. Attributes are Age, Gender, TB, DB, ALP, Sgpt, Sgot, TP, ALB and A/G Ratio. This dataset contains Liver Function Test details (LFT).

Naïve Bayes A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independent assumption. A more descriptive term for the underlying probability model would be the self-determining feature model. In basic terms, a Naive Bayes classifier assumes that the presence of a particular feature of a class is unrelated to the presence of any other feature [11]. The Naive Bayes classifier performs reasonably well even if the underlying assumption is not true. The advantage of the Naive Bayes classifier is that it only requires a small amount of training data to estimate the means and variances of the variables necessary for classification. Because of independent variables are unspecified, only the variances of the variables for each label need to be determined and not the entire covariance matrix. In contrast to the Naive Bayes operator, the Naive Bayes (Kernel) operator can be applied on numerical attributes. This can be able in a clear-cut fashion using kernel density estimation and Bayes' theorem:

This is essentially the same idea as discriminant analysis, only instead of assuming normality, were estimating the probability density of the classes using a nonparametric method Patrick Support Vector Machine

Support Vector Machine was first found by Vapnik in 1979 [5]. It was again recommended by Vapnik in 1995 for regression and classification [4]. Support vector can be used for pattern classification [8] which has multilayer perceptron and radial-basis function networks [12]. The SVM is the advanced technology with maximum classification algorithms embedded in statistical learning theory [9]. SVM methods are used in classification of linear and non-linear data. It transforms the original training data into higher dimension using non-linear mapping. Within this new dimension it searches for linear optimal separating hyperplane. Data from two classes can be separated by hyperplane with an appropriate nonlinear mapping to a sufficiently high dimension. Using support vectors and margins the SVM finds these hyperplane [6]. SVM implements the classification task by maximizing the margin classifies both class while minimizing the classification errors. Although the SVM can be applied to various optimization problems such as regression, the classic problem is that of data classification. The basic idea is shown in figure 2. The data points are identified as being positive or negative, and the problem is to find a hyperplane that separates the data points by a maximal margin

数据集

印度肝病患者数据集（ILPD）已从UCI存储库中获取。该数据集具有五百七十六个实例和十个属性。属性是年龄，性别，TB，DB，ALP，Sgpt，Sgot，TP，ALB和A / G比率。此数据集包含肝功能测试详细信息（LFT）。

Naive Bayes朴素贝叶斯分类器是一个基于应用具有强独立假设的贝叶斯定理的简单概率分类器。基本概率模型的更具描述性的术语将是自决定特征模型。在基本术语中，朴素贝叶斯分类器假定类的特定特征的存在与任何其他特征的存在无关[11]。朴素贝叶斯分类器相当好地执行，即使基本假设不是真的。朴素贝叶斯分类器的优点是，它只需要少量的训练数据来估计分类所需的变量的均值和方差。因为独立变量是未指定的，所以只需要确定每个标签的变量的方差，而不是整个协方差矩阵。与朴素贝叶斯算子相反，朴素贝叶斯（Kernel）算子可以应用于数值属性。这可以使用核密度估计和贝叶斯定理以清晰的方式：

这本质上是与判别分析相同的想法，只是不使用正态性，而是使用非参数方法估计类的概率密度Patrick支持向量机

支持向量机是由Vapnik在1979年首次发现[5]。它再次被Vapnik在1995年推荐用于回归和分类[4]。支持向量可以用于模式分类[8]，其具有多层感知器和径向基函数网络[12]。 SVM是在统计学习理论中嵌入最大分类算法的先进技术[9]。 SVM方法用于线性和非线性数据的分类。它使用非线性映射将原始训练数据转换为更高维度。在这个新的维度内，它搜索线性最优分离超平面。来自两个类的数据可以通过具有适当的非线性映射到足够高的维度的超平面来分离。使用支持向量和边距SVM找到这些超平面[6]。 SVM通过最大化边缘来实现分类任务，同时最小化分类误差。虽然SVM可以应用于各种优化问题，如回归，经典的问题是数据分类。基本思想如图2所示。数据点被标识为正或负，问题是找到一个超平面，它将数据点分开最大限度

IV. EXPERIMENTAL RESULTS

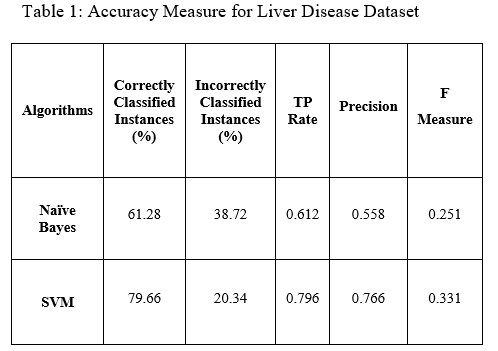
In this section, the results are analyzed which are given by the classification algorithms such as naïve Bayes and Support Vector Machine. This work is implemented in Matlab 2013 tool.

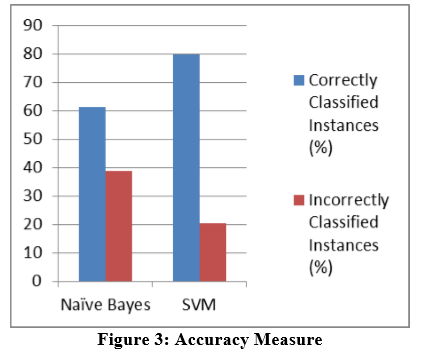
Figure 3 represents the accuracy measure for the Naïve Bayes and SVM classification algorithms. An experimental result shows the performance of SVM is better than Naïve Bayes algorithm.

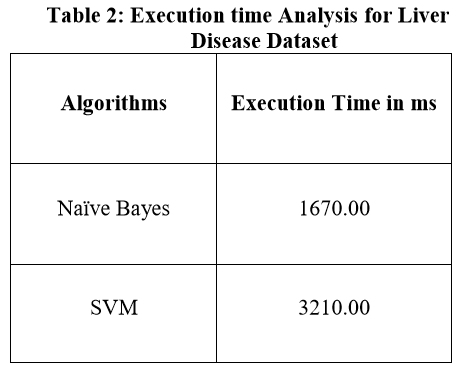
IV。 实验结果

在本节中，分析结果由分类算法给出，如天真贝叶斯和支持向量机。 这项工作在Matlab 2013工具中实现。

图3表示Naive Bayes和SVM分类算法的准确度测量。 实验结果表明SVM的性能优于NaïveBayes算法。







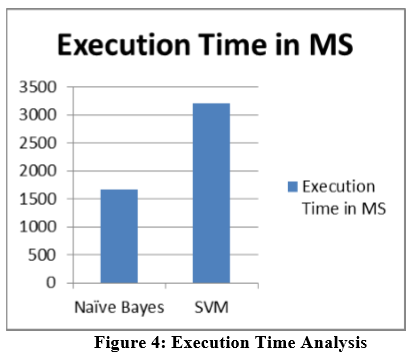
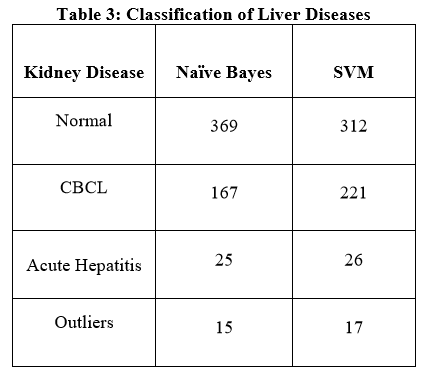


Table 2 represents the execution time requirement of classification algorithms for predicting liver diseases from liver function test dataset. Figure 4 represents the time taken for execution process. Naïve Bayes performs with minimum period of execution time than SVM.

表2表示用于从肝功能测试数据集预测肝脏疾病的分类算法的执行时间要求。 图4表示执行过程所花费的时间。 Naive Bayes执行时间比SVM最短。



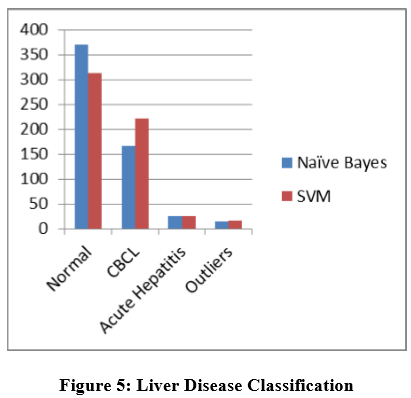


Table 3 represents and describes the classification of liver diseases such as cirrhosis, bile duct, chronic hepatitis, liver cancer and acute hepatitis. (CBCL) derives first four diseases. CBCL are the liver diseases affected when the liver function test data slightly increased with normal range. Acute hepatitis is a severe liver disease occurs when the liver function test data heavily increases than the normal range. Outliers are predicted in this work based on the moderate range of the liver function test results. Figure 5 represents the classification of liver diseases by classifiers such as Naïve Bayes and SVM. By analyzing the results, SVM gives the overall best classification result than Naïve Bayes classifier.

表3表示并描述了肝病如肝硬化，胆管，慢性肝炎，肝癌和急性肝炎的分类。 （CBCL）产生前四种疾病。 CBCL是当肝脏功能测试数据略有增加时与正常范围相关的肝脏疾病。 急性肝炎是一种严重的肝脏疾病发生时，肝功能测试数据大大增加超过正常范围。 在本工作中基于肝功能测试结果的中等范围预测离群值。 图5表示通过分类器如Naive Bayes和SVM对肝脏疾病的分类。 通过分析结果，SVM给出的总体最佳分类结果比朴素贝叶斯分类器。

V. CONCLUSION

Classification is the major data mining technique which is primarily used in healthcare sectors for medical diagnosis and predicting diseases. This research work used classification algorithms namely Naïve bayes and Support Vector Machine (SVM) for liver disease prediction. Comparisons of these algorithms are done and it is based on the performance factors classification accuracy and execution time. From the experimental results, this work concludes, the SVM classifier is considered as a best algorithm because of its highest classification accuracy. On the other hand, while comparing the execution time, the Naïve Bayes classifier needs minimum execution time.

五，结论

分类是主要用于医疗领域的医学诊断和预测疾病的主要数据挖掘技术。 本研究工作使用分类算法，即Naive Bayes和支持向量机（SVM）进行肝脏疾病预测。 这些算法的比较完成，它是基于性能因素分类的精度和执行时间。 从实验结果，这项工作的结束，SVM分类器被认为是一个最佳的算法，因为其最高的分类精度。 另一方面，在比较执行时间时，NaïveBayes分类器需要最小的执行时间。

A Novel Multiinstance Learning Approach for Liver Cancer Recognition on Abdominal CT Images Based on CPSO-SVM and IO

摘要