Towards Content-Based Image Retrieval: From Computer Generated Features to Semantic Descriptions of Liver CT Scans

Assaf B. Spanier and Leo Joskowicz

The Rachel and Selim Benin School of Computer Science and Engineering
The Hebrew University of Jerusalem, Israel.
{assaf.spanier,leo.josko}@mail.huji.ac.il
http://www.cs.huji.ac.il/~caslab/site/

Abstract. The rapid increase of CT scans and the limited number of radiologists present a unique opportunity for computer-based radiological Content-Based Image Retrieval (CBIR) systems. However, the current structure of the clinical diagnosis reports presents substantial variability, which significantly hampers the creation of effective CBIR systems. Researchers are currently looking for ways of standardizing the reports structure, e.g., by introducing uniform User Express (UsE) annotations and by automating the extraction of UsE annotations with Computer Generated (CoG) features. This paper presents an experimental evaluation of the derivation of UsE annotations from CoG features with a classifier that estimates each UsE annotation from the input CoG features. We used the datasets of the ImageCLEF-Liver CT Annotation challenge: 50 training and 10 testing CT scans with liver and liver lesion annotations. Our experimental results on the ImageCLEF-Liver CT Annotation challenge exhibit a completeness level of 95% and accuracy of 91% for 10 unseen cases. This is the second best result obtained in the Liver CT Annotation challenge and only 1% away from the rst place.

1 Introduction

About 68 million CT scans are performed in the USA each year. Clinicians are struggling under the burden of diagnosis and follow-up of such an immense amount of scans. This phenomenon gave rise to a plethora of methods to improve and assist clinicians with the diagnosis process.

Content based image retrieval (CBIR) is a growing and popular research topic [8]. The goal of CBIR is to assist physicians with the diagnosis of tumors or other pathologies by finding similar cases to the case at hand. Therefore, CBIR requires efficient search capabilities in a huge database of medical images. The matching criteria are based on image properties and features extracted from image and pathology [1] and on searching in the clinical reports database.

Besides the known problem with diagnosis and follow-up of this huge number of scans, there is substantial variability in the structure of the clinical reports provided by the clinicians for each case. This variability hampers the ability to establish an efficient and consistent CBIR system, as uniform reports structure is a major need for such an application.

The task of standardization the clinical reports for liver and liver lesions has recently been proposed by Kokciyan et al. [1]. The ONLIRA ontology constitutes a standard that is used to generate multiple-choice User Express (UsE) annotations [9] consisting of features that clinically characterize the liver and the liver lesion. Note that the UsE annotations are provided by the radiologist, as they cannot be extracted automatically from the image itself. However, the image descriptors, called Computer Generated (CoG) features, can be automatically derived from the image with image processing algorithms [9].

The goal of this work is therefore to use the CoG features to automatically generate the UsE annotations. A major part of this work deals with designing and building a machine learning algorithm that link CoG features to UsE annotations. Training datasets provided by the ImageCLEF-Liver CT challenge [6] which is part of the ImageCLEF-2014 evaluation campaign [7]. Additional contributions of this work consist of extending the CoG available features and optimal selection of the most relevant ones to the liver task. Experimental results on the ImageCLEF-Liver CT Annotation challenge exhibit estimation of UsE annotations at a completeness level of 95% and accuracy of 91% for 10 unseen cases. This is the second best result obtained in the Liver CT Annotation challenge [6] which is part of the ImageCLEF-2014 evaluation campaign [7] and only 1% away from the first place.

2 Method

A major part of this work would deal with developing a machine learning algorithm that would best link CoG features to UsE annotations based on training data-sets. Developing a machine learning algorithm involves four main steps [2]: (1) Data collection; (2) Feature extraction; (3) Model selection/Fitting classifier parameters and; (4) Training the selected model. A diagram illustrating the phases of the process is shown in Fig 1. We describe each step in detail below.

2.1 Data Collection

The input of our algorithm is a set of 50 datasets provided by the imageCLEF 2014 Liver Annotation Task data and has been collected by CaReRa Project (TUBITAK Grant 110E264), Bogazici University, EE Dept., Istanbul, Turkey (www.vavlab.ee.boun.edu.tr). Each dataset includes:

1. A CT scan that consists the liver region and liver tumors
2. A segmentation of the liver
3. The lesion's bounding box
4. A set of 60 Computer Generated features (CoG) features
5. A set of 73 User Express (UsE) annotations

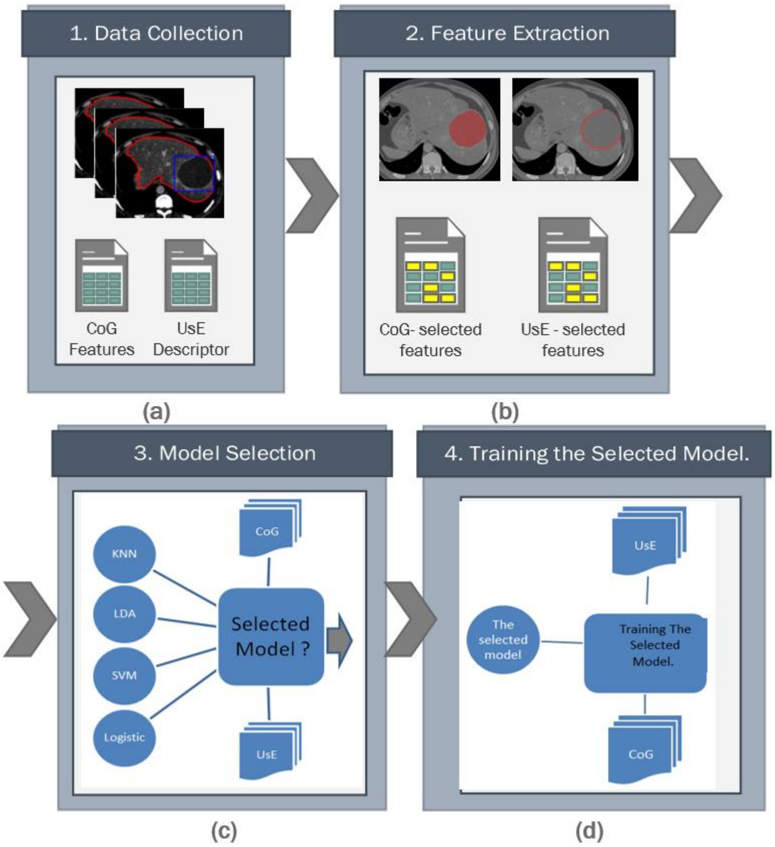

Fig. 1. Four main steps in designing and building a machine learning algorithm for the estimation of UsE from CoG and CT images: (a) Data collection: the input to our system consists of CT scans and CoG features and UsE annotation ; (b) Feature extraction: only the most informative CoG features are selected; (c) Model selection: Estimating the most appropriate model for the UsE annotations from CoG features. The model is selected after testing a variety of prediction models and; (d) Training the selected model: The selected models parameters are trained using all 50 cases.

The CoG features can be divided to global image descriptors and pathology descriptors. The global image descriptors cover the basic and liver-wide global statistical properties, such as the mean and variance of the gray-level values, and the liver volume. They are extracted directly from the CT scans and the associated segmentation. The pathology descriptors are computed for each liver lesion. They reflect finer levels of visual information related to individual lesions.

UsE annotations are also divided into global and pathology descriptors. Global descriptors are divided into Liver and Vessel groups. These two groups include annotations about the liver itself and its hepatic vasculature. Pathol-

Table 1. Example of the Computer Generated features (CoG) features. The CoG can be divided into two main Descriptor type (a) Global - liver, vessel and all lesions annotations. (b) Pathology - the selected annotated lesion

Descriptors type	Group	**Name**	**Type**	**Value**
Global	Liver	LiverVolume	double	12987.6
	Vessel	LiverVariance	double	297.683
	AllLesions	NumberofLesions	int	5
Pathology	Lesion	HaarWaveletCoef	VectorOfDouble	8.4, 3.9, 2.1,

ogy descriptors include two groups: Lesion and Lesion Component and contain annotations about the selected lesion in the liver.

A complete list of all UsE and CoG features with their associated descriptor type can be found at [6]. Representative examples of CoG and UsE are shown in Table 2.1 and Table 2.1 respectively.

Table 2. Example of the User Express (UsE) annotations . The UsE can be divided into two main groups (a) Global - liver and vessel annotations (b) Pathology - lesion annotations

Descriptors Type	**Group**	**Concept**	**Properties**	**Values**	**Indices**
Global	Liver	Right Lob	Right Lobe Size Change	Normal	2
	Vessel	Hepatic Artery	Hepatic Artery Lumen Diameter	normal	2
Pathology	Lesion	Lesion	is Close to Vein	Other	8
	Lesion	Lesion	Segment	SegmentI, SegmentII	1,2
	Lesion	Capsule	Is Calcified?(Capsule)	True	1

2.2 Feature Extraction

In this step we aim at extracting the optimal set of CoG features for the estimation of each UsE annotation. Note that due to the diversity of lesions between and within patients, estimating pathology descriptors is a much more challenging task than estimating global descriptors. Thus, our analysis includes developing two distinct classification models: one for the global CoG features and one for the CoG pathology descriptors.

First, we reduce problem dimensionality by omitting 21 high dimension features from the 60 provided CoG (e.g. *FourierDescriptors, BoundaryScaleHistogram, BoundaryWindowHistogram* etc.). Thus, our analysis includes only CoG features with scalar values. This results in 39 features divided between global and pathology related features.

The 18 global CoG features are: *LiverVolume , LiverMean, LiverVariance, VesselRatio, VesselVolume, MinLesionVolume, MaxLesionVolume, LesionRatio, AllLesionsMean, AllLesionsVariance, AllLesionsSkewness, AllLesionsKurtosis, AllLesionsEnergy, AllLesionsSmoothness, AllLesionsAbcssia, AllLesionsropy, AllLesionsThreshold, NumberofLesions*

The 21 pathology related CoG features are: *LesionMean, LesionVariance, LesionSkewness, LesionKurtosis, Lesionnergy, Lesionmoothness, Lesionbcssia, LesionEntropy, LesionThreshold, Lesion2VesselMinDistance, Lesion2VesselTouchRatio, VesselTotalRatio, VesselLesionRatio, Volume, SurfaceArea, MaxExtent, AspectRatio, Sphericity, Compactness, Convexity, Solidity*

We added 9 features to the 21 pathology features to describe the statistics of the lesion itself. The new features are derived from a refined segmentation of the liver obtained by thresholding the given lesion bounding box with its mean gray level followed by morphological operations. The added CoG features are:

1. The average gray level intensity values of the healthy part of the liver. *(LiverGrayMean)*
2. The standard deviation of gray level intensity values of the healthy part of the liver. *(LiverGrayStd)*
3. The average gray level intensity values of the lesion. *(LesionGrayMean)*
4. The standard deviation of gray level intensity values of the lesion. *(LesionGrayStd)*
5. The lesion's contour mean gray levels. *(LesionBounderyGrayMean)*
6. The standard deviation of the lesion's contour gray levels. *(LesionBounderyGrayStd)*
7. The average gray level difference between the healthy part of the liver and the lesion. *(LesionLiverGrayDiff)*
8. The average gray level difference between the healthy part of the liver and the lesion' contour. *(BounderyLiverGrayDiff)*
9. The average gray level difference between the the lesion and its contour. *(lesionBounderyGrayDiff)*

The result is a modified CoG list with 18 global image descriptors and 30 pathology descriptors.

2.3 Model Selection

In this section the classification algorithm to be evaluated is presented: Predictive models can be characterized by two properties: parametric/non-parametric and generative/discriminative. Parametric models have a fixed number of parameters and have the advantage of often being faster to use. However, they tend to rely on stronger assumptions about the nature of the data distributions. In non-parametric classifiers, the number of parameters grows with the size of the training data. Non-parametric classifiers are more flexible but are often computationally intractable for large datasets.

Table 3. Four classifiers examined

	Generative	Discriminative
Parametric	Linear Discriminant Analysis (LDA)	Logistic Regression (LR)
Non-Parametric	K-Nearest Neighbors (KNN),	Support Vector Machine (SVM)

As to the generative/discriminative property, the main focus of generative generative models, is not the classification task but to correctly model the probabilistic model. They are called generative since sampling can generate synthetic data points. Discriminative models, however, do not attempt to model the underlying probability distributions, but rather focus on the given task, i.e. the classification itself. Therefore, they may achieve better performance in terms of overall accuracy of the classification task. In general, when the probabilistic distribution assumptions made are correct, the generative model requires less training data than discriminative methods to provide the same performance, but if the probabilistic assumptions are incorrect, discriminative methods will do better [4]. Table 2.3 shows the characteristics of each classifier.

For real world datasets, so far there is no theoretically correct, general criterion for choosing between the different models. Therefore, we examined four classifiers, representative of the 4 different families of models [3]: K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Logistic Regression (LR), Support Vector Machine (SVM). Note that for each UsE annotation, the outcome of each predictive model consists of classification and subset selection of the optimal CoG features. Therefore, the selection of the CoG features is unique for each model.

We used the Python scikit-Learn Machine learning tool [5] to examine the four selected classifiers. For each UsE descriptor, the best predicting classifier and its features was based on leave-one-out cross validation with exhaustive search – i.e. systematically enumerating all possible Combinations CoG features. Note that since we develop two distinct classication models, one for the 18 global CoG features and one for the 30 CoG pathology descriptors, the exhaustive search was performed for each CoG group separately. Three UsE features (Cluster size, Lobe and Segment) were estimated from the image itself and were not part of the learning process (Section 2.6).

For simplicity, each model was tested with a set of default parameters as define by scikit-learn package [5]. The parameters for each model are:

- KNN: K=5, Distance: euclidean distance with no threshold for shrinking.
- LDA: Euclidean distance, regularization strength of 1.0.
- LR: L2 penalty, regularization strength of 1.0, tolerance for stopping criteria is 0.0001.
- SVM: Penalty parameter of 1.0, RBF kernel with degree of 3 and gamma of 0, tolerance for stopping criteria is 0.0001.

2.4 Training

Once the classifier which produced the highest classification was found in the previous step, we trained it using all 50 cases. As a result, for each UsE, we obtain a trained model, consisting of classifier along with optimized sets of CoG features (i.e. the selected features)

2.5 Evaluation

The evaluating phase of the challenge consists of the estimation of UsE annotation from the given CoG features and the images. Unlike the training phase, UsE annotation are not given here to examine our accuracy.

To apply the resulting classifier in the testing phase on an unseen dataset, we first extract and extend the CoG features according to a scheme described in Section 2.3. Then, for each test case, we apply the prediction model – the one with the highest score – according to the training phase results. The result is a UsE annotation for each unseen case.

2.6 Estimation of the Lesion Lobe, Lesion Segment, and Cluster Size

As noted in Section 2.3, the *Cluster Size* (i.e. the number of lesions inside the lesion bounding box), the *Lesion Lobe* and the *Lesion Segment* containing the lesion were not part of the general learning process but were rather estimated from the image itself. The estimation of these fields was performed as follows:

For the *Lesion Lobe*: we measure the center of the lesion and the liver. The lesion lobe is estimated as the right lobe if the lesion center is on the right part of the liver, and as the left lobe if the lesion center is on the left part of the liver. In case they were overlapping we estimated it to be the Caudate Lobe.

The *Lesion Segment* is estimated as follows: If at a previous stage the estimation was that the lesion is on the right lobe, we assessed that the lesion is located on the fourth segment. Alternatively, if at the previous stage the estimation was that the lesion is on the left lobe, we analyzed whether the lesion is located above or below the center of the liver. If it is located above the center of the liver, we assessed that the lesion is in segment 5-6; and if it is located below the center of the liver - we assessed that the lesion is on segment 7-8.

For the *Cluster Size*: We define the *Cluster Size* to the number of lesion – the value listed in the CoG field *NumberofLesions*. Except in cases when the value listed in number of lesion is higher than 6, and then we define that the number of lesions to be 6. For the *Cluster Size*: We define the *Cluster Size* to the number of lesion – the value listed in the CoG field *NumberofLesions*. Except in cases when the value listed in number of lesion is higher than 6, and then we define that the number of lesions to be 6.

Table 4. Training and test results for each of the classifiers. The values indicate the percentage of accuracy of the leave-out-out for each UsE. For convenience, results are grouped s.t. UsE annotations that exhibit the same results are shown in one line in the table.

Training Results							Test Results
Group	UsE Annotations	**KNN**	**LR**	**LDA**	**SVM**	**Average**	**Average**
Liver	All	0.92	0.92	0.92	0.92	0.92	0.92
Vessel	All	1	1	1	1	1	1
Lesion-Lesion	Contrast Uptake	0.7	0.62	**0.79**	0.66		
Lesion-Lesion	Contrast Pattern	**0.75**	0.7	0.58	0.67		
Lesion-Lesion	Others	0.92	0.92	0.92	0.92	0.8	0.68
Lesion-Area	is contrasted	0.74	0.75	**0.79**	0.76		
Lesion-Area	Density	0.9	0.9	**0.92**	0.9		
Lesion-Area	Density Type	0.75	0.76	**0.8**	0.76		
Lesion-Area	Is Peripherical Localized	**0.8**	0.76	0.72	0.74		
Lesion-Area	Is Central Localized	**0.8**	0.76	0.72	0.74	0.83	0.74
Lesion-Area	Others	0.91	0.91	0.91	0.91		
Lesion Component	All	0.93	0.93	0.93	0.93	0.93	0.93

3 Results

Experimental results of applying our method to the ImageCLEF-Liver CT Annotation challenge datasets results with estimation of UsE annotations at a completeness level of 95% and accuracy of 91% for 10 unseen cases.

Training and test results for each of the classifiers are shown in Table 2.6. Due to the simplicity of estimating the global UsE annotations, we present their result per group. A detailed presentation is provided for the pathology related features, which is indeed a much more challenging task. It can be seen that all classifiers successfully estimated the global features.

As mentioned, three additional features were estimated from the images and were not part of the learning process. These features are: *ClusterSize, LesionLobe, LesionSegment* and their accuracy on the training datasets was 0.75, 0.9, 0.7 respectively. The completeness level of our method is 0.95 due to omitting 3 UsE annotations from the analysis: *LesionComposition*, *Shape* and the *MargenType*.

The optimized sets of CoG features (i.e. the selected features) that were obtained by the model with the highest score after the leave-one-out procedure are shown in Table 5. Note that the set of added features (Section 2.2) were indeed selected by the model, which proves their necessity.

Table 5. The optimized sets of CoG features selected, as a result of the exhaustive procedure, for the classifier that exhibits the highest score by the leave-one-out cross validation . For convenience, results are grouped s.t. UsE annotations that exhibit the same results are shown in one line in the table.

Group	**UsE Annotations**	**Selected Classifier**	**Selected CoG Features**
Liver	All	Any	All
Vessel	All	Any	All
Lesion-Lesion	Contrast Uptake	LDA	*BounderyLiverGrayDiff*, *LesionGrayStd*, *Entropy*, *SurfaceArea*
Lesion-Lesion	Contrast Pattern	KNN	*LesionGrayMean*, *LesionBounderyGrayMean*, *BounderyLiverGrayDiff*, *LesionBounderyGrayStd*,
Lesion-Lesion	Other	Any	All
Lesion-Area	Is Contrasted	LDA	*LesionLiverGrayDiff*, *Solidity*, *Entropy*, *Kurtosis*,
Lesion-Area	Density	LDA	*lesionBoundryGrayMean*
Lesion-Area	Density Type	LDA	*LesionBounderyGrayMean*, *Solidity*, *LesionGrayStd*
Lesion-Area	Is Peripherical Localized	KNN	*LiverGrayStd*, *LesionGrayStd*, *lesionBounderyGrayStd*
Lesion-Area	Is Central Localized	KNN	*LiverGrayStd*, *LesionGrayStd*, *LesionBounderyGrayStd*
Lesion-Area	Other	Any	All
Lesion-Component	All	Any	All

4 Conclusion and future work

We have presented an approach to estimate UsE annotations from CoG features and associated CT scans. We extended the CoG features with 9 additional features to enhance the learning process. In the ImageCLEF-Liver CT Annotation challenge. Our approach provides an average accuracy level of 91% with completeness level of 95% when applied on 10 unseen test cases. This work provides reliable estimation of uniform clinical report from imaging features and therefore it constitutes another step toward automatic CBIR system by enabling efficient search in clinical reports. Future work consists of examining an additional set of classifiers and extending the completeness of our algorithm to estimate the full UsE annotations set and values (e.g. estimating segment 1-3 of the Lesion Segment feature)

References

1. Kokciyan, N., Turkay, R., Uskudarli, S., Yolum, P., Bakir, B., Acar, B. . Semantic Description of Liver CT Images: An Ontological Approach. IEEE Journal of Biomedical and Health Informatics, vol. 2194 (2014): 11, .
2. Duda, Richard O., Peter E. Hart, and David G. Stork. "Pattern classification." New York: John Wiley, Section 10 (2001).
3. Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT Press, (2012)
4. Jordan, A.: On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems, 14, 841. (2002)
5. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Duchesnay E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12:28252830, (2011)
6. Marvasti, N., Kökciyan, N., Türkay, R., Yazı, A., Yolum, P., Üsküdarlı, S. and Acar, B.: imageCLEF Liver CT Image Annotation Task 2014 In: CLEF 2014 Evaluation Labs and Workshop, Online Working Notes (2014)
7. Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., Marvasti, N., Üsküdarlı, S., Paredes, R., Cazorla, M., Garcia-Varea, I. and Morell, V.: ImageCLEF 2014: Overview and analysis of the results. Springer Berlin Heidelberg. (2014)
8. Akgl, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., Acar, B.: Content-based image retrieval in radiology: current status and future directions. Journal of Digital Imaging, 24(2), 208-222. (2011).
9. Barzegar Marvasti, N., Akgl, C. B., Acar, B., Kkciyan, N., skdarl, S., Yolum, P., Trkay, R., Bakr, B.: Clinical experience sharing by similar case retrieval. In: Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare (pp. 67-74). ACM. (2013)