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

Lei Meng; Changyun Wen
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
Lmeng1@e.ntu.edu.sg; ecywen@ntu.edu.sg

Guoqi Li
Institute of high performance computing
A*Star
Singapore
gqli@ihpc.a-star.edu.sg

***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**

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.

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

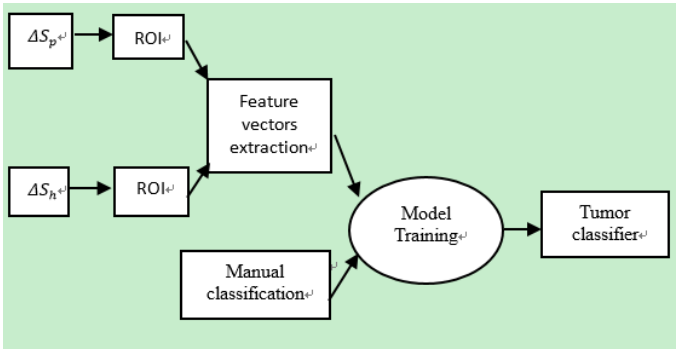

Fig. 1 Flow chart of classification model construction procedures

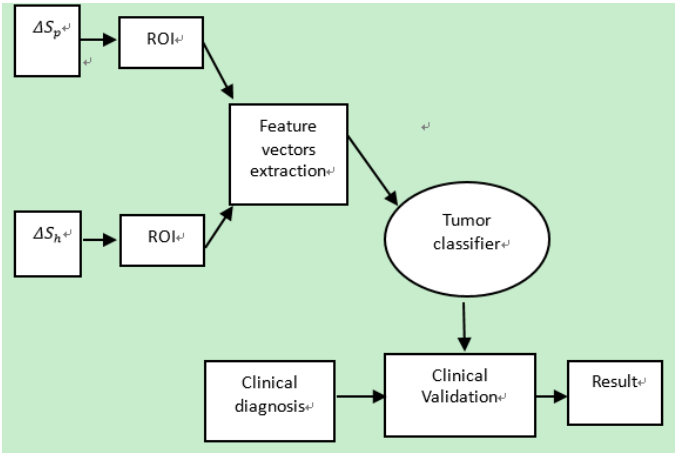

Fig. 2 the flow chart of the clinical model validation

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, S_0, is MR image without contrast agent. The other two images are captured under the effect of contrast agent. S_h, is MR image captured when the peak value of contrast agent appears in the hepatic artery. S_p 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 (S_p, and S_h) and MR images without contrast agent (S_0) are used as the input data of our system. The changes of the MR signal intensities ΔS_h and ΔS_p are calculated as follows:

$$\Delta S_h = (S_h - S_0)/S_0 \times 100 \quad (1)$$

$$\Delta S_p = (S_p - S_0)/S_0 \times 100 \quad (2)$$

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 ΔS_h or ΔS_p 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.


3a 3b 3c

Fig. 3 Sample MR images of early stage liver cancer with the ROI selected. Where 3a is the image of S_0, 3b is image for the S_h and 3c is the S_p image.

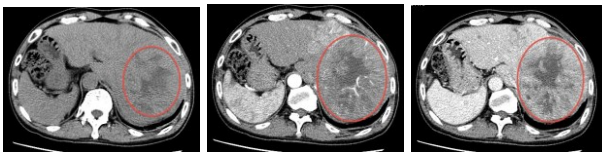
4a 4b 4c

Fig. 4 Sample MR images of late stage liver cancer with the ROI selected. Where 4a is the image of S_0, 4b is image for the S_h and 4c is the S_p image.

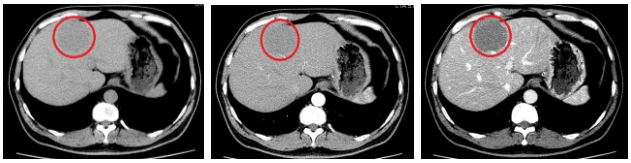
5a 5b 5c

Fig. 5 Sample MR images of BLT with the ROI selected. Where 5a is the image of S_0, 5b is image for the S_h and 5c is the S_p image.

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

$$K_{rbf} = \exp\left(-\gamma \times \|x_i, x_j\|^2\right)$$

where x_i and x_j 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.

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.

Table 1 Distribution of liver cancer patients

	Age below 40	*40-60*	*60-80*	*Above 80*	*Total*
Male	3	19	25	5	52
Female	2	17	26	3	48
Sum	5	36	51	8	100

Table 2 Distribution of BLT patients

	Age below 40	*40-60*	*60-80*	*Above 80*	*Total*
Male	9	21	17	4	51
Female	11	19	16	3	49
Sum	20	40	33	7	100

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.

Table 3 Distribution of 30 confirmed early stage liver cancer samples

	Age below 40	*40-60*	*60-80*	*Above 80*	*Total*
Male	4	7	4	2	17
Female	4	5	3	1	13
Sum	8	12	7	3	30

Table 4 Distribution of 30 BLT samples

	Age below 40	*40-60*	*60-80*	*Above 80*	*Total*
Male	7	2	4	5	18
Female	3	3	4	2	12
Sum	10	5	8	7	30

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:

$$f_k^* = \frac{f_k - u_{f_k}}{\sigma_{f_k}}$$

where f_k^* is the normalized feature vector, f_k is the histogram feature vector of each ROI sample, u_{f_k} and σ_{f_k} are the mean and standard deviation values of each element in the feature vector f_k.

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.

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 naked-eye 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 higher-level texture analysis features and other advanced classification techniques to improve the classification results of our system.

REFERENCES

[1]:Liver Cancer Information, http://www.disabled-world.com/health/cancer/liver

[2]: Jerry L. Prince, Jonathan M. Links, 2006, Medical Imaging Signals And Systems, Pearson Prentice Hall, ISBN 9780130653536

[3]: Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R. and Davatzikos, C. (2009), Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magnetic Resonance Medicine, 62: 1609–1618. doi: 10.1002/mrm.22147.

[4]:Poza, J., Hornero, R., Escudero, J., Fernández, A., & Sánchez, C. I. (2008). Regional Analysis of Spontaneous MEG Rhythms in Patients with Alzheimer's Disease Using Spectral Entropies. Annals of Biomedical Engineering, 36(1), 141-152.

[5]: Schröder, F. H., Hermanek, P., Denis, L., Fair, W. R., Gospodarowicz, M. K. and Pavone-Macaluso, M. (1992), The TNM classification of prostate cancer. Prostate, 21: 129–138.

[6]: Freiman, M. Edrei, Y. Sela, Y. Shmidmayer, Y. Shmidmayer, Y. Joskowicz, L. and Abramovitch, R. (2008), Classification of Suspected Liver Metastases Using fMRI Images: A Machine Learning Approach. 93-100. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008.

[7]: Sela Y, Freiman M., Dery E., Edrei Y., Safadi R., Pappo O., Joskowicz L. and Abramovitch R. "FMRI-based hierarchical SVM model for the classification and grading of liver fibrosis", IEEE Trans. Biomed. Eng., vol. 58, no. 9, pp.2574 -2581 2011

[8]: Freiman, M.; Edrei, Y.; Gross, E.; Joskowicz, L.; Abramovitch, R., "Liver metastasis early detection using fMRI

based statistical model," Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on , vol., no., pp.584,587, 14-17 May 2008
[9]: Vapnik V., Statistical learning theroy, New York Wiley, 1998
[10]: Lachenbruch PA. Discriminant Analysis. Hafner Press; 1975.
[11]: Dasarathy BV, editor. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. 1991.
[12]: Novakovic, J.; Veljovic, A., "C-Support Vector Classification: Selection of kernel and parameters in medical diagnosis," Intelligent Systems and Informatics (SISY), 2011 IEEE 9th International Symposium on , vol., no., pp.465,470, 8-10 Sept. 2011
[13]: Boser E., Guyon I., and Vapnik V. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pages 144-152. ACM Press, 1992.
[14]: Cortes and Vapnik V., Support-vector network. Machine Learning, 20:273-297, 1995.
[15]: Guoqi Li; Changyun Wen; Zheng Guo Li; Aimin Zhang; Feng Yang; Kezhi Mao, "Model-Based Online Learning With Kernels," Neural Networks and Learning Systems, IEEE Transactions on , vol.24, no.3, pp.356,369, March 2013