Identifying Morphological Markers for Differential Diagnosis of Liver Lesion

Francisco Gimenez, Ting Liu, Yi Liu
Biomedical Informatics, Stanford University

1. Introduction

Liver lesions stem from a variety of causes ranging from benign to malignant. The ability to efficiently and accurately differentiate these lesions as malignant or benign using non-invasive medical imaging is important to patient treatment and outcome. Contrast-enhanced computed tomography (CT) imaging is the dominant technology used for liver lesion diagnosis [1]. This modality takes advantage of the fact that the liver receives blood from two main sources, the portal vein and the hepatic artery. About 80% of blood in the liver comes from portal vein and the other 20% comes from the hepatic artery, whereas liver cancer receive all their blood from the hepatic artery [2, 3]. Injecting contrast enhancing agents into the portal vein and performing a liver CT scan shortly thereafter will cause the liver to appear brighter than the cancer because it receives the entire dosage of contrast agent. This technique is called *Portal Venous Phase (PVP) imaging*. Unfortunately the specificity of this method is a function of the size of the tumor and prone to a high false positive rate because other types of benign liver lesions have similar manifestation on CT images [4]. Human variability has been shown to be an issue regarding lesion differential diagnosis, and automated methods are being investigated to facilitate the diagnosis of these lesions [5].

Recent research has investigated computer-aided methods to facilitate diagnosis by providing a database of annotated images that can be retrieved by similarity [6]. It suggests that using *semantic annotations* and *computational imaging features* can lead to highly accurate diagnosis. Semantic annotations are a controlled terminology used by radiologists to convey image content, called semantic features. Computational features are quantitative markers derived using image-processing algorithm that are independent of the semantic features. Machine-learning methods have been used to explore different feature combinations that produce the best diagnosis results. However, investigation of different methods is still in preliminary stages.

In this study we will compare the performance of an array of widely used machine-learning methods to examine the imaging feature space comprising semantic and computational features. In addition, we will apply feature selection to identify the most predictive semantic and computational imaging features for diagnosis.

2. Methods

2.1 Dataset

We have obtained 79 CT images of liver lesions, which consist of eight types of lesion diagnoses: *metastasis*, *hemangioma* (abnormal buildup of blood vessels in the liver), *hepatocellular carcinoma (HCC)*, *focal nodule* (unknown cause), *abscess* (inflammation), *laceration* (injury or tear), fat *deposition and cyst* (Figure 1). Among these, HCC and metastasis are malignant and cancerous.

2.1.1 Semantic Features

For each scan, the axial slice with the largest lesion volume was selected for analysis. Two radiologists drew a Region of Interest (ROI) around the lesion on these images. Radiologists then made annotations on these images with the imaging physician's annotation device (iPAD) [7].

The iPAD system uses a controlled radiological vocabulary called RadLex to define observations [8]. A binary semantic feature vector was created from these annotations to indicate presence or absence of a specific observation.

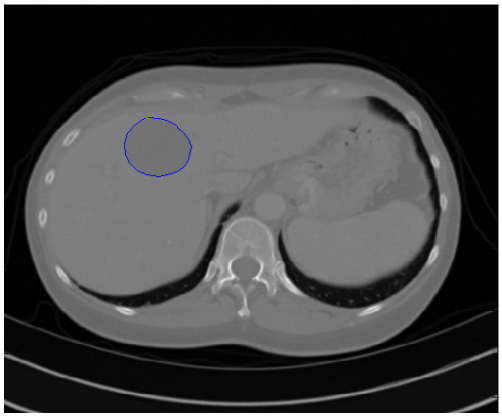

Figure 1: An example of an axial spiral CT slice of the liver with a delineated cystic lesion.

2.1.2 Computational features

The segmented liver lesions were processed using a variety of computer vision algorithms to quantify texture, lesion boundary, lesion shape, and contrast (TABLE 1). The extracted computational features were concatenated into a feature matrix and normalized to have means zero and standard deviations one.

Table 1. Twelve groups of Computational Features

Computational feature group description	Feature group dimension
Proportion of pixels with intensity larger than 1100	1
Entropy of histogram	1
Peak Position	1
Histogram	9
Difference In and Out	1
Variance	1
Gabor	12
Edge Sharpness	60
Haar on Histogram	1
Daube on Histogram	324
Histogram on Edge	1
Shape	19

2.2. Cancer classification

Lesions were predicted as malignant/benign using a variety of classifiers over three different sets of features: computational, semantic, and both. Results were evaluated using leave one out cross validation (LOOCV). The classifiers we used included Majority Rule (MR), Naïve Bayes (NB), K-nearest neighbor (KNN), Gaussian Discriminant Analysis (GDA), and Support Vector Machines (SVM). As our feature space far exceeds the number of patient samples available, GDA and SVM did not converge using the full feature sets. Thus, we selected the first 25 principal component vectors of the feature sets that account for 99% of the variance and trained GDA and SVM classifiers.

2.3. Semantic feature selection

We used feature selection to determine the most descriptive semantic features for predicting cancer. The sequential forward search (FS) algorithm was used over several classifiers with 10-fold cross validation as the optimization function [9]. Classifiers we used were KNN, SVM, and general linear models (GLM). For each classifier, we selected a series of different parameters for both the forward search criterion and the prediction evaluation.

KNN was parameterized with odd integer $k \in [1,15]$ as the neighbor count. SVMs were trained using varying values of $\sigma \in [1,8]$ in the Gaussian kernel $K(x,y) = Exp(-\frac{\| x - y \|^2}{\sigma})$. GLM was run with *normal*, *binomial*, and *poisson* distribution models with *identity*, *logit*, and *probit* link functions in Matlab. The results of classification using these subsets were evaluated using LOOCV. Frequencies of being selected were tracked for each semantic feature. Features with higher frequencies were considered to have more

predicative power. The top six semantic features were then used to examine correlation between semantic features and computational features.

2.4 Predicting semantic features

Prediction of top six semantic imaging features was performed using GDA over the first K = 1…50 principal component vectors of the computational features. This allows us to investigate whether automatically generated computational features can be used to predict semantic features that require much more time and effort. Evaluation was performed using LOOCV.

3. Results

3.1 Different types of lesions identified by different imaging features

PCA using the first two major components shows most groups of patients with the same lesion causes are clustered together (Figure 2). The first two principal components account for 66.8% variance. Thus, different types of lesions are generally separable. In particular, cyst lesions are well separated from other types of lesions.

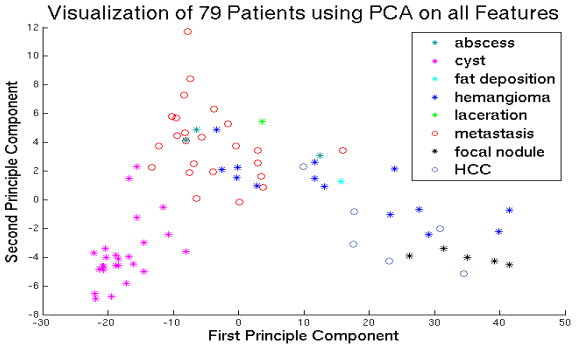

Figure 2 Visualization of 79 liver lesions using first two principal components in PCA. Circles correspond to cancerous diagnoses.

3.2 Comparison of computational and semantic features in diagnosis

Our primary goal is to accurately differentiate cancerous lesions from benign ones using different sets of imaging features on the liver CT images. The accuracy table with different classifiers consistently shows that semantic features outperform computational ones (Table 2). This suggests that the computational features may contain redundant information when given semantic features.

Table 2. Accuracy for Predicting Diagnosis

LOOCV Accuracy	Computational	Semantic	All Features
Majority Rule	62.03%	62.03%	62.03%
Naïve Bayes	74.68%	94.94%	74.68%
K-NN (k=1)	78.48%	92.41%	91.14%
K-NN (k=3)	78.48%	87.34%	87.34%
K-NN (k=5)	82.28%	84.81%	86.08%
K-NN (k=7)	79.75%	83.54%	81.01%
K-NN (k=9)	77.22%	83.54%	81.01%
GDA (25 PCVs)	86.08%	94.94%	91.14%
SVM (25 PCVs)	72.15%	84.81%	88.61%

3.3 Predictive semantic features

We applied forward search to generate a list of most predictive features for differential diagnosis of liver lesions. We empirically searched for the minimization of error in the 10-fold cross-validation of KNN, SVM and GLM.

The subsequent figure lists the most common semantic features across our forward search feature selection iterations (Figure 3). We identified the top six features: *Homogeneous retention, soft tissue density, necrosis, centripetal fill-in, central scar, and cluster-of-grapes*. Furthermore, four of these have been independently validated in the literature as being anatomical imaging markers in differential diagnosis [10,11].

3.4 Using computational features to predict semantics features

As discussed above that computational features may be redundant to semantic features in predicting diagnosis, thus, we were interested in evaluating whether semantic features can be predicted, if given automated computational features.

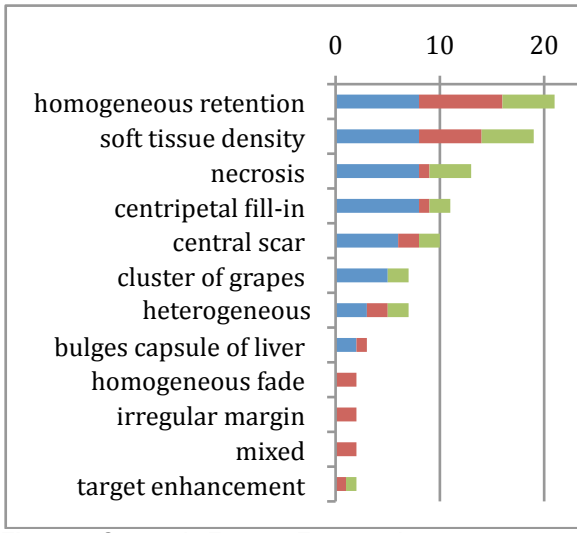

Figure 3 Semantic Feature Frequencies

We applied GDA on different numbers of principal components of the computational features. Features such as *homogeneous retention* (blue) and *central scar* (purple) achieve particularly low error rates (Figure 4).

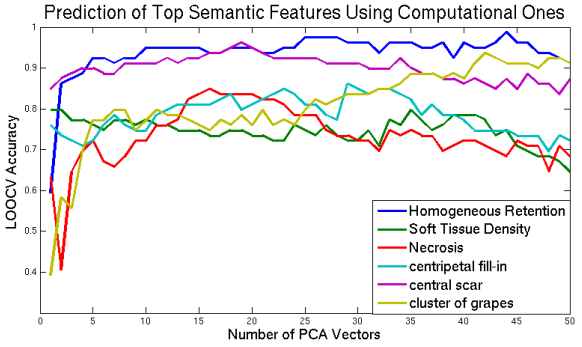

Figure 4 Top six semantic features prediction accuracy (y-axis) using Principal Components (x-axis) of computational features

3.5 Evaluation of forward search in semantic feature selection

When selecting predictive semantic features for diagnosis (Section 3.3), we applied sequential forward search, a heuristic for feature selection algorithm. We are interested in evaluating the relative degree to which over-fitting may occur when we apply different machine learning algorithms inside the forward search error-minimization criterion.

For SVMs with each radial width from 1 to 8, we performed FS to select semantic features based on their performances in predicting diagnosis. We then validated each selected feature set using SVMs with each radial widths ranging from 2 to 12.

For example, let SVM(σ) denote a SVM learning instance with a Gaussian kernel radial width of σ, and KNN(k) denote a KNN learning instance that takes the majority vote of k neighbors. SVM(8) classifies with 94.9% accuracy when using features selected by an SVM(1) (Table 3a). In contrast, suppose we minimize the empirical error of KNN(5) in our feature selection process: yet, KNN(1) would achieve only 84.8% accuracy when using only the selected features (Table 3b).

This pattern suggests that SVMs are more robust in comparison to KNN when used in forward feature selection (Table 3).

Table 3. (a) Accuracy of classification using SVM with forward feature selection (b) Accuracy of classification using KNN with forward feature selection

LOOCV Accuracy		Kernel Radial Width for Classification					
		2	4	6	8	10	12
Kernel Radial Width Used in Forward Search	1	96.2%	94.9%	94.9%	94.9%	82.3%	82.3%
	2	100.0%	100.0%	97.5%	97.5%	88.6%	88.6%
	3	96.2%	94.9%	94.9%	94.9%	82.3%	82.3%
	4	100.0%	100.0%	97.5%	97.5%	88.6%	88.6%
	5	100.0%	100.0%	97.5%	97.5%	88.6%	88.6%
	6	97.5%	97.5%	97.5%	97.5%	97.5%	81.0%
	7	97.5%	97.5%	97.5%	97.5%	97.5%	81.0%
	8	97.5%	97.5%	97.5%	97.5%	97.5%	81.0%

LOOCV Accuracy		K (Neighbor Counts in KNN) for Classification					
		1	3	5	7	9	11
K used in Forward Search	1	93.7%	92.4%	82.3%	82.3%	81.0%	79.8%
	3	91.1%	98.7%	97.5%	94.9%	91.1%	79.8%
	5	84.8%	97.5%	97.5%	96.2%	87.3%	88.6%
	7	81.0%	93.7%	93.7%	94.9%	88.6%	89.9%
	9	79.8%	77.2%	89.9%	89.9%	89.9%	88.6%
	11	79.8%	77.2%	88.6%	88.6%	88.6%	88.6%
	13	79.8%	77.2%	88.6%	88.6%	88.6%	88.6%
	15	89.9%	89.9%	89.9%	89.9%	86.1%	91.1%

4. Discussion

We have shown that diagnosis of malignant cancer using a rich, controlled radiological vocabulary (semantic features) outperforms general computational features. This finding should be of little surprise given than these radiological observations are optimized for such diagnostics.

Subsequent analysis of semantic features revealed six semantic features with high predictive power for differential diagnosis: *homogeneous retention*, *soft tissue density*, *centripetal fill*, *necrosis*, *central scar*, and *circumscribed margin*. These findings are consistent with published literature as being indicative of either malignant or benign lesions.

Moreover, we showed that these highly descriptive semantic features could be accurately classified with automatically extracted computational features. This result can be applied to Computer Aided Diagnostic (CAD) systems by providing suggestions of semantic features to radiologists.

In addition, we investigated over-fitting problem in forward feature selection for diagnosis with different classifiers. SVM appears more robust than KNN.

Finally, representative ROC curves show that we achieved excellent sensitivities and specificities simultaneously when predicting diagnoses with semantic features (Figure 5).

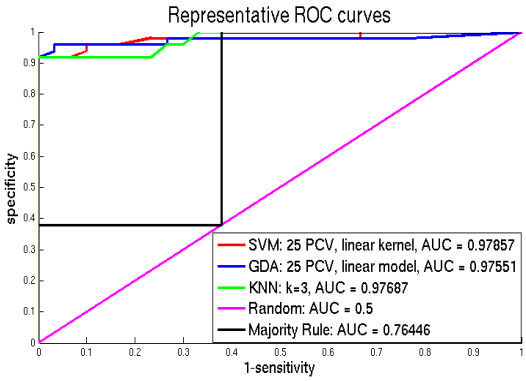

Figure 5. ROC Curves of different classifiers for predicting diagnosis

References

1. Baron, R. Understanding and optimizing use of contrast material for CT of the liver. *Am. J. Roentgenol.* **163**, 323-331 (1994).
2. Lautt, W.W. & Greenway, C.V. Conceptual review of the hepatic vascular bed. *Hepatology* **7**, 952-963 (1987).
3. Matsui, O. et al. Benign and malignant nodules in cirrhotic livers: distinction based on blood supply. *Radiology* **178**, 493-497 (1991).
4. Lencioni, R., Cioni, D., Della Pina, C., Crocetti, L. & Bartolozzi, C. Imaging diagnosis. *Semin. Liver Dis* **25**, 162-170 (2005).
5. Armato, S.G. et al. The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. *Acad Radiol* **14**, 1409-1421 (2007).
6. Napel, S. A. et al. Automated Retrieval of CT Images of Liver Lesions on the Basis of Image Similarity: Method and Preliminary Results. *Radiology* 257 (3), (2010).
7. Rubin, D.L., Rodriguez, C., Shah, P. & Beaulieu, C. iPad: Semantic Annotation and Markup of Radiological Images. *AMIA Annu Symp Proc* 2008, 626-630 (2008).
8. Langlotz, C.P. RadLex: a new method for indexing online educational materials. *Radiographics* 26, 1595-1597 (2006).
9. Kohavi, R. & John, G.H. Wrappers for feature subset selection. *Artificial Intelligence* 97, 273-324 (1997).
10. Youk, J.H., Kim, C.S. & Lee, J.M. Contrast-enhanced agent detection imaging: value in the characterization of focal hepatic lesions. *J Ultrasound Med* **22**, 897-910 (2003).
11. El-Serag, H.B., Marrero, J.A., Rudolph, L. & Reddy, K.R. Diagnosis and treatment of hepatocellular carcinoma. *Gastroenterology* **134**, 1752-1763 (2008).

Acknowledgements

We would like to thank Sandy Napel and Daniel Rubin for providing data sets, and Jiajing Xu for his work in extracting computational feature.