

# A Discriminative Distance Learning–Based CBIR Framework for Characterization of Indeterminate Liver Lesions

María Jimena Costa<sup>a</sup>, Alexey Tsymbal<sup>a</sup>,  
Matthias Hammon<sup>c</sup>, Alexander Cavallaro<sup>c</sup>,  
Michael Sühling<sup>a</sup>, Sascha Seifert<sup>a</sup>, Dorin Comaniciu<sup>b</sup>

<sup>a</sup> Siemens Corporate Technology, Erlangen, Germany

<sup>b</sup> Siemens Corporate Research, Princeton, USA

<sup>c</sup> University Hospital Erlangen

**Abstract.** In this paper we propose a novel learning–based CBIR method for fast content–based retrieval of similar 3D images based on the intrinsic Random Forest (RF) similarity. Furthermore, we allow the combination of flexible user–defined semantics (in the form of retrieval contexts and high–level concepts) and appearance–based (low–level) features in order to yield search results that are both meaningful to the user and relevant in the given clinical case. Due to the complexity and clinical relevance of the domain, we have chosen to apply the framework to the retrieval of similar 3D CT hepatic pathologies, where search results based solely on similarity of low–level features would be rarely clinically meaningful. The impact of high–level concepts on the quality and relevance of the retrieval results has been measured and is discussed for three different set–ups. A comparison study with the commonly used canonical Euclidean distance is presented and discussed as well.

**Keywords:** CBIR, distance learning, liver lesion, random forest, context–specific retrieval

## 1 Introduction

Possible causes for liver lesions are varied; they may originate from both malignant (e.g. metastases) or benign (e.g. cysts or hemangiomas) pathologies [1]. The characterization of these abnormal masses constitutes an essential task on which both diagnosis and the eventual treatment of the patient are based. Factors such as size, number, shape, margin definition or enhancement pattern in different contrast phases have a strong impact on the subsequent decisions.

A contrast agent is often administered to the patient, and several consecutive CT scans are then acquired within some minutes. The degree of contrast agent present in the liver at each acquisition’s time (i.e., the contrast agent phase) can have a severe impact on the appearance of abnormal masses. While the same lesion may look radically different in each of these consecutive CT images, two lesions originated by very different pathologies that pose different risks for the patient may have almost identical appearance in some of them (see Figure 1).

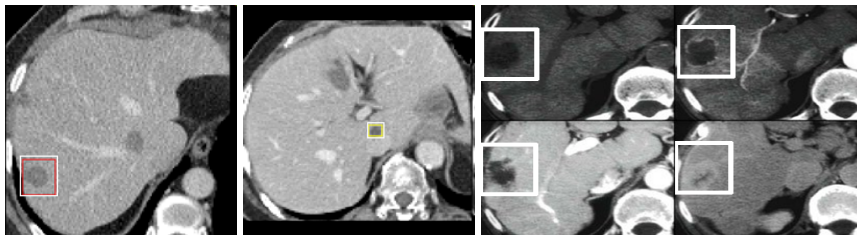


Fig. 1: While the two leftmost lesions have similar appearance, the leftmost is malignant (metastasis) while the one in the center is benign (cyst). The rightmost image shows the varying appearance of the same lesion in 4 different contrast agent phases. In such scenarios, a search relying solely on image-based similarity criteria (i.e. appearance) would probably fail to yield meaningful results.

While similar lesion retrieval can be a powerful decision support tool, searches based solely on appearance criteria would seriously limit the clinical meaningfulness of the retrieved results. Much more relevant results can be produced if the retrieval system is able to respond to user requests such as "Find lesions with comparable benignancy, given that the input lesion is in Portal Venous phase" or "Find lesions that are likely to be of the same type given that the input lesion is located in liver segment 4a and that it has a defined margin". The possibility to specify a desired *context* (i.e. similarity criterion, e.g. "comparable benignancy" or "similar margin definition") as well as additional *semantic* (i.e. *high-level*) features or concepts (e.g. "given that the input lesion is in Portal Venous phase" or "given that the input lesion is located in liver segment 4a") play an important role when tailoring the results to the user expectations.

Towards this goal, we introduce a fast, flexible, learning-based semantic CBIR framework that can accommodate combinations of appearance-based and user-defined semantic similarity criteria in order to retrieve the lesions most relevant in the given clinical context. Furthermore, once the learning phase has been completed, the method does not rely on previously annotated volumes, which allows for a fast and simple addition of new volumes to the search database if needed.

The outline of our method (the application phase) is as follows:

- 1. ROI selection**

The user selects a region of interest within an image, containing a structure (e.g. liver lesion) that he wishes to use as input for the search.

- 2. Specification of semantic context(s) and additional features**

The desired relevance/similarity context(s) that are to be used for the search are specified (e.g. benignancy criterion, lesion type criterion, etc.). The user may also choose to add semantic information as input to further characterize the structure of interest (e.g. margin enhancement or location within the liver).

- 3. Image-based feature computation**

A set of appearance-based features is computed on the selected region.

#### 4. Distance/Similarity computation

A set of Random Forest classifiers associated with the chosen similarity context(s) computes distances (i.e. similarity measures) between the input lesion (characterized by the given high–level concepts and low–level features computed) and the samples contained in an image database.

#### 5. Ranking and presentation of results

Results are grouped and ordered based on the desired combination of criteria. A tailored, ranked list of retrieval results is finally produced in the form of an *html* report.

The complete workflow has been tested in three different context scenarios: lesion density, lesion benignancy, and lesion type, and the impact of additional input semantic concepts has been measured, with promising results.

## 2 Related Work

According to Akgül et al. [2], radiology images pose specific challenges compared with images in the consumer domain; they contain varied, rich, and often subtle features that need to be recognized in assessing image similarity. Moreover, radiology images also provide rich opportunities for CBIR: rich metadata about image semantics are often provided by radiologists, and this information is not yet being used to its fullest advantage in CBIR systems.

In a closely related work, Napel et al. [3] study how low–level computer–generated features can be combined with semantic annotations in order to improve automatic retrieval of CT images of liver lesions. Combining the features has been shown to improve the overall performance in terms of discriminating lesion types. The study has been carried out on 10 cystic and 13 metastatic volumes to evaluate one context (lesion type); 95% mean retrieval accuracy is reported. We, on the other hand, use 1103 annotated metastases and 98 cysts in our study, and test 2 additional contexts, obtaining comparable accuracy results.

Automated decision support and lesion characterization are important in particular for small indeterminate lesions where uncertainty in characterization and treatment planning may remain even after collective discussion in a multidisciplinary setting [4]. In comparison to [3], we also consider the performance with small sub–centimeter lesions separately.

A measure of image similarity can be improved through the use of classification models that categorize new instances using a training set of instances with high–level semantic annotations, which constitutes a promising attempt to close the so–called *semantic gap* between the content–based description of an image and its meaning [5]. Segal et al [6] demonstrate that high–level information extracted from imaging for liver lesions may include valuable information, and in particular may reconstruct as much as 78% of the global gene expression profiles, revealing cell proliferation, liver synthetic function, and patient prognosis.

Seifert et al. [7] propose a learning–based ”search by scribble” system and apply it to the retrieval of similar liver lesions in 3D CT scans. In this case, the search context is fixed and based on pairwise similarity scores annotated for 160 hepatic lesions, with the similarity criterion left to the discretion of the

annotator. Additional information about the lesion’s anatomical location (i.e. the liver) is used to eliminate retrieval results that are located in other structures (e.g. in the spleen). The reported mean average precision is of 88%.

Although most statistical pattern recognition techniques are based on the Euclidean distance, one often requires non-Euclidean and non-metric similarity especially when contours, spectra or shapes are compared, for robust object matching [8]. Two non-metric distance measures based on *discriminative distance learning* has been introduced in [9]; the intrinsic Random Forest similarity and learning from equivalence constraints, with an application to anatomy surface mesh retrieval for decision support, among other applications including retrieval of clinical records, microarray gene expression and mass spectroscopy data.

The *intrinsic RF similarity* can be used for different tasks related to the classification problem. Thus, Shi et al. [10, 11] successfully use it for hierarchical clustering of tissue microarray data. First, the unlabeled original data are expanded with a synthetic class of evenly distributed instances, then an RF is trained to discriminate the original instances and the synthetic class, so that the intrinsic RF similarities can be determined and instances clustered. The resulting clusters are shown to be clinically more meaningful than the Euclidean distance based clustering with regard to post-operative patient survival. Hudak et al. [12] use the RF similarity for nearest neighbour imputation on forestry sensor data. They conclude that the RF distance based imputation is the most robust and flexible among the imputation techniques tested. It is interesting that using this similarity for the most immediate task, case retrieval and nearest neighbour classification or regression, is rather uncommon compared with its use for clustering. In one of related works, Qi et al. [13] use it for protein–protein interaction prediction, and the results compare favourably with all previously suggested methods for this task. In one imaging application, Vitanovski et al. [14] study the use of RF similarity-based regression for similar pulmonary trunk model retrieval in order to reconstruct the full surface of pulmonary trunk from incomplete sparse MRI data.

RF has been successfully applied to a number of different imaging tasks over the last several years. However, to the best of our knowledge, we are the first to consider the intrinsic RF similarity in the context of CBIR in this paper.

## 3 Method

### 3.1 Intrinsic RF similarity

For a Random Forest trained for a certain classification problem, the proportion of the trees where two instances appear together in the same leaves can be used as a measure of similarity between them [15]. For a given forest  $f$  the similarity between two instances  $x_i$  and  $x_j$  is calculated as follows. The instances are propagated down all  $K$  trees within  $f$  and their terminal positions  $z$  in each of the trees ( $z_i = (z_{i1}, \dots, z_{iK})$  for  $x_i$ , similarly  $z_j$  for  $x_j$ ) are recorded. The similarity between the two instances then equals to:

$$S(x_i, x_j) = \frac{1}{K} \sum_{k=1}^K I(z_{ik} = z_{jk}) \quad (1)$$

where  $I$  is the indicator function. When dissimilarity or distance is needed and not a similarity (e.g., for clustering or multi-dimensional scaling) it is normally calculated as suggested by Breiman [15]:

$$D(x_i, x_j) = \sqrt{1 - S(x_i, x_j)} \quad (2)$$

The intrinsic RF dissimilarity (equation 2) is known not to be metric [9], as the triangular inequality (equation 3) is often violated:

$$\forall x_i, x_j, x_k : D(x_i, x_j) \leq D(x_i, x_k) + D(x_k, x_j) \quad (3)$$

Several reasons motivate the choice of learning algorithm in our framework. First, RF was demonstrated to work well with and be robust to high-dimensional data with many weakly relevant, redundant and noisy features, without the need for additional data pre-processing and feature selection. Next, RF-based models are relatively fast to train and to apply comparing for example with Support Vector Machines. Then, RFs can be trained both for classification and regression problems, support supervised learning from multiple categories, and can easily handle missing values. Last but not least, they are able to provide the intrinsic RF similarity, which helps to combine the discriminative power and robustness of RFs with the transparency of case retrieval and nearest neighbour classification or regression.

Thanks to the appealing properties of RF, the RF similarity (1) can be easily calculated for different tasks, including classification and regression problems, tasks with heterogeneous feature vectors with possible missing values, and multiclass tasks.

### 3.2 Online RF

In order to speed up our extensive experiments, make it possible to validate more trends and make the models adaptive to learn from new samples we implement an incrementalization of RF similar to Saffari [16]. While successful lossless incrementalizations exist for many learning algorithms, most strong techniques applied in real applications are still difficult to incrementalize, and among them perhaps the most prominent example that received considerable attention lately is the ensemble of randomized decision trees, the most famous representative of which is the Random Forest algorithm [15]. A few algorithms for online ensembles of randomized decision trees have been recently proposed, and their application to vision tasks (in particular tracking) have been considered, despite the fact that they are not lossless and often require considerably more training cases than the corresponding batch technique in order to converge.

In particular, the online RF of Saffari et al. [16] trains decision trees of fixed depth and has a fixed structure which does not change with the observation

of new cases once the tree depth limit is reached. We address this issue in our online RF algorithm with

- The use of primed off-line learning to speed up convergence to a reasonable accuracy.
- Different sources of randomness (including bagging modeled with Poisson distribution of instance weights, and a random sample of observed features at each node).
- Memory management to avoid exceeding a specified memory limit for the model.
- Restructuring of the trees according to observed changes in the data distribution.

The training process starts with constructing a canonical RF [15] with 100 trees. Each decision tree is subsequently refined using online training. Online learning allows the framework to be scalable to the number of training instances. For each tree, a random sample of 200 lesions with even class distribution is used for the primed training. Primed off-line training is a simple but effective technique to improve the predictive performance of the final model (see [17] for an example).

We exploit the memory management scheme proposed before for online Hoeffding trees that dynamically activates most promising nodes, for tracking feature distributions and a split attempt, and deactivates and removes the less promising ones [18]. Similar to Saffari et al. [16] and different to the Hoeffding tree, a split is simply generated after observing a certain specified number of instances (40 is the default value normally leading to best performance). For each feature, Gaussian distribution is assumed and is tracked online, and a split threshold value which maximizes the *Gini Index* value is selected.

### 3.3 Low-level liver lesion descriptors

Each liver lesion in our experimental setting is described by a set of low-level computer-generated imaging features as follows:

- Relative frequency histogram and four first central normalized moments on it for the Hounsfield Units (HU) distribution in the bounding box for the lesion;
- The set of eight invariant Hu moments of order up to 3 [19];
- Six invariant Zernike moments [20];
- HU histogram and the four first central moments for the whole liver;

The first three feature types describe the lesion itself, while the last one describes the image of the whole liver. Inclusion of whole liver features has been shown to always lead to an improved discrimination performance in our set of experiments.

We use 2D Hu and Zernike invariant moments [19, 20]. In order to adapt them to the 3D liver lesion, for each lesion we generate 3 orthogonal 2D cuts intersecting at the center of the lesion ROI. The invariant moments are then calculated for each cut, and the feature vector includes both the moment for each separate cut and the averaged moments.

### 3.4 High-level concepts

Each liver lesion in a CT scan has been also annotated by 2 clinical experts with a set of 20 semantic labels. Some of these descriptors correspond to or are similar to features identified in Segal et al. [6] and Napel et al. [3]. Semantic features identified in these two studies were used as inspiration during our meetings with the clinical partners.

While some features are relatively inter-observer stable (lesion margin, rim continuity), some discrepancies arise concerning mostly lesion type (i.e. diagnosis). In such cases, information from additional studies (e.g. MRI reports or biopsies) has been used to resolve differences.

The semantic features used in our studies are as follows:

- Contrast agent phase (Native, Arterial, Portal Venous, Late, N/A);
- Lesion focality (Single, Multiple, N/A);
- Lesion surrounding (Complete, Incomplete, Absent, N/A);
- Rim continuity (Continuous Bright Rim, Discontinuous Bright Rim, Continuous Dark Rim, Discontinuous Dark Rim, N/A);
- Margin (Regular, Irregular, N/A);
- Margin definition (Defined, Diffuse, N/A);
- Lesion density (Hypodense, Hyperdense, N/A);
- Benignancy (Benign, Rather Benign, Malignant, Rather Malignant, N/A);
- Lesion type (Cyst, HCC, Hemangioma, Metastasis, N/A);

While benignancy, lesion type and density annotations were used as class labels to train the RFs associated with each search context, the other annotated high-level features were tested as additional semantic input during the training phase of the models. After the models have been trained, at the retrieval stage, both the low-level features computed for the given lesion ROI and the high-level feature(s) are provided as input to the RF. Figure 2 in the next section shows a simple example illustrating the system workflow.

## 4 Framework Application to Similar Hepatic Lesion Retrieval

We apply the proposed retrieval framework to the search of similar hepatic lesions. The scheme shown in Figure 2 illustrates the search for hepatic lesions that have comparable benignancy given that the lesion specified in the ROI is known to have a defined margin.

In Figure 2, the input sample is represented by both a set of low-level features computed from the given ROI and by the given additional high-level concepts (in the example "Defined Margin"). The input sample is then run through the RF associated to (and trained for) the desired context ("Benignancy"). The database samples that occur in the same terminal nodes as the input sample will be noted and later ranked according to the frequency of co-occurrence with the input sample.

For hepatic lesion retrieval, three search contexts have been tested: lesion density, benignancy and type. For each of them, an associated Random Forest

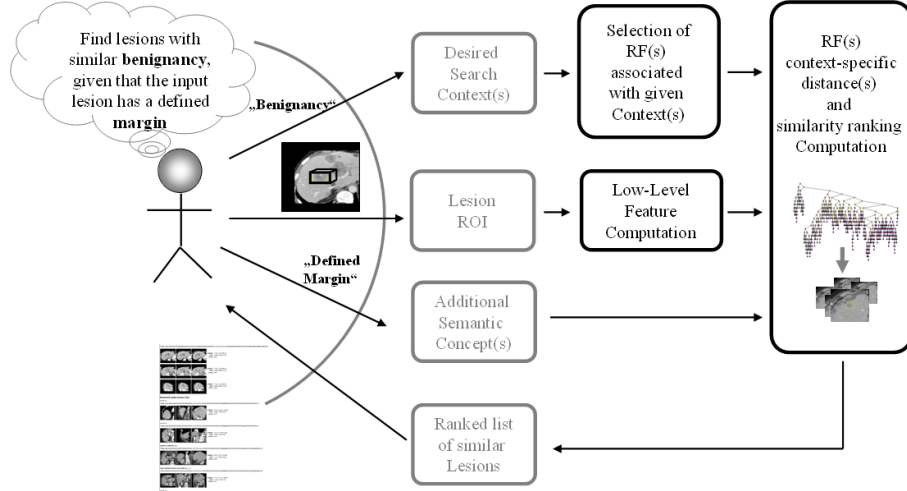


Fig. 2: Overview of the framework with an example application to the search of hepatic lesions with comparable benignancy (context), given that the lesion specified in the ROI has a defined margin (semantic concept or high-level feature).

has been trained with the following annotated lesions from 523 3D CT scans corresponding to 244 patients:

**Density** (Hypodense, Hyperdense): 1948 hypodense and 358 hyperdense.

**Benignancy** (Benign, Malignant): 761 malignant and 93 benign.

**Type** (Cyst, Metastasis): 1103 metastases and 98 cysts.

Our experimental setting is flexible enough to accommodate any combination of high-level and low-level features, and has been tested in the 3 aforementioned contexts in a *Leave-One-Patient-Out* (LOPO) evaluation. We assess the retrieval results in the next section.

## 5 Empirical Study

We evaluate the retrieval performance of our framework in the three specified contexts, in terms of discrimination accuracy of the RF-based distance learnt. We use LOPO ROC AUC (*Area Under the Curve*) values to evaluate the ultimate performance. The intrinsic RF similarity is compared with the canonical Euclidean distance. Both for the intrinsic RF similarity and the Euclidean metric, predictive performance is measured via  $k$ -nearest neighbor classification, with  $k=7$  and with weighting the votes of the neighbors proportional to the RF similarity or inversely proportional to the Euclidean distance. This parameter setting has been shown to always lead to competitive results. For the Euclidean distance all feature values were also scaled to lie in the same range  $[0,1]$ .

For each of the 3 search criteria and each representation of the input sample (ROI), Table 1 shows the ROC AUC measure of the retrieval results with the RF similarity, and Table 2 shows similar results for the Euclidean metric. The ROC curves pertaining to RFs trained in the three different contexts are



shown in Figure 3. Table 3 shows semantic feature importance, i.e., the average appearance of a semantic feature given as input in a RF tree. Figure 4 illustrates the top 5 ranked results for 2 different input lesions.

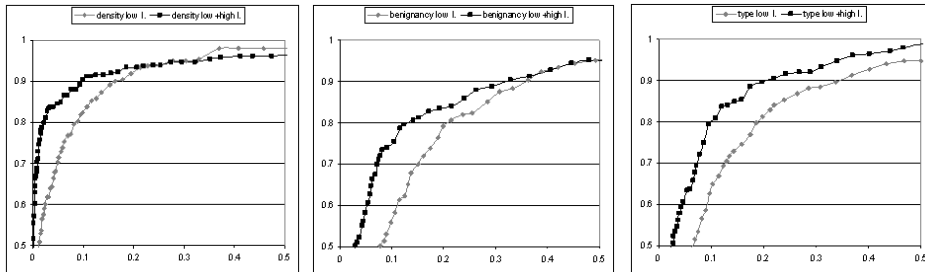


Fig. 3: ROCs for the training of the RFs for each context. From left to right, Density, Benignancy and Type context ROCs of the corresponding trained classifiers. In each case, the grey line shows the curve with the low–level features only, while the black line shows the curve when additional high–level features are included as input.

An additional high–level feature was tested in all three contexts: the location of the lesion (i.e. the segment number within the liver) was given as additional input to the framework. However, in all 3 contexts the feature had little or no influence on the results, which seems to suggest that the location of a lesion has little influence on its benignancy, density or type (a reasonable assumption).

Table 1: Quantitative evaluation of retrieval results in the three contexts: Lesion Density, Benignancy and Type using, in each case, low–level features alone or a combination of low– and high–level features. The table shows the results for all lesion sizes combined, and sub–centimeter lesions are studied in the two rightmost columns.

context	All Lesions		Lesions $\leq 1000mm^3$	
	low l.	(low+high) l.	low l.	(low+high) l.
Density	0.94	0.933	0.959	0.955
Benignancy	0.854	0.891	0.732	0.855
Type	0.872	0.906	0.755	0.838
Average	0.889	0.91	0.828	0.8827

Table 2: Retrieval results with the Euclidean distance metric. Table has the same structure as Table 1.

context	All Lesions		Lesions $\leq 1000mm^3$	
	low l.	(low+high) l.	low l.	(low+high) l.
Density	0.55	0.543	0.525	0.542
Benignancy	0.72	0.713	0.655	0.655
Type	0.675	0.650	0.625	0.618
Average	0.648	0.635	0.601	0.605

Table 1 shows the results obtained using the intrinsic RF similarity. It is interesting to observe that the low–level features alone produce very reasonable retrieval results in the lesion density context, which is an inherently “low–level” task. High–level features seem indeed to have less impact in this context. How-

ever, when higher level clinical contexts such as benignancy or lesion type are considered, the positive impact of the additional high-level features given as input becomes more prominent, and results in a better accuracy and more relevant retrieval results. This benefit is particularly visible in the case of sub-centimeter lesions, which are often very hard to characterize even for the expert eye. In this case, the addition of one or more high-level features has a significant influence on the accuracy and meaningfulness of the results. The difference in performance in these cases (low-level features versus low- plus high-level features in the benignancy and lesion type contexts) is always significant according to the McNemar’s test on difference between two proportions conducted for the specificity of interest (0.9). In a clinical setting where small lesions are difficult to assess, the proposed framework has great potential to become a significant decision support tool.

In the same 3 contexts we have also evaluated the Euclidean distance metric. The obtained results, shown in Table 2, are clearly inferior to those produced by the intrinsic RF similarity (the difference is always significant). Moreover, the Euclidean distance is not able to benefit from the addition of the high-level features, as opposed to the RF similarity, and the performance is often even worse after the addition of high-level features. The performance is especially poor in the lesion density context.

Table 3: *Importance of the input high-level concepts for retrieval.*

Context	High-level feature given as additional input						Average
	Contrast Agent Phase	Focality Surrounding	Rim	Margin	Margin Definition		
Density	1.358	1.393	1.443	1.081	1.052	0.964	1.215
Benignancy	1.801	1.641	1.771	1.091	2.293	1.89	1.749
Type	1.501	1.371	1.559	0.936	1.926	1.641	1.489
Average	1.553	1.468	1.591	1.0359	1.757	1.50	

The time required for retrieval, from low-level feature computation to running the instance through the appropriate RF to similarity calculations and lesion sorting to the output of the ranked list of lesions, is less than a second on a standard PC. Although we consider three isolated user-specified search contexts in this study, a certain number of them can be easily combined when needed, for instance in order to address a query "Find lesions that are likely to be of same lesion type and have comparable density".

## 6 Conclusion and Future Work

We have presented a fast, flexible framework for semantic CBIR based on discriminative distance learning. The speed of retrieval, promising retrieval performance especially for difficult, small lesions, and the flexibility to define relevance criteria combinations with a considerable positive impact on the quality of the results open the door for higher level searches that yield results that are meaningful to the the user in any given context. The addition of semantic features as input for search consistently improves the retrieval results, with a particularly positive impact on small lesions, making results for all lesion sizes comparable in terms of quality and meaningfulness to the user.

We believe the framework presented in this paper has a large potential, as

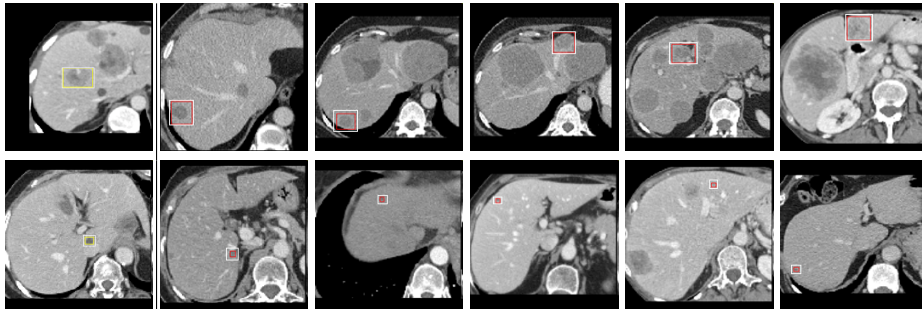


Fig. 4: Results of the search for similar lesions in the benignancy similarity context. Top row: A malignant lesion used as input (leftmost image), and the top 5 retrieved lesions, all of which are malignant as well, even though their appearances are not always comparable. Bottom row: Input benign lesion (leftmost image) and top 5 retrieval results, 4 of which are benign as well.

well as room for improvement. It would be interesting to allow not only the manual input of semantic features by the user, but also their automatic prediction, before they are used for retrieval. Using RF for regression tasks as well as multiclass–classification would allow the inclusion, for instance, of more lesion types to the similarity search. The presentation of results in the *html* report can also be complemented with more detailed statistics on the quality of the search results, which would allow the user to use the tool in a more informed way when making decisions.

Benefits of this framework include inconsistency checks in radiology reports (i.e. as a second opinion with respect to the reported diagnosis, benignancy, or any other lesion characteristics) and abnormality or relevance pointers highlighting critical lesions. In the future, the system will be combined with automatic lesion detection as a preprocessing step. This allows presenting the detected lesions to the radiologist in the form of a ranked list upon opening the case, prioritizing, for instance, the most likely malignant lesions. This will render diagnostic reading even more efficient by speeding up the experts review of the ever increasing amount of image data.

## References

1. Lencioni, R., Cioni, D., Bartolozzi, C., and Baert, A., [*Focal Liver Lesions: Detection, Characterization, Ablation*], Springer (2005).
2. Akgül, C., Rubin, D., Napel, S., Beaulieu, C., Greenspan, H., and Acar, B., “Content–based image retrieval: current status and future directions,” *Journal of Digital Imaging* (2010).
3. Napel, S., Beaulieu, C., Rodriguez, C., Cui, J., Xu, J., Gupta, A., Korenblum, D., Greenspan, H., Ma, Y., and Rubin, D., “Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results,” *Radiology* **256** 1 (2010).
4. Spencer, J., “Indeterminate lesions in cancer imaging,” *Clinical Radiology* **63**, 843–852 (2008).

5. Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R., "Content-based image retrieval at the end of the early years," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **22**(12) (2000).
6. Segal, E., Sirlin, C., Ooi, C., Adler, A., Gollub, J., Chen, X., Chan, B., Matchuk, G., Barry, C., Chang, H., and Kuo, M., "Decoding global gene expression programs in liver cancer by noninvasive imaging," *Nature Biotechnology* (2007).
7. Seifert, S., Thoma, M., Stegmaier, F., Hammon, M., Kramer, M., Huber, M., Kriegel, H.-P., Cavallaro, A., and Comaniciu, D., "Combined semantic and similarity search in medical image databases," **7967**, 7967–2 (2011).
8. Pekalska, E., Harol, A., Duin, R., and Spillmann, B., "Non-euclidean or non-metric measures can be informative," in [*Proceedings International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, SSPR*], Lecture Notes in Computer Science, Springer (2006).
9. Tsymbal, A., Huber, M., and Zhou, S., "Learning discriminative distance functions for case retrieval and decision support," *Transactions on CBR* **3**(1), 1–16 (2010).
10. Shi, T., Seligson, D., Belldgrun, A., Palotie, A., and Horvath, S., "Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma," *Mod Pathol.* **18**(4), 547–557 (2005).
11. Shi, T. and Horvath, S., "Unsupervised learning with random forest predictors," *Computational and Graphical Statistics* **15**(1), 118–138 (2006).
12. Hudak, A., Crookston, N., Evans, J., Hall, D., and Falkowski, M., "Nearest neighbour imputation of species-level, plot-scale forest structure attributes from lidar data," *Remote Sensing of Environment* **112**(5), 2232–2245 (2008).
13. Qi, Y., Klein-Seetharaman, J., and Bar-Joseph, Z., "Random forest similarity for protein-protein interaction prediction from multiple sources," *Proceedings of Pacific Symposium on Biocomputing* (2005).
14. Vitanovski, D., Tsymbal, A., Ionasec, R., Georgescu, B., Zhou, S., and Comaniciu, D., "Learning distance function for regression-based 4d pulmonary trunk model reconstruction estimated from sparse MRI data," *Proc. SPIE Medical Imaging* (2011).
15. Breiman, L., "Random forests," *Machine Learning*, 5–32 (2001).
16. Saffari, A., Leistner, C., Santner, J., Godec, M., and Bischof, H., "Online random forests," *3rd IEEE ICCV Workshop on Online Computer Vision* (2009).
17. Oza, N. C. and Russell, S., "Experimental comparisons of online and batch versions of bagging and boosting," 359–364 (2001).
18. Pfahringer, B., Holmes, G., and Kirkby, R., "New options for Hoeffding trees," *Australian Conference on AI* (2007).
19. Hu, M., "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory* **8** (1962).
20. Pejnovic, P., Buturovic, L., and Stojiljkovic, Z., "Object recognition by invariants," *Proceedings of Int. Conf. on Pattern Recognition* (1992).