Journal of Digital Imaging
https://doi.org/10.1007/s10278-019-00298-w

ORIGINAL PAPER

Generate Structured Radiology Report from CT Images Using Image Annotation Techniques: Preliminary Results with Liver CT

Samira Loveymi[1] · Mir Hossein Dezfoulian[1] · Muharram Mansoorizadeh[1]

Abstract

A medical annotation system for radiology images extracts clinically useful information from the images, allowing the machines to infer useful abstract semantics and become capable of automatic reasoning and making diagnostic decision. It also supplies human-interpretable explanation for the images. We have implemented a computerized framework that, given a liver CT image, predicts radiological annotations with high accuracy, in order to generate a structured report, which includes predicting very specific high-level semantic content. Each report of a liver CT image is related to different inhomogeneous parts like the liver, lesion, and vessel. We put forward a claim that gathering all kinds of features is not suitable for filling all parts of the report. As a matter of fact, for each group of annotations, one should find and extract the best feature that results in the best answers for that specific annotation. To this end, the main challenge is discovering the relationships between these specific semantic concepts and their association with the low-level image features. Our framework was implemented by combining a set of the state-of-the-art low-level imaging features. In addition, we propose a novel feature (DLBP (deep local binary pattern)) based on LBP that incorporates multi-slice analysis in CT images and further improves the performance. In order to model our annotation system, two methods were used, namely multi-class support vector machine (SVM) and random subspace (RS) which is an ensemble learning method. Applying this representation leads to a high prediction accuracy of 93.1% despite its relatively low dimension in comparison with the existing works.

Keywords Deep local binary pattern · Medical image annotation · Computed tomography · Liver CT images

Introduction

In modern healthcare, medical imaging plays a crucial role in paving the way for quick and reliable diagnosis, medical reference, training, treatment planning, and analysis of response to therapy. With the advent of medical imaging systems and further employment of them in practice, the amount of medical image data has constantly been growing. This raises the need for building powerful algorithms which can efficiently process such data and extract the useful and relevant information hidden in the pixels and voxels. This, in turn, makes the medical imaging informatics research such as annotation, semantic-based image retrieval, modality-classification, content-based retrieval, case-based retrieval, and classification more significant [1].

It is well established from previous works on image annotation that text queries are more comfortable for users to explain what they want to retrieve from the databases than visual queries [2]. To reach this goal, it is of particular importance to effectively extract information embedded in the patients' medical imaging and annotate the images accurately and consistently. On the other hand, manual annotation of the medical images with human experts is time-consuming, costly, and vulnerable to human mistakes. Moreover, the manual classification, given the large and constantly growing databases of medical images, leads to errors in the tag assignment, which means that a part of the available knowledge will not be accessible to the physicians anymore. To tackle the issue, there is a need for reliable automatic annotation algorithms, capable of performing the task effectively.

✉ Muharram Mansoorizadeh
mansoorm@basu.ac.ir; muharram@gmail.com

Samira Loveymi
s.loveymi@basu.ac.ir

Mir Hossein Dezfoulian
dezfoulian@basu.ac.ir

[1] Computer Engineering Department, Bu-Ali Sina University, Shahid Fahmideh blvd., Hamedan, Iran

Published online: 14 November 2019

One of the key research challenges in the semantic image annotation is to detect discriminative and robust low-level features and discover the relations between them and the higher-level labels derived from a standard terminology. The performance of such pattern recognition systems highly depends on employing a front-end that supplies rich and representative features, capable of capturing different high-level visual properties and nuances of the images. It also involves applying effective learning algorithms for training the back-end for carrying out a reliable image annotation and classification. A considerable amount of literature in this regard has been dedicated to the methods that only annotate the medical images based on the general properties of the image. For example, Tommasi et al. proposed an image annotation system, which annotates the medical images in terms of the properties of the image, such as the image modality, body orientation, body region, and biological system. They combined a set of local and global image features, such as raw pixel values and SIFT with different fusion schemes [3, 4]. They achieved high recognition rates in the ImageCLEF 2009 medical image annotation task [5]. Dimitrovski et al. proposed a hierarchical multi-label classification (HMC) system for annotating medical images in which an image can simultaneously belong to multiple classes with a hierarchical arrangement [6]. Fushman et al. employed features extracted from image and text to retrieve clinically relevant images using supervised machine learning methods [7]. Camlica et al. [8] have proposed medical image classification and retrieval system using IRMA dataset, based on support vector machine (SVM) classifier trained with local binary pattern (LBP) features derived from detected salient regions of images. Bouslimi et al. [9] have suggested an automatic system for medical image annotation that extracts the co-occurrence between textual terms and visual descriptors using latent semantic vector using a bag of visual words (BOVW). Actually, for automatically annotating a new medical image, they compare the vector describing the source image with the similar vectors that already exist in the database.

The literature in medical image annotation can be grouped into two categories: First, general medical image annotation that provides information like modality and body region as the image annotation [3–12]; second, the systems that try to annotate images with clinically relevant content, which is a more difficult objective and can improve clinical workflows or the performance of automatic tools already incorporated by the patients or physicians in those workflows [13].

Another application of medical image annotation is in building medical computer-aided diagnosis (CAD) systems that fill in the structured radiology reports about computed tomography (CT) images. Radiologists investigate CT scans and record their observations in some reports to communicate with physicians; if these reports have a standard and structured format, it will be more effective and more beneficial to maintain [14]. Such structured reports are very instrumental in a wide range of medical applications such as formal reporting, image retrieval, and CAD systems built using state-of-the-art machine learning techniques. To this end, many research studies have focused on efficient information extraction tasks from radiology reports [15, 16]. However, producing structured reports with standard format is not a trivial task, since a high level of domain-specific knowledge is required. Also, the mismatch between experts' opinions is likely to happen during manual labeling of the data which leads to the so-called teacher error. As a result, in practice, such structured medical reports may be incomplete and are not available in large amounts [17, 18] which, in turn, hinder building effective pattern recognition systems for such tasks.

The goal of the liver annotation task [18] is automatically filling in a structured radiology report about the liver CT volumes. The primary purpose of this task is describing the semantical and physical features of the liver, its vascularity, and the types of lesions in the liver.

In [17], two methods are proposed for liver CT annotation task. The first one utilizes a combination of shape, texture, and lesion features using random forest classifier. The second one uses the signature of the liver and Hamming distance for retrieval. This combination leads to the best results in ImageCLEF 2015 [18]. In [19], techniques from content-based image retrieval (CBIR) were employed to annotate liver CT images. They used the dataset containing computer-generated (CoG) image features, BOVW features, a two-stage SVM, and weighted the nearest neighbor for annotation.

One of the important issues in the clinically relevant annotation is that there is a wide range of semantic labels to learn and often very few labeled training samples are available, e.g., [13]. This, in turn, poses two significant challenges: developing representative feature extraction methods and designing annotation techniques that are less affected by a limited amount of training data.

To compensate for the lack of training data, Xue et al. [20] annotated abdominal CT images using CBIR to identify/retrieve a set of labeled images that bear visual resemblance to a given unlabeled query image. Then, they managed to annotate the unlabeled images using annotations taken from visually similar labeled images.

Some published studies in liver CT analysis domain only focus on the characterization of liver lesions, which is a part of the information that should be generated for filling in a clinical radiology report. Depeursinge et al. [21] tried to learn the semantic terms that describe the visual appearance of liver lesions. The visual elements are modeled by a linear combination of high-order Riesz multi-scale wavelet features. Yingying Xu et al. [22] used a texture-based version of BOVW and spatial information of the visual words in the ROI for retrieval of the focal liver lesion (FLL) using multi-phase CT images. Cirujeda et al. [23] investigated the application of a number of covariance-based descriptors for shape

and texture fusion of 3D surfaces and tissue characterization in 3D CT imaging.

One of the main considerations in feature extraction step is taking into account the volumetric nature of liver CT images; [24] studied the use of LBP for 2D and 3D image in detecting lung cancer, and the results indicate that the accuracy of using 3D images outperforms 2D images.

In this paper, we present a novel method for annotation of liver CT images, which generate a structured radiology report. Given an image to annotate, we propose an informative visual feature extraction method in combination with texture, shape, and BOVW features and employ two types of back-ends for achieving the best annotations: SVM classification and ensembles using nearest neighbors. The proposed approach is evaluated on CLEF liver CT annotation dataset [18] and compared to the state-of-the-art solutions for this task.

The rest of this paper is organized as follows. "Dataset" is dedicated to the applied dataset and reviews its main properties. "Methods" is dedicated to the methods employed for image analysis and annotation, and the proposed feature extraction method is introduced and explained. The experimental results and discussion are presented in "Evaluation," and finally, "Conclusion" concludes the paper and provides some point of attention for future work.

Dataset

In our experimental setup, we have used a publicly available database of volumetric (3D) CT images of the liver from the ImageCLEF 2015 liver CT annotation task [18]. The training part contains 50 cases where each one consists of:

- A cropped CT image of the liver 3D matrix. The volumes had various resolutions (x 190–308 pixels, y 213–387 pixels, slices 41–588) and spacing (x, y 0.674–1.007mm, slice 0.399–2.5 mm).
- A liver mask that specifies the part corresponding to the liver (a 3D matrix).
- The bounding box (ROI) corresponding to the region of the selected lesion within the liver.
- The dataset annotations are RDF files generated using Liver Case Ontology (LiCO), representing manually transcribed imaging observations by a radiologist. These features clinically characterize the liver, hepatic vascularity, and liver lesions. The ontology describes the different sets of possible labels L_Q for different questions Q.

In this paper, only 43 of the questions have been considered for the task. Questions with unbounded labels and questions with empty labels for some images are not included in learning and evaluation processes. In contrast to the liver CT dataset used in ImageCLEF 2014 annotation task, this dataset does not include features extracted from the images. Since the whole dataset was available only to ImageCLEF participants, following the approach of Kumar et al. [19], we used the released 50 images for training and testing in a 10-fold setup.

Methods

Method Overview

The proposed annotation task is motivated by our research interests in visual question answering (VQA). We use the following terminology and notations in this paper; every element of the structured report that should be filled in automatically named a question Q that is a specific annotation task. An answer to each question is the label that the annotation system automatically assigns to that question based on the analysis of the image features. The answer is selected from a set of labels that are unique to each question. Thus, a label L_Q is an annotation that could possibly be assigned to a question.

One of the important challenges in image annotation is to decide on what kind of visual features should be taken into account and how to extract them. Such features should be discriminative. That is, in the features space, the instances belonging to different classes should be as distant as possible and the instances with the same label should occupy the same neighborhood. In addition, the features should be robust, that is, they should be sensitive to data attributes which are relevant to the task and, at the same time, bear a minimum level of sensitivity to irrelevant and nuisance aspects of the data. Furthermore, the computational complexity of the feature extraction process should not be problematic for practical applications with limited computational power and memory. These issues make the feature engineering a challenging task. Although deep neural networks can potentially address these issues [25] and effectively solve the data representation problem, due to lack of enough training data, it cannot be effectively used in such tasks. Figure 1 presents an overview of the proposed method.

The question/annotation of each liver CT image is related to different in-homogeneous parts like the liver, lesion, and vessel. We put forward a claim that gathering all kinds of features is not suitable for answering all the questions. As a matter of fact, for each group of questions, one should find and extract the best feature that results in the best answers for that specific question.

Another challenge in the feature extraction phase is that liver CT images are 3D volumes, and the radiologists say that they use all the slices for answering the questions. Hence, the features should be maintained within the 3D domain.

To investigate information load of each feature type, different kinds of visual parameterizations have been examined. The 3D version of texture feature extractors includes Gabor

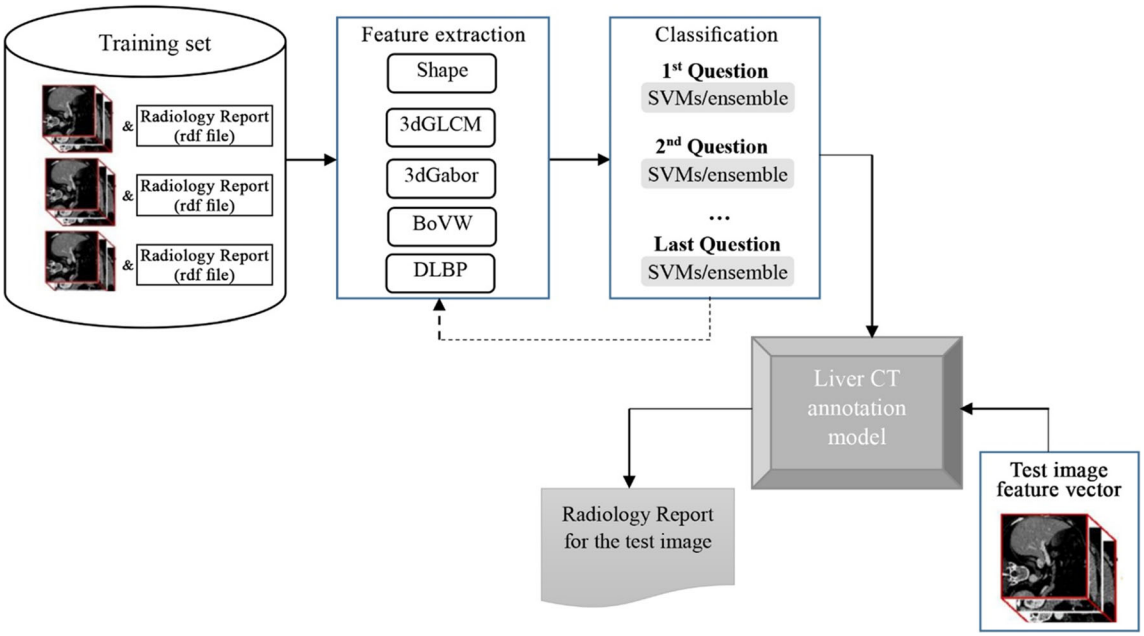

Fig. 1 Overview of the proposed liver CT annotation method

filters, co-occurrence matrix, and our proposed version of LBP. The proposed feature extraction method has been motivated by the procedure employed in [26]. It is called GOLD and is based on the gradient of the local binary patterns.

To find the answers and classify images, multi-class SVM and randomized-supspace k-nearest neighbor (RSKNN) have been used as the two back-end methods. Multi-class SVM classification has been utilized because SVMs are effective in high-dimensional spaces where the dimensionality is higher than the training sample size. In addition, employing different kernel functions increases their capability in successfully carrying out different classification tasks through paving the way for modeling complex non-linear decision borders. Moreover, they offer reasonable robustness due to the maximum margin criterion and can effectively handle data sparsity as well as class-imbalanced training data. These properties of the SVMs make them an optimal choice for our particular dataset, in which there is a wide variety of semantic labels with a small number of training samples.

Since in medical care, one of the ways to treat the patients is comparing the new cases with the previous patients and using experiments for assessment and diagnosis, one way to simulate this process with computers is to compare the new cases with the most similar samples in the annotated archives or datasets to facilitate the decision making process by leveraging the available data.

To reach this goal, the k-nearest neighbor (KNN) has been widely employed in medical applications, since it uses the most similar samples in the dataset for labeling or annotating the new inputs [27]. Moreover, empirical work has shown that combining classifiers can improve accuracy and reduce generalization error. Unfortunately, for the nearest neighbor classifiers, improvement via combination methods is not impressive [28]. This is because this method is very robust to the variations of a dataset, although it is sensitive to extracted features and distance metric. To investigate this issue, experimental results of combining RSKNN have been reported and relatively high performance has been achieved. Based on these results, we chose to use RSKNNs.

Feature extraction

Shape and texture

Shape and texture features play an important role in parameterizing the images, since they are indications for the human to identify and recognize real-world objects [29, 30]. We extracted region and shape descriptors including area, moments, circularity, solidity, weighted centroids, major and minor axis lengths, and eccentricity from liver images. For texture features, we have calculated the 3D gray-level co-occurrence matrix for four directions corresponding to 0, 45, 90, and 135° with a distance $d = 1$, and we extracted GLCM-based feature vector that includes 48 elements [31]. In addition, we used the mean and the standard deviation of the 3D Gabor wavelet transform as texture features, with different orientations and scales, so that this part of the feature vector includes 48 entries, 24 for the means, and 24 components for standard deviation.

BOVW

BOVW is a modified version of the bag of words (BOW) technique which is mainly used in the natural language

processing and image retrieval. In BOVW, each image and the corresponding features are treated as a document and words, respectively. Unlike documents which consist of words, images are not composed of discrete symbols. Therefore, a codebook containing the visual vocabulary (code words) is constructed by clustering features such as SURF (speeded up robust features). Having generated the dictionary, for visual vocabulary computation, the number of features reduced via quantizing the feature space by *k*-means clustering. To measure the frequency of the visual words in an image, a particular encoding scheme is employed, which produces a histogram that encodes the image into a feature representation. Finally, the classifier is trained using encoded training images from each category.

DLBP

As the radiologists employ the whole CT scan and check liver and lesions slice-by-slice (on original 2D CT slices) for analysis and detection, we try to produce a new feature extraction method that maintains multi-slice analysis and encodes the information about the correlation and variance of the attributes between slices. To reach this goal and develop a new feature extraction method, three critical points should be considered:

- How to extract appropriate features for effective parameterization of each slice?
- How to discover changes of these features between different slices of one CT image?
- How to analyze and resolve these changes for producing a feature vector that contains useful information about each CT image?

Our proposed front-end is motivated by [26], which presents texture descriptors called Gaussians of local descriptors. Our approach includes four steps, which is also depicted in Fig. 2 :

Step 1. Liver and lesion extraction: After preprocessing the 3D image, liver and lesion regions are extracted from the slices. The liver gradually appears in the slices. The liber region must be larger than a certain threshold to be detected.

Step 2. Extracting LBP descriptors: Assuming a regular grid over the 3D input image, the LBP descriptors are extracted from liver and lesion regions in each slice.

Step 3. Distribution estimation: Each region (liver or lesion) in a slice is represented as a histogram of rotation invariant local binary patterns (RI-LBP). These histograms are used to estimate the mean and covariance of a multivariate Gaussian distribution.

Step 4. Feature generation: The covariance matrix is treated as an image and hence image features such as texture descriptors are extracted from it.

The LBP feature that is extracted in step 2 is a method for image representation in which different patterns and window sizes are placed on the image pixels, and then the occurrences of the patterns are summarized in a histogram. For every pixel with grey value g_c, a neighborhood of P pixels of grey level, g_p, at the radial distance R is considered and the LBP descriptor defined as follows:

$$\mathrm{LBP}(P,R)=\sum_{p=0}^{P-1}\mathrm{sign}\left(g_p-g_c\right)2^p \tag{1}$$

where the sign function is as follows:

$$sign(x)=\begin{cases}0, & x<0\\ 1, & x\geq 0\end{cases}$$

Then, the extracted binary codes, or patterns, are represented as histograms. The patterns that were observed were divided into two categories: uniform and non-uniform. Therefore, in a more advanced definition of LBP, uniformity, U, of a pattern is defined by the number of spatial transitions between 0s and 1s in the pattern. For example, patterns 00000000_2 and 11111111_2 have U value of 0. If this value is less than or equal to the threshold U_T, the pattern is defined as uniform; otherwise, it is considered as non-uniform pattern. The number of possible uniform patterns that may occur in a circularly symmetric neighborhood of P pixels is $P + 1$ [32].

There is a modification on original LBP formulation in [32] to achieve rotation invariance. In this regard, a unique identifier was assigned to each rotation invariant LBP. The following formula underpins this mathematically:

$$\mathrm{LBP}_{P,R}^{r_i}=\min_i ROR\left(\mathrm{LBP}_{P,R},i\right) \tag{2}$$

Here, $ROR(x, i)$ performs a circular bitwise right shift on the P-bit number x, i times. In terms of image pixels, the above equation means rotating the neighbor set clockwise so many times that a maximal number of the most significant bits, starting from $g_P - 1$, is 0. The following operator for grayscale and rotation invariant texture description can be proposed instead of $LBP_{P,R}$:

$$\mathrm{LBP}_{P,R}^{r_iU_i}=\begin{cases}\sum_{p=0}^{P-1}\mathrm{sign}\left(g_p-g_c\right)2^p & \text{if } U<U_t\\ P+1 & \text{otherwise}\end{cases} \tag{3}$$

Here, superscript, r_iU_t, shows utilization of rotation-invariant uniform patterns that have U value less than or equal to U_T. In the step 3, if the extracted rotation invariant LBP (RI LBP) is the set RLH = {RLH1, …, RLHN} where the RLH is RI LBP histograms of n slices in a CT image, $\mu \in \mathrm{R_n}$ is the

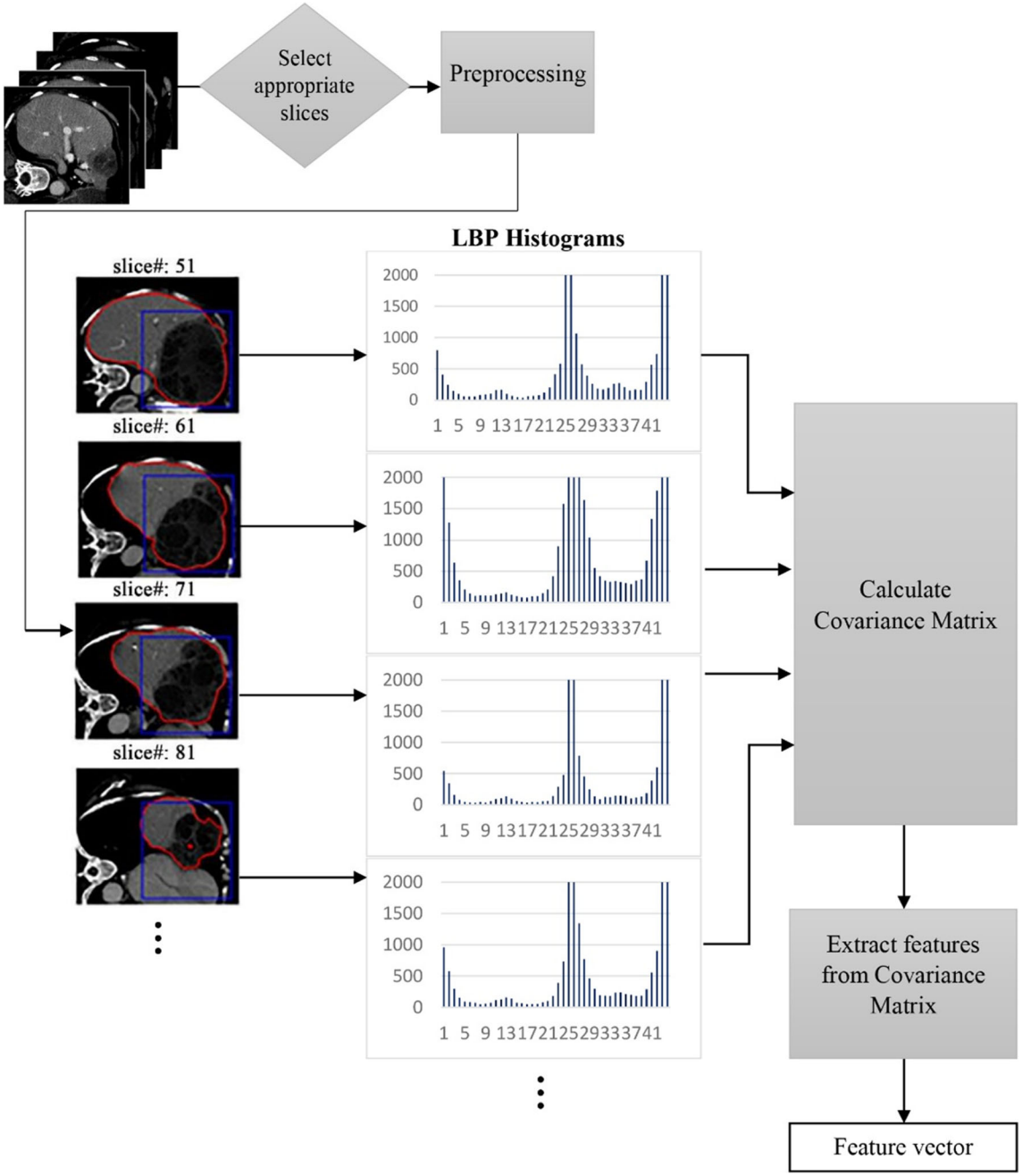

Fig. 2 Steps of the DLBP feature extraction

mean vector, and $C \in R_{n\times n}$ is the covariance matrix. The covariance matrix of a multivariate Gaussian distribution N belonging to the local descriptors of a 3D CT image can be defined as follows:

$$N(\mathrm{RLH};\mu,C)=\frac{1}{2\pi^{n/2}|C|^{1/2}}\exp\left(-\frac{1}{2}(\mathrm{RLH}-\mu)^{T}C^{-1}(\mathrm{RLH}-\mu)\right)$$
$$\mu=\frac{1}{N}\sum_{i=1}^{N}\mathrm{RLH}_i$$
$$C=\frac{1}{N-1}\sum_{i=0}^{N}(\mathrm{RLH}_i-\mu)(\mathrm{RLH}_i-\mu)^{T} \quad (4)$$

Finally, in step 4, after exploring several standard texture and visual features, RI LBP is used as the descriptor for the covariance matrix.

Classification

Support vector machines

SVMs solve binary classification problems via maximization of the margin separating two classes and kernelization enhances their capability in tackling non-linear problems [33].

The radial basis function (RBF), namely$K(x, y) = \exp(-\gamma\|x - y\|^2)$, is among the most popular and powerful kernels. The multi-class classifiers generated by binary SVMs using one-versus-one or one-versus-all approaches and the former are adopted here. Matlab®'s built-in library used to train the SVM classifiers [34].

The ensemble of random subspaces

To improve the accuracy of k-nearest neighbor classifiers, we have used random subspaces. Subspace ensembles also have the advantage of using less memory than ensembles with all predictors. The basic RS algorithm uses the following parameters:

- m is the number of dimensions (variables) to sample in each learner (i.e., the number of selected features in every subspace).
- d is the number of dimensions in the data, which is the number of columns in the data matrix X.
- n is the number of learners in the ensemble (i.e., number of KNNs in our method).
- k is the number of neighbors used in KNNs.

The basic RS algorithm performs the following steps:

1. Select, without replacement, m out of d available features, randomly.
2. Train a weak learner using just the selected features.
3. Repeat steps 1 and 2, n times to train n weak learners.
4. Find the average of the prediction scores for the weak learners and choose the class with the highest average score as the classification result.

For fitting the ensemble model, predictors that have been already chosen for a learner will be down-weighted. This decreases the chance of utilizing a predictor that was previously employed. Such a weighting scheme leads to distributing the predictors more evenly [35].

In order to build our question-answering or annotation system, we had to select the proper number of the nearest neighbors, then train an ensemble of KNN learners, and find an optimal size for subspaces for each question:

For choosing, k, the number of neighbors, KNN classifiers with different k values are trained. All of the extracted features are used in this stage. In the next step, an RSKNN is trained, and the appropriate subspace for every question is selected. This kind of ensemble has two advantages: First, the performance of the learned model in terms of the classification rate will be higher as demonstrated in the results. Second, we find the appropriate subspace for every question as well as the optimal size of the final model. As a result, the number of features to be extracted from the query image will be smaller.

Evaluation

In this section, we present the experimental setup that is used to evaluate the proposed feature extraction methods for medical image annotation and answering questions with

Table 1 Mean of 10-fold cross-validation accuracy and standard deviation

Feature set[1]	Dimensionality	Classifier	Accuracy (%)	Std
DLBP(Basic+LesionROI)	26 + 18 = 44	RS KNN	90.97	3.678
		SVM(RBF)	91.44	4.050
DLBP_4 block	4 × 18 = 72	RS KNN	90.79	4.095
		SVM(RBF)	91.58	3.530
Group1: Cooc3d+Gabor3d+Shape	48 + 48 + 36 = 132	RS KNN	90.67	3.762
		SVM(RBF)	89.44	8.787
Group2: DLBP(Basic+LesionROI) +DLBP_4block	26 + 18 + 72 = 116	RS KNN	90.55	3.185
		SVM(RBF)	91.58	4.109
Group3: Cooc3d+Gabor3d+Shape +DLBP_4block	132 + 72 = 204	RS KNN	**92.23**	4.840
		SVM(RBF)	87.02	6.841
Group4: Cooc3d+Gabor3d+Shape +DLBP(Basic+LesionROI)	132 + 44 = 176	RS KNN	91.72	3.774
		SVM(RBF)	85.62	5.505
Group5: Cooc3d+Gabor3d+Shape +DLBP(Basic+LesionROI) +DLBP_4block	132 + 26 + 18 + 72 = 248	RS KNN	90.28	4.393
		SVM(RBF)	86.42	7.278
The best feature for each question[2]			*93.14*	*3.013*

[1] All the mentioned features in this table are extracted from the CT image

[2] In answering each question, only features that had the highest accuracy for that question in the cross-validation were used

preliminary results on liver CT images. First, we present the learning algorithms and classifiers that were used in our experiments. Next, the procedure for the extraction and selection of the optimal visual descriptors is explained. We then define the feature combination schemes employed to improve the predictive performance of the image annotation. Finally, the experimental results are presented along with the discussion.

Experimental procedure

Given the size of the dataset and to achieve reliable results from generalization perspective, we have used 10-fold cross-validation to evaluate the accuracy of our annotation system. The experimental results are reported using the mean and standard deviation of the accuracy across 10-fold.

For classification, as mentioned earlier, SVM and RSKNNs are used. As shown in Table 5, some questions have binary (true, false) and some have multi-choice answers. For SVM, to solve the multi-class classification questions, we employ the one-vs-one (OvO) approach with different kernels, which, as will be shown in next subsection, the radial basis function (RBF) kernel achieves the best accuracy among all utilized kernels.

For tuning the parameters of the RSKNN, we repeated our experiments with different parameters. The number of dimensions to sample in each KNN was tuned for every question individually. The number of neighbors, k, in every KNN learner was tuned too (Fig. 4).

One of the major considerations in the annotation of the liver CT is the appropriate combination of features and classification methods. To investigate this issue thoroughly, we investigated different combinations of features and classifiers. For assessing the performance of the classifiers, the overall classification accuracy was used as the performance measure. It was calculated by comparing the automatic answers generated by each approach with the ground-truth labels. For each image, different feature sets were extracted and then the methods were tested on each set separately, as well as their combinations:

Table 2 The best accuracy obtained for every question with feature and classifier names

Question no.	Accuracy (%)	Feature	Classifier
1	94	Shape	SVM/RSKNN
4	55	BOVW	SVM
5	52	BOVW	SVM
6	64	DLBP(Basic+LesionROI)	SVM/RSKNN
8	72	Shape	RSKNN
9	82	DLBP(Basic+LesionROI)	SVM/RSKNN
10	68	DLBP(Basic+LesionROI)+Shape+Cooc3d+ Gabor3d	RSKNN
11	96	Shape	RSKNN
13	98	DLBP(Basic+LesionROI)	SVM/RSKNN
14	98	DLBP(Basic+LesionROI)	SVM/RSKNN
20	98	DLBP(Basic+LesionROI)	SVM/RSKNN
21	98	DLBP(Basic+LesionROI)	SVM/RSKNN
22	96	Shape	SVM/RSKNN
23	96	DLBP(Basic+LesionROI)	SVM/RSKNN
24	96	DLBP(Basic+LesionROI)	SVM/RSKNN
29	98	Cooc3d	SVM/RSKNN
30	100	Cooc3d	SVM/RSKNN
31	92	DLBP(Basic+LesionROI)	SVM/RSKNN
33	98	DLBP(Basic+LesionROI)	SVM/RSKNN
34	98	DLBP(Basic+LesionROI)	SVM/RSKNN
35	98	DLBP(Basic+LesionROI)	SVM/RSKNN
36	98	Cooc3d	SVM
37	92	DLBP(Basic+LesionROI)	SVM/RSKNN
38	94	Shape	SVM/RSKNN
39	94	Shape	SVM/RSKNN
40	98	DLBP(Basic+LesionROI)	SVM/RSKNN
41	90	Shape	SVM/RSKNN
42	92	DLBP(Basic+LesionROI)	SVM/RSKNN

Table 3 Feature types and their number of winning questions

Feature	BOVW	Cooc3d	Shape	DLBP
#Wins	2	3	7	16

- The feature set 1: This collection contains the 3D co-occurrence matrix (cooc3d) and 3D Gabor and shape features. In this case, the length of the feature vector is 132 (48 + 48 + 36).
- The feature set 2: This set includes the proposed feature extraction method (DLBP) utilized in three different scenarios; we will call them DLBP Basic, DLBP 4block and DLBP LesionROI.

1. For the DLBP Basic case, the following steps are employed to extract the proposed feature extraction method from each CT image:

   (a) Select relevant slices from each CT image based on the observed lesion in each slice. The number of the selected slices in each CT image ranges from 50 to 300.
   (b) Extract the liver using the liver contour in each slice.
   (c) Compute the rotation invariant LBP histogram (RILBP) with $P = 24$, $R = 3$ parameter setting using the whole image in every slice.
   (d) Accumulate the block histograms to obtain a matrix of LBP for every CT.
   (e) Calculate the covariance matrix from the histograms extracted in the previous step.
   (f) Extract the RILBP (with $P = 24$, $R = 3$) from the Covariance matrix. The feature dimensionality is 26.
2. In case of DLBP 4block, the images are resized into 128 × 128 and each slice is divided into four blocks of size 64 × 64 pixels. Then, DLBP basic features are extracted from each block separately and will be accumulated in one feature vector of size 72 (4 × 18).
3. For computing DLBP LesionROI feature set, again DLBP features are extracted, but from the lesion ROI, and the parameters of LBP are set to $P = 16$, $R = 2$, in which case the total dimensionality is 18.

- The feature set 3: This set only used for answering two types of questions about lesions. Evaluating the results showed that using this kind of features improves the accuracy in answering these two specific questions. SURF bag of visual words was done using k-means clustering with $k = 500$, which leads to total dimensionality of 500.

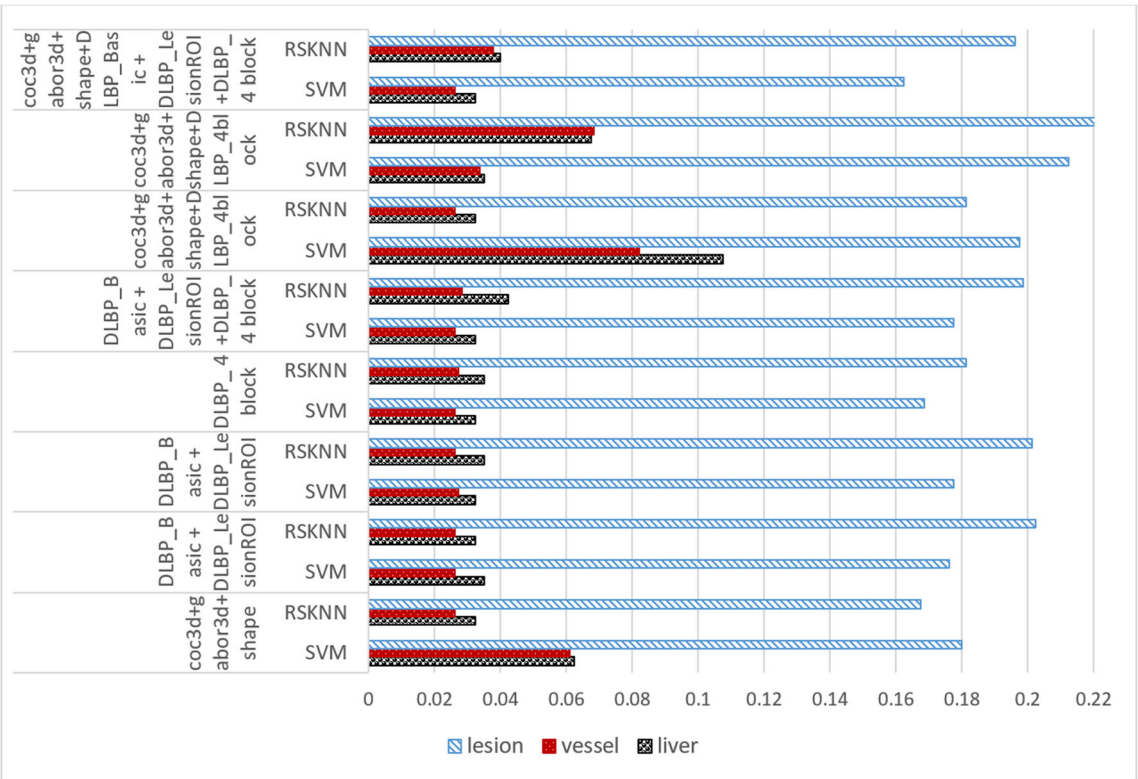

Fig. 3 Average classification error of different groups of questions using shared features

Results

Two different strategies are followed in designing the classifier models. The first and the most popular one learns the classifiers using a shared set of features for all the questions. This approach has been adopted in several related works (e.g., [17, 19, 36]). The second strategy is to design individual classifiers for the questions. This approach selects the best set of features for each question independently. Consequently, the recognition accuracy is at least as the previous strategy. However, the first strategy has lower training overhead and generalizes well in the practical scenarios. In the sequel, we will present results of both strategies in details.

Shared features for all the questions

Table 1 presents the average accuracy and standard deviation for 10-fold cross-validation in predicting answers for all the 43 questions. The results show the effect of using different feature extraction methods including the proposed parameterization method along with the two classification techniques.

Question-specific features

In this section, results of question-wise experiments are presented and analyzed. To be exhaustive, we report results of almost all the combinations of questions, classifiers, and features. Questions such as *AreahasAreaLengthFirst*, *AreahasAreaLengthSecond*, and *LesionhasLesionQuantity* are excluded from the results since they refer to the static size measurements and do not depend on the features and classifiers. Further, few questions take the same value for all the instances in the datasets. This issue has been mentioned in earlier works as a case of imbalanced sample problems (e.g., [19, 36]). Tables 6 and 7 report the detailed results. In these tables, *question number* refers to the number in the rightmost column of Table 5. In Table 6, results of using each individual set of features are presented. The single feature or combination leading to the highest accuracy for each question is reported in Table 2. Whenever two or more types of features get the same performance, we have reported the lower-dimensional features. All the results are more or less better than the aggregated accuracy (Table 1). Another fact that Table 2 reveals is that the proposed features (DLBP) got the highest rate in 16 out of 28 inspected questions (see Table 3).

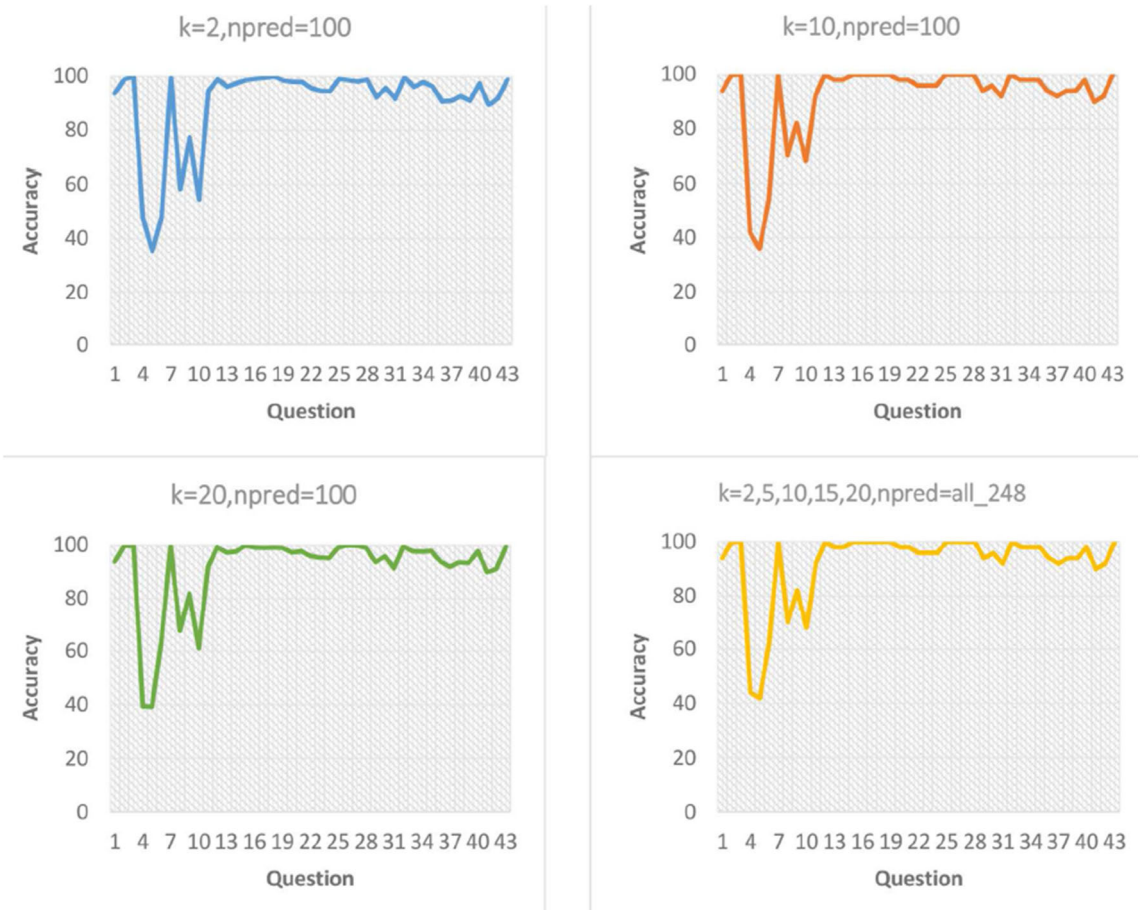

Fig. 4 Accuracy of the RSKNN model. k is the number of neighbours and npred is the size of the subspace of features.

Discussion

In this section, we focus on discussing two main contributions of this paper and their importance.

- The proposed approach offers a robust and computationally efficient technique for feature extraction with the potential to be scaled up and used for big data processing.
- The proposed method focuses on annotations rather than on diagnosis, which allows for greater coverage and flexibility [37].

A key challenge in medical imaging applications and analysis of 3D CT images which is addressed and investigated in this paper is to provide an efficient feature extraction method that characterizes different concepts in a data, like the liver properties, vessel, and lesions. Because of the wide range of questions that should be automatically answered with the proposed system, a variety of features were chosen.

One of the main novelties of this paper is the presentation of the DLBP feature which is based on the LBP histograms. Using the histogram makes the parameterization process computationally effective and makes the implementation relatively easy. Furthermore, despite relying on only low-level image representations, the information encoded in the histograms leads to the extraction of features with high expressivity power which results in high performance. In addition, since the spatial relations between image segments are not taken into account, the histogram-based features are rotation and translation invariant. This enhances the robustness of this kind of features. These properties also make them instrumental for liver lesions, as they can appear anywhere within the liver [37]. Another advantage of the proposed DLBP is that it processes every slice separately and explores through the whole 3D CT image and maintains multi-slice analysis.

Figure 3 depicts the classification errors for the proposed descriptors and feature extraction methods deployed in prediction, considering each group of questions about liver, vessel, and lesion concepts, separately (Table 5). The reported error is the average error of the questions in each group. Figure 4 shows the mean accuracy of the different configurations of the RSKNN method for individual questions, and Fig. 5 illustrates the same information for different SVM configurations. As Table 1 and Fig. 3 suggest, DLBP achieves higher accuracy and lower classification error than the combination of the shape and texture features (using Gabor filter and co-occurrence matrix), despite the fact that its dimensionality is relatively small, in comparison with alternative methods in medical applications [18, 19]. Such relatively low dimensionality, among others, makes this technique computationally efficient and could make the learning process easier. As seen in Table 1, the DLBP feature set works well along with both SVM and RSKNN back-ends.

For measuring the changes between the slices, instead of the covariance matrix, affinity matrix was used which leads to accuracy reduction. This implies that the covariance matrix plays a crucial role in detecting the changes in an image. The results reported in Table 1 show that the RSKNN outperforms SVM (higher accuracy and lower standard deviation) when the dimensionality increases. This is due to the fact that the ensemble system explores different feature subspaces and k values (for KNN) which ultimately results in finding an optimal setup resulting in a lower classification error (Fig. 4).

The best accuracy achieved by RSKNN is higher than the best performance yielded by the SVM. To further improve the performance, we tried to optimize the system for each question separately. This was carried out through using unique classification methods and features for each individual question. To this end, by exploring the performance for each individual question (Figs. 4 and 5), we extracted the best feature set and classifier for every question. Results show that the question-specific methods achieve slightly better accuracy, as compared to the shared features (Table 2).

One of the remaining challenges was slightly low accuracy in answering questions about the lesion concept. As shown in Fig. 3, the highest classification error is about 0.22. BOVW achieved higher accuracy for questions about the shape and margin of the lesion. This feature is extracted from the key points and has a higher capability to encapsulate the most important visual characteristics of each image [19]. Another advantage of BOVW is that, for extracting this feature, there is no dependence on segmentation; therefore, this feature set can be applied in purely automatic annotation approaches.

Table 4 shows the accuracy of the proposed method, compared to the best performing works on ImageCLEF liver CT annotation dataset. One of the major considerations in the results is the feature dimensionality, which shows that the proposed method in this paper outperforms the other alternative methods, although the number of features is relatively low. This confirms the claim that the proposed feature extraction method provides a more discriminative and informative representation for the CT images.

Conclusion

In this paper, a novel approach aiming to generate a structured radiological report to annotate liver CT images was presented. In contrast to the methods which simply carry out the annotation task based on the image modality or body regions, the proposed method annotates images with specific semantic contents about the liver, lesion, and vessels. The main contributions of this paper are the proposed feature extraction method (called DLBP) that incorporates multi-slice analysis in 3D CT images as well as finding the best performing method and feature set combination for answering and annotating each question separately.

Our experiments have been conducted on the ImageCLEF 2015 annotation dataset. Our best-achieved result is 93.1, which

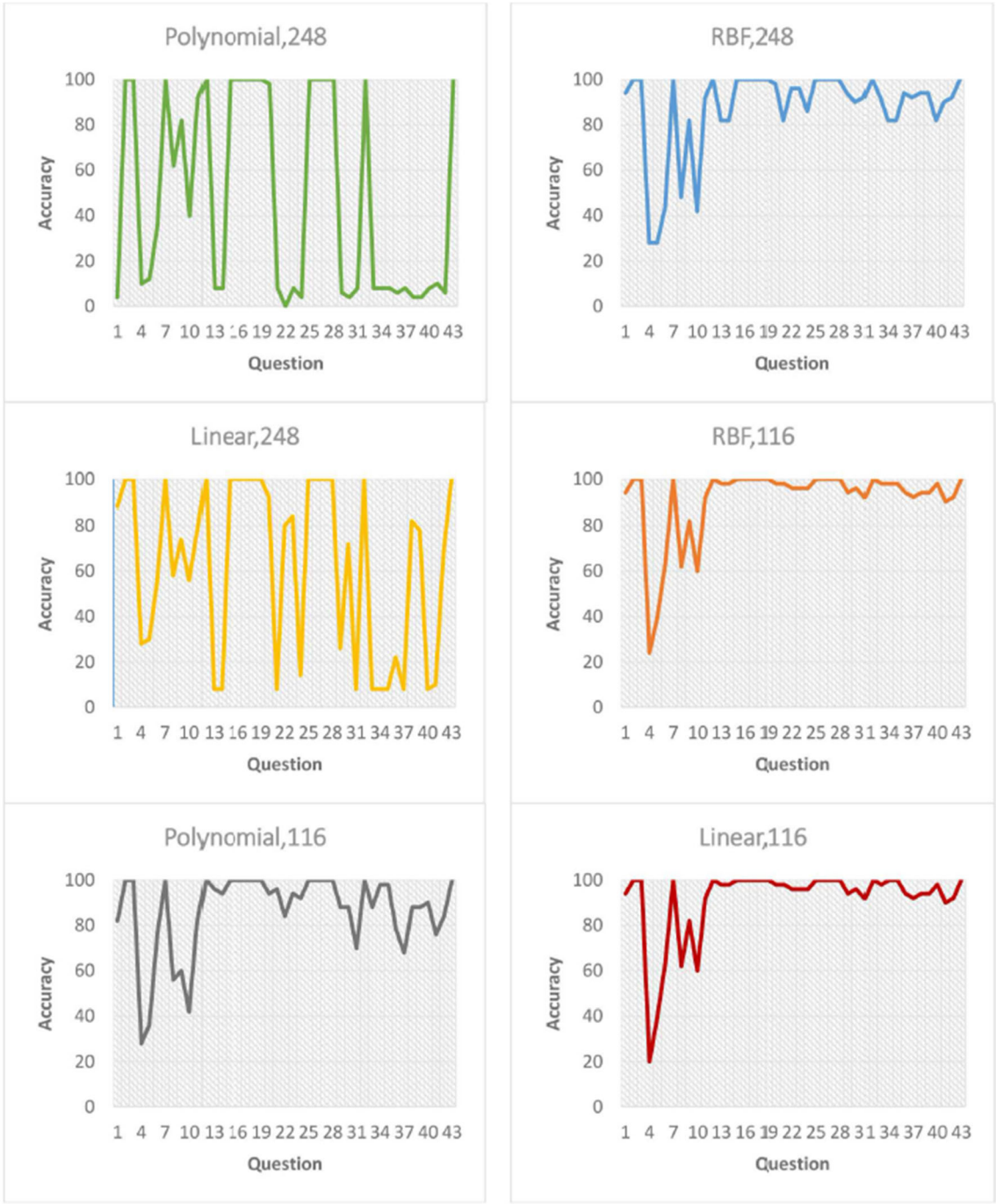

Fig. 5 Accuracy of the SVM on different questions using the proposed feature extraction method DLBP and combination of it with another type of features. Title of each chart is the type of the kernel and number of features

Table 4 Accuracy of the proposed method and the results presented by different authors on the ImageCLEF Liver CT annotation dataset

Ref	Feature	Dimension	Classifier	Accuracy (%)
[19]	SIFT	1000	Weighted nearest neighbor	88.7
[17]	GLCM + 3D Gabor	111	Random forest	84.0
[36]	Gabor of BIMFs	960	CBIR/majority voting	88.9
Our work selected feature for every question dimension of feature set varies from 36 to 500 for every question				93.1

is noticeably higher than the best reported result in the literature. Incorporating complementary information such as non-image information, e.g., the semantic distance between the related terms in an ontology [38, 39], is recommended for future work. Another avenue for future work is investigating the usefulness of the proposed handcrafted features in a DNN-based framework where the DNNs either serve as a classifier trained with these features (which may not be promising due to the lack of data) or combine the representations learned by DNNs with the handcrafted features along with the classic classifiers such as SVM.

Appendix

Table 5 List of liver CT annotations [18]

Group	Concept	Properties	Values	Question no.
Vessel	HepaticArtery	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	13
	HepaticArtery	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	14
	HepaticPortalVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	15
	HepaticPortalVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	16
	HepaticPortalVein	isCavernousTransformationObserved	NA (− 1), true (1), false (0)	17
	HepaticVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	18
	HepaticVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	19
	LeftHepaticVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	20
	LeftHepaticVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	21
	LeftPortalVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	23
	LeftPortalVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	24
	LeftPortalVein	isCavernousTransformationObserved	NA (− 1), true (1), false (0)	25
	MiddleHepaticVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	34
	MiddleHepaticVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	35
	RightHepaticVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	38
	RightHepaticVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	39
	RightPortalVein	hasLumenDiameter	Decreased (0), increased (1), normal (2), other (3)	41
	RightPortalVein	hasLumenType	Obliterated (0), open (1), partially obliterated (2), other (3)	42
	RightPortalVein	isCavernousTransformationObserved	NA (− 1), true (1), false (0)	43
Liver	LeftLobe	hasSizeChange	Decreased (0), increased (1), normal (2), other (3)	22
	RightLobe	hasSizeChange	Decreased (0), increased (1), normal (2), other (3)	40
	CaudateLobe	hasSizeChange	Decreased (0), increased (1), normal (2), other (3)	12
	Liver	hasDensity	Heterogeneous (0), homogeneous (1), other (2)	29
	Liver	hasLiverContour	Irregular (0), lobulated (1), nodular (2), regular (3),other (4)	30
	Liver	hasLiverDensityChange	Decreased (0), increased (1), normal (2), other (3)	31
	Liver	hasLiverPlacement	Downward displacement (0), normal placement (1), leftward displacement (2), upward displacement (3), other (4)	32
	Liver	hasSizeChange	Decreased (0), increased (1), normal (2), other (3)	33
Lesion	Lesion	hasLesionQuantity	1 (1), 2 (2), 3 (3), 4 (4), 5 (5), multiple (6)	26
	Lesion	LesionisDebrisObserved	True (1), false (0), NA (− 1)	27
	Lesion	LesionisLevelingObserved	True (1), false (0), fluid fluid (0), fluid gas (1), fluid solid (2), gas solid (3), other (4)	28
	Parenchyma	hasDensity	Heterogeneous (0), homogeneous (1), other (2)	36
	Parenchyma	hasParenchymaDensityChange	Decreased (0), increased (1), normal (2), other (3)	37
	Area	hasAreaDensity	NA (− 1), hyperdense (0), hypodense (1), isodense (2), other (3)	1
	Area	hasAreaLengthFirst	A number in mm which represents the width of the lesion	2
	Area	hasAreaLengthSecond	A number in mm which represents the width of the lesion	3
	Area	hasAreaMarginType	Geographical (0), ill defined (1), irregular (2), lobular (3), serpiginous (4), speculative (5), well defined (6), other(7)),	4
	Area	hasAreaShape	Band (0), fusiform (1), irregular (2), linear (3), nodular (4),ovoid (5), round (6), serpiginous (7), other (8)	5
	Area	hasDensityType	NA(− 1), heterogeneous (0), homogeneous (1), other (2)	6
	Area	isCalcified	True(1), false(0),NA(− 1)	7
	Area	isCentralLocalized	True(1), false(0)	8
	Area	isGallbladderAdjacent	True(1), false(0)	9
	Area	isPeripherical Localized	True(1), false(0)	10
	Area	isSubcapsular Localized	True(1), false(0)	11

Table 6 Accuracy of feature groups and classifiers for each question

Q no.	Cooc3d (48)		Shape (36)		Gabor3d (48)		DLBP1 (44)		DLBP1 (72)		BOVW
	SVM	RSKNN	SVM	RSKNN	SVM	RSKNN	SVM	RSKNN	SVM	RSKNN	SVM
1	94	94	94	92	94	92	94	94	94	92	
4	42	42	36	32	36	40	24	24	30	30	55
5	42	26	42	40	42	46	50	42	40	44	52
6	62	62	60	60	64	62	64	60	64	64	
8	62	60	62	72	62	58	62	60	62	58	
9	78	78	78	78	82	82	82	82	82	82	
10	50	52	60	60	60	52	60	52	60	56	
11	92	90	92	96	92	94	92	92	92	86	
13	92	94	92	96	92	98	98	98	98	94	
14	92	94	92	92	92	96	98	98	98	98	
20	94	92	92	92	92	98	98	98	98	98	
21	92	92	92	98	92	98	98	98	98	98	
22	96	96	96	96	96	96	96	96	96	96	
23	94	94	94	94	96	96	96	96	96	96	
24	94	94	92	94	96	96	96	96	96	96	
29	98	98	94	90	94	94	94	94	94	94	
30	100	100	96	96	96	96	96	96	96	92	
31	92	90	90	90	92	84	92	92	92	86	
33	92	98	92	94	92	98	98	98	98	98	
34	92	94	92	94	92	96	98	98	98	98	
35	94	94	92	94	92	96	98	96	98	98	
36	98	94	94	94	94	94	94	94	94	92	
37	92	90	90	90	92	90	92	92	92	90	
38	94	94	94	94	94	94	94	94	94	92	
39	94	94	94	94	94	92	94	94	94	92	
40	96	98	92	98	92	92	98	96	98	98	
41	90	90	90	90	90	86	90	90	90	86	
42	92	92	90	90	92	90	92	92	92	90	

DLBP1: DLBP(Basic+LesionROI)
DLBP2: DLBP_ 4 block

Table 7 Accuracy of feature groups and classifiers for each question

Q no.	Group1		Group2		Group3		Group4		Group5	
	SVM	RSKNN	SVM	RSKNN	SVM	RSKNN	SVM	RSKNN	SVM	RSKNN
1	94	94	94	94	94	94	94	94	94	94
4	36	40	28	38	28	42	28	38	24	42
5	42	42	46	42	22	46	30	32	18	42
6	60	60	64	58	64	56	64	62	64	56
8	62	72	62	56	62	70	58	66	62	66
9	82	82	82	82	82	82	82	82	82	82
10	60	60	60	50	56	66	58	**68**	56	68
11	92	96	92	92	92	92	92	92	92	92
13	92	96	98	98	92	98	2	98	2	98
14	92	92	98	98	92	98	2	98	2	98
20	92	96	98	98	98	98	98	98	98	98
21	92	92	98	98	92	98	92	98	92	98
22	96	94	96	96	96	96	96	96	96	96
23	96	94	96	96	96	96	96	96	96	96
24	96	94	96	96	96	96	96	96	96	96
29	94	92	94	94	88	98	94	94	88	94
30	96	96	96	96	4	96	84	96	4	96
31	92	90	92	92	92	92	92	92	92	92
33	92	92	98	98	92	98	92	98	92	98
34	92	96	98	98	92	98	92	98	92	98
35	92	96	98	98	92	98	92	98	92	98
36	94	94	94	94	88	94	94	94	88	94
37	92	90	92	92	92	92	92	92	92	92
38	94	92	94	94	94	94	94	94	94	94
39	92	92	94	94	94	94	94	94	94	94
40	92	92	98	98	92	98	92	98	92	98
41	90	90	90	90	90	90	90	90	90	90
42	92	92	92	92	92	92	92	92	92	92

Group1: Cooc3d+Gabor3d+Shape(132)

Group2:DLBP(Basic+LesionROI)+DLBP_4block(116)

Group3: cooc3d+gabor3d+shape+DLBP_4block(204)

Group4: cooc3d+gabor3d+shape+DLBP(Basic+LesionROI) (176)

Group5: cooc3d+gabor3d+shape+DLBP(Basic+LesionROI)+DLBP_4block(248)

References

1. Atam P Dhawan. *Medical image analysis*, volume 31. John Wiley & Sons, 2011.
2. Loveymi S, Shadgar B, Osareh A: An efficient approach to automated medical image annotation. *International Review on Computers and Software* 6(5):749–759, 2011
3. Tommasi T, Orabona F, Caputo B: Discriminative cue integration for medical image annotation. *Pattern Recognition Letters* 29(15): 1996–2002, 2008
4. Tatiana Tommasi, Francesco Orabona, and Barbara Caputo. An svm confidence-based approach to medical image annotation. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 696–703. Springer, 2008.
5. Tatiana Tommasi, Barbara Caputo, Petra Welter, Mark Oliver Guldld, and Thomas M Deserno. Overview of the clef 2009 medical image annotation track. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 85–93. Springer, 2009.
6. Dimitrovski I, Kocev D, Loskovska S, Dzeroski S: Hierarchical annotation of medical images. *Pattern Recognition* 44(10-11): 2436–2449, 2011
7. Demner-Fushman D, Antani S, Simpson M, Thoma GR: Annotation and retrieval of clinically relevant images. *International Journal of Medical Informatics* 78(12):e59–e67, 2009
8. Zehra Camlica, Hamid R Tizhoosh, and Farzad Khalvati. Medical image classification via SVM using LBP features from saliency-

based folded data. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 128–132. IEEE, 2015.

9. Riadh Bouslimi, Abir Messaoudi, and Jalel Akaichi. Using a bag of words for automatic medical image annotation with a latent semantic. *arXiv preprint arXiv:1306.0178*, 2013.
10. Alaidine Ben Ayed, Mustapha Kardouchi, and Sid-Ahmed Selouani. Rotation invariant fuzzy shape contexts based on eigenshapes and Fourier transforms for efficient radiological image retrieval. In *2012 International Conference on Multimedia Computing and Systems*, pages 266–271. IEEE, 2012.
11. Zare MR, Mueen A, Seng WC: Automatic medical x-ray image classification using annotation. *Journal of Digital Imaging* 27(1): 77–89, 2014
12. Mueen A, Zainuddin R, Baba MS: Automatic multilevel medical image annotation and retrieval. *Journal of digital imaging* 21(3): 290–295, 2008
13. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, NaveedAfzal SL, Zeng Y, Mehrabi S, Sohn S et al.: Clinical information extraction applications: a literature review. *Journal of biomedical informatics* 77:34–49, 2018
14. 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* 18(4):1363–1369, 2014
15. Gao H, Aiello Bowles EJ, Carrell D, Buist DSM: Using natural language processing to extract mammographic findings. *Journal of biomedical informatics* 54:77–84, 2015
16. Castro SM, Tseytlin E, Medvedeva O, Mitchell K, Visweswaran S, Bekhuis T, Jacobson RS: Automated annotation and classification of bi-rads assessment from radiology reports. *Journal of biomedical informatics* 69:177–187, 2017
17. Imane Nedjar, Said Mahmoudi, Mohammed Amine Chikh, Khadidja Abi-Yad, and ZouheyrBouafia. Automatic annotation of liver ct image: ImageCLEFmed 2015. In *CLEF (Working Notes)*, 2015.
18. Neda Barzegar Marvasti, Maria del Mar Roldan Garcia, Suzan Uskudarli, Jose Francisco Aldana Montes, and Burak Acar. Overview of the ImageCLEF 2015 liver ct annotation task. In *CLEF (Working Notes)*, 2015.
19. Kumar A, Dyer S, Kim J, Li C, Leong PHW, Fulham M, Feng D: Adapting content-based image retrieval techniques for the semantic annotation of medical images. *Computerized Medical Imaging and Graphics* 49:37–45, 2016
20. Zhiyun Xue, Sameer Antani, L Rodney Long, and George R Thoma. Automatic multi-label annotation of abdominal ct images using CBIR. In *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, volume 10138, page 1013807. International Society for Optics and Photonics
21. Depeursinge A, Kurtz C, Beaulieu C, Napel S, Rubin D: Predicting visual semantic descriptive terms from radiological image data: preliminary results with liver lesions in ct. *IEEE transactions on medical imaging* 33(8):1669–1676, 2014
22. Xu Y, Lin L, Hu H, Wang D, Zhu W, Wang J, Han X-H, Chen Y-W: Texture-specific bag of visual words model and spatial cone matching based method for the retrieval of focal liver lesions using multiphase contrast-enhanced ct images. *International journal of computer-assisted radiology and surgery* 13(1):151–164, 2018
23. Pol Cirujeda, Henning Muller, Daniel Rubin, Todd A Aguilera, Billy W Loo, MaximilianDiehn, Xavier Binefa, and Adrien Depeursinge. 3d Riesz-wavelet-based covariance descriptors for texture classification of lung nodule tissue in ct. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7909– 7912. IEEE, 2015.
24. Arai K, Herdiyeni Y, Okumura H: Comparison of 2d and 3d local binary pattern in lung cancer diagnosis. *Int J Adv Comput Sci Appl* 3(4):89–95, 2012
25. Zhang J, Xia Y, Xie Y, Fulham M, Feng DD: Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. *IEEE Journal of biomedical and health informatics* 22(5):1521–1530, 2018
26. Nanni L, Paci M, Brahnam S, Ghidoni S: An ensemble of visual features for Gaussians of local descriptors and non-binary coding for texture descriptors. *Expert Systems with Applications* 82:27–39, 2017
27. Gang Wang, David Forsyth, and Derek Hoiem. Comparative object similarity for improved recognition with few or no examples. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3525–3532. IEEE, 2010.
28. Carlotta Domeniconi and Bojun Yan. Nearest neighbor ensemble. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 1, pages 228–231. IEEE, 2004.
29. Castellano G, Fanelli AM, Sforza G, Torsello MA: Shape annotation for intelligent image retrieval. *Applied Intelligence* 44(1):179–195, 2016
30. Varish N, Pal AK: A novel image retrieval scheme using gray level co-occurrence matrix descriptors of discrete cosine transform based residual image. *Applied Intelligence* 48(9):2930–2953, 2018
31. Soh L-K, Tsatsoulis C: Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on geoscience and remote sensing* 37(2):780–795, 1999
32. Janis Fehr and Hans Burkhardt. 3d rotation invariant local binary patterns. In *2008 19th International Conference on Pattern Recognition*, pages 1–4. IEEE, 2008.
33. Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.
34. Wu T-F, Lin C-J, Weng RC: Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research* 5(Aug):975–1005, 2004
35. Ludmila I Kuncheva. *Combining pattern classifiers: methods and algorithms, 2nd Ed.* John Wiley & Sons, 2014.
36. Image Nedjar, Said Mahmoudi, and Mohammed Amine Chikh. Content-based medical image retrieval for liver ct annotation. *Transactions on Machine Learning and Artificial Intelligence*, 5(4), 2017.
37. Assaf B, Spanier NC, Sosna J, Acar B, Joskowicz L: A fully automatic end-to-end method for content-based image retrieval of ct scans with similar liver lesion annotations. *International journal of computer-assisted radiology and surgery* 13(1):165–174, 2018
38. Kurtz C, Depeursinge A, Napel S, Beaulieu CF, Rubin DL: On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. *Medical image analysis* 18(7):1082–1100, 2014
39. Spanier AB, Cohen D, Joskowicz L: A new method for the automatic retrieval of medical cases based on the RadLex ontology. *International journal of computer assisted radiology and surgery* 12(3):471–484, 2017