A Retrieval-Based Computer-Aided Diagnosis System for the Characterization of Liver Lesions in CT Scans

Peter Dankerl, MD, Alexander Cavallaro, MD, Alexey Tsymbal, PhD, Maria Jimena Costa, PhD, Michael Suehling, PhD, Rolf Janka, MD, Michael Uder, MD, Matthias Hammon, MD

Rationale and Objectives: To evaluate a computer-aided diagnosis (CADx) system for the characterization of liver lesions in computed tomography (CT) scans. The stand-alone predictive performance of the CADx system was assessed and compared to that of three radiologists who were provided with the same amount of image information to which the CADx system had access.

Materials and Methods: The CADx system operates as an image search engine exploiting texture analysis of liver lesion image data for the lesion in question and lesions from a database. A region of interest drawn around an indeterminate liver lesion is used as input query. The CADx system retrieves lesions of similar histology (benign/malignant), density (hypodense/hyperdense), or type (cyst/hemangioma/metastasis). The system’s performance was evaluated with leave-one-patient-out receiver operating characteristic area under the curve on 685 CT scans from 372 patients that contained 2325 liver lesions (193 <1 cm³). Sensitivity, specificity, and positive and negative predictive values were evaluated separately for subcentimeter lesions. Results were compared to those of three radiologists who rated 83 liver lesions (20 hemangiomas, 20 metastases, 20 cysts, 20 hepatocellular carcinomas, and 3 focal nodular hyperplasias) displaying only the liver.

Results: The CADx system’s leave-one-patient-out receiver operating characteristic area under the curve was 97.1% for density, 91.4% for histology, and 95.5% for lesion type. For subcentimeter lesions, input of additional semantic information improved the system’s performance. The CADx system has been proved to significantly outperform radiologists in discriminating lesion histology and type, provided the radiologists have no access to information other than the image. The radiologists were most reliable in diagnosing hemangioma given the limited image data.

Conclusions: The CADx system under study discriminated reliably between various liver lesions, even outperforming radiologists when accessing the same image information and demonstrated promising performance in classifying subcentimeter lesions in particular.

Key Words: Imaging technology; computer-assisted diagnosis; special interest; hepatobiliary.

Despite intense efforts to cure or to control cancer through advances in imaging, surgery, chemotherapy, and radiation therapy, treatment of most malignancies continues to be challenging. This is reflected, for example, by the fact that in the past 5 years, overall cancer deaths in the European Union (EU) have only decreased by 10% in men and 7% in women, and for liver cancer in particular, no decrease in the rates is expected in 2013 (1). Most patients die not because of the growth of the primary cancer but because of its spread to other sites. Various types of malignant primary tumors spread to the liver, which is the second most common site for cancer metastases. Liver metastases have been proved to significantly worsen the survival rate compared to patients without hepatic involvement (2). Furthermore, the confirmed presence of liver metastases compels the crucial choice of a suitable treatment, such as chemotheraphy, surgery, radiofrequency ablation (RFA), transarterial chemoembolization (TACE), or selective internal radiotherapy treatment (SIRT). Despite their pivotal importance, liver lesions are at risk of being missed by clinicians in images of the most commonly used modalities (2). Not only detection but also characterization and risk assessment are difficult, particularly for small lesions, further hindering prompt and personalized patient management.

Focal liver lesions are usually detected in routine computed tomography (CT) scans. Different types of liver lesions often display similar image features and general appearance, and
both hypodense and hyperdense lesions may have various benign and malignant differential diagnoses. Therefore, correctly classifying liver lesions is a challenging task that requires medical expertise, training, concentration, and time. In the era of thin slice imaging, radiological reading has become even more time consuming (3). With the rapid advances in the related technology, smaller, even subcentimeter, lesions can be identified on the scans; their characterization, however, remains challenging—many of them would be called indeterminate. It is well known that the interpretation of findings may suffer from interrreader variability (4–7) and is prone to error (8,9), especially when benign and malignant lesions display similar visual appearances (10). As demonstrated by Ganeshan et al (11), computer-based texture analysis might be one approach to assist in the diagnosis of such indeterminate hepatic lesions. Recently, Napel et al (12) suggested a radiological image search to improve the diagnosis of indeterminate liver lesions in CT examinations.

The goal of the present study is the evaluation of the performance of a user-friendly (clinically practicable) computer-aided diagnosis (CADx) system for liver lesion characterization. The evaluated system is based on texture analysis of liver and lesion tissues and operates like a radiological image search engine, with the purpose of clinical decision support in indeterminate liver lesion characterization.

As input, the radiologist supplies a region of interest (ROI) drawn around a liver lesion and has the opportunity to input additional semantic information (e.g., lesion fociality, margin continuity, etc) into the CADx system’s query. The system subsequently retrieves lesions with likely similar histology (benign/malignant), density (hypodense/hyperdense), and/or type (cyst/hemangioma/metastasis) from the database of annotated liver lesions (Fig 1). The key points in our study are whether the system is able to (automatically) accurately retrieve similar lesions for the considered tasks and whether the system’s stand-alone performance can outperform that of the radiologists, in particular when both the system and the radiologist have access to the same information.

MATERIALS AND METHODS

Institutional review board approval was obtained for retrospective analysis, and all procedures were in accordance with the Declaration of Helsinki.

Data Acquisition and Scan Technique

The radiological information system (RIS) of the University Hospital Erlangen was searched for radiological CT reports describing focal liver lesions in either consecutive CT scans or complementary examinations. We selected 372 patients (174 women; mean age 62.3 ± 9.8 years; 198 men, mean age 60.4 ± 8.4 years) with 685 CT scans for our research. All patients were examined using a 64-row multidetector CT scanner (Somatom 64; Siemens, Erlangen, Germany), the tube voltage was 120 kV, Care Dose (Siemens, Erlangen, Germany) was used, the pitch was 0.75, the collimation was (0.6 mm), and selected examinations consisted of 1-mm slices in the portal venous contrast agent phase and, if present, in the native and arterial phase. A weight-dependant dose of warmed contrast agent (Imeron; Bracco Imaging, Konstanz, Germany) was administered intravenously to all patients with a flow rate of 3 ml/sec. If indicated by a clinical issue, oral or rectal contrast agent (1.5 L of diluted Gastrografin; Bayer, Leverkusen, Germany) was given in accordance with the guidelines.

All scans were performed in the Department of Radiology at the University Hospital Erlangen under supervision of experienced radiologists.

Lesions in the Study

A total of 685 CT scans from 372 patients containing 2325 liver lesions in the portal venous phase were included in our study. From these, 193 lesions were subcentimeter (their volume was less than the volume of a sphere with a diameter of 1 cm). The group of larger liver lesions comprised 353 hyperdense and 1779 hypodense lesions, 1416 malignant and 716 benign lesions, and 1381 metastases, 449 cysts, and 229 hemangiomas. The group of 193 smaller liver lesions comprised 147 hypodense and 46 hyperdense lesions. These consisted of 89 malignant and 104 benign lesions containing 91 cysts, 5 hemangiomas, and 74 metastases (Table 1). Lesions from this repository were used both in the cross-validation performance evaluation of the content-based image retrieval (CBIR) system and in the evaluation of radiologist’s performance.

To establish the true diagnosis of the lesions (i.e., the ground truth for evaluation purposes), all evaluated lesions were described in radiological reports and classified before the experiment, by consensus reading performed by experienced radiological residents (with at least 6 years of work experience). Moreover, all available information from additional examinations, such as magnetic resonance imaging (MRI), (contrast-enhanced) ultrasound, positron emission tomography (PET)-CT, together with prior studies (long-term lesion monitoring), in addition to biopsy results and patient-specific clinical history (tumor markers, etc), were taken into consideration when available.

CADx System

We evaluated a machine learning-based system for fast retrieval of similar three-dimensional images of liver lesions, with the purpose of decision support in characterization of liver lesions. The framework is based on distance learning, with the intrinsic random forest similarity as the core distance function used to retrieve lesion examples from the repository of preannotated liver lesions described earlier. The system allows the combination of flexible user-defined search contexts and high-level semantic features to yield search results that are both meaningful to the user and relevant in the given context.

Further details concerning the CBIR framework and the related theoretical background are presented in the
supplemental electronic material and have been described previously (10), focusing on the machine learning-based CBIR framework and (13) introducing the problem of distance learning for retrieval-based decision support, and the intrinsic random forest similarity in particular.

The proposed workflow begins when a radiologist inputs an ROI containing the query lesion, with the option to include complementary information (which will be subsequently called high-level semantic features) into the CADx system query. High-level semantic features are pieces of information that are routinely mentioned in radiological reports (e.g., focality or rim continuity) as part of the query and then determines in which context he or she wishes to have discrimination support (e.g., lesion type). The investigated system is based on texture analysis, a database of annotated liver lesions, and a training process, which from a technical standpoint relies on a random forest similarity model and content-based image retrieval algorithms (in the image, blue represents benign lesions, while red represents malignant lesions) and operates much like a radiological image search engine. The system analyzes the image features of the input lesion (here, a malignant lesion, red ROI) and compares the input lesion’s features with the features of already classified database liver lesions. The CADx system then retrieves and displays annotated database lesions most similar to the input lesion in the given context. Both the images and the associated records and files of the retrieved patients are presented.

Table 1. Breakdown of the Underlying Database of 2325 Liver Lesions from the Evaluated Computer-Aided Diagnosis System

<table>
<thead>
<tr>
<th>Liver Lesions</th>
<th>Liver Lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1 cm³</td>
<td>&lt;1 cm³</td>
</tr>
<tr>
<td>Lesion density</td>
<td></td>
</tr>
<tr>
<td>Hyperdense</td>
<td>353</td>
</tr>
<tr>
<td>Hypodense</td>
<td>1779</td>
</tr>
<tr>
<td>Lesion histology</td>
<td></td>
</tr>
<tr>
<td>Benign</td>
<td>716</td>
</tr>
<tr>
<td>Malignant</td>
<td>1416</td>
</tr>
<tr>
<td>Lesion type</td>
<td></td>
</tr>
<tr>
<td>Benign primary liver tumors</td>
<td>35</td>
</tr>
<tr>
<td>Malignant primary liver tumors</td>
<td>38</td>
</tr>
<tr>
<td>Cysts</td>
<td>449</td>
</tr>
<tr>
<td>Hemangiomas</td>
<td>229</td>
</tr>
<tr>
<td>Metastases</td>
<td>1381</td>
</tr>
<tr>
<td>Total number of lesions</td>
<td>2132</td>
</tr>
</tbody>
</table>

Figure 1. Radiological work flow using the computer-aided diagnosis (CADx) system: The radiologist detects an indeterminate liver lesion in a computed tomography (CT) scan and seeks computer assistance in discrimination. He then draws a region of interest (ROI) around the lesion and inputs this lesion image from a CT scan into the CADx system. The radiologist has the option to input additional semantic information describing the lesions (e.g., focality or rim continuity) as part of the query and then determines in which context he or she wishes to have discrimination support (e.g., lesion type). The investigated system is based on texture analysis, a database of annotated liver lesions, and a training process, which from a technical standpoint relies on a random forest similarity model and content-based image retrieval algorithms (in the image, blue represents benign lesions, while red represents malignant lesions) and operates much like a radiological image search engine. The system analyzes the image features of the input lesion (here, a malignant lesion, red ROI) and compares the input lesion’s features with the features of already classified database liver lesions. The CADx system then retrieves and displays annotated database lesions most similar to the input lesion in the given context. Both the images and the associated records and files of the retrieved patients are presented.

Evaluation

In our study, we evaluated the predictive performance of the CADx system in retrieving lesions similar to the input lesion according to the tissue density, histology, and type; conducted receiver operating characteristic analysis as well as qualitative analysis of false-positive and false-negative findings for
each related classification task; and separately evaluated the performance for subcentimeter liver lesions.

In the experimental setting, each of the liver lesions from the underlying repository were separately taken as an input item into the CADx system, in the leave-one-patient-out (LOPO) cross-validation evaluation, once without further input of high-level semantic features and once with the additional input of high-level semantic features. Based on the input query, the system was asked to retrieve lesions from the repository displaying similar density (hypodense, hyperdense), histology (benign, malignant), or type (cyst, metastasis, hemangioma). A subset of 35 malignant lesions originating from primary liver tumors (e.g., hepatocellular carcinoma [HCC], cholangiocellular carcinoma [CCC]) as well as 38 benign primary liver tumors (e.g., adenoma, focal nodular hyperplasia [FNH]) and 15 malignant and 8 benign subcentimeter lesions were not used in the type evaluation, due to the small sample size of their corresponding category. The true (manual and confirmed) classification of the input lesion and the classification inferred from the lesions retrieved by our CADx system were then compared to assess the system’s accuracy.

The predictive performance of lesion retrieval for each characterization task was examined using LOPO evaluation and receiver operating characteristic area under the curve (ROC AUC) as the key criterion for evaluation (15). For lesion type classification, ROC was constructed and evaluated for the task of discriminating metastasis from hemangioma, which has proved to be the most difficult lesion type discrimination task in comparison to the tasks involving cysts.

In LOPO evaluation, repeatedly, the lesions corresponding to a single patient are left out from the repository and are used for performance evaluation. The system is retrained using the data from the reduced repository. LOPO is a common choice for performance evaluation in computer-aided detection (CADx) and CADx systems because it helps to avoid the optimistic bias of overlapping or dependent training and testing sets and because the loss of training information in the cross-validation is minimal, corresponding to data from a single patient only, which is important with learning-based studies with relatively small samples.

After the lesions are retrieved by the system, the ultimate classification decision is made with the weighted majority voting rule, using the proximity information and semantic annotations from the subset of the seven most similar lesions. This setting has been shown to work reliable in different contexts with different data sets (10).

The accuracy of the CADx system was evaluated by calculating the sensitivity and specificity for classifications resulting from the retrieved lesions with respect to two similarity criteria (also called contexts in this text), namely density and histology, as well as the per-class sensitivity for the retrieval-based lesion type classification. Moreover, positive predictive value (PPV) and negative predictive value (NPV) were examined for lesion retrieval according to histology.

The system run time per lesion retrieval was measured, and the mean run time with a standard deviation was calculated with usual clinical hardware Fujitsu Celsius [Fujitsu, Tokyo, Japan], (dual core Xeon 2.66 GHz [Intel, Santa Clara, CA, USA]; Windows XP, 32 bit [Microsoft, Redmond, WA, USA]).

To better understand the retrieval performance of the CADx system and to be able to compare it to the diagnostic conclusiveness of radiologists in the same experimental setup, both human experts and the CADx system were exposed to the same image data. Concretely, three radiologists with different levels of work experience (2, 3, and 16 years) independently categorized 83 liver lesion images in the portal-venous contrast agent phase according to their histology and type. The lesion image set was unlabeled and included 20 hemangiomas, 20 metastases, 20 cysts, 20 HCCs, and 3 FNHs. Both the CADx system and the three radiologists were given sets of contiguous 1-mm slices of CT volumes that contained the liver only. The software used by the human participants includes a 3D viewer (standard axial, sagittal, and coronal views) and allowed for the interactive assessment of both the liver and the lesion in question (Fig 2). If more than one lesion were present inside the liver, the lesion to be characterized was displayed with a surrounding box. We estimated the diagnostic accuracy of the resulting characterizations of each lesion’s histology and type, which were considered the two most clinically relevant contexts. Furthermore, we evaluated related interreader agreement in terms of both percentage of same decisions and Cohen’s $k$.

RESULTS

CADx

The retrieval accuracy measured with LOPO ROC AUC for the density characterization context (hypodense, hyperdense) was 95.2% when using only the lesion image as input and 97.1% when high-level semantic features were added into the CADx system’s query (Fig 3a). For small lesions, the system showed an accuracy of 98.7% without and 98.5% with additional input of high-level semantic features. When investigating the retrieval results with respect to similar histology (benign/malignant), we found an accuracy of 75.1% with only lesion image as input and 91.4% with additional high-level semantic features (Fig 3b). In the same context, small lesions were retrieved with an accuracy of 73.3% with only lesion image as input and 84.6% when including high-level semantic features into the query. Retrieval results for lesion type classification (hemangioma/metastasis) showed an accuracy of 85.8% with only a lesion image as input and 95.5% with additional input of high-level semantic features (Fig 3c). For small lesion retrieval, the accuracy with respect to lesion type was 81.9% with only lesion image as input and 92.7% when including high-level semantic features into the query (see Table 2 for detailed information). Figure 4
includes examples of correctly classified lesions and the sets of retrieved lesions for the lesions of each type.

The sensitivity and specificity of the CADx system when retrieving similar density lesions were found to be 89.8% and 91.9%, respectively. For retrieval according to histology, the sensitivity was 96.9% with a specificity of 59.9%. Retrieval according to type showed a per-class sensitivity of 85.0% for cysts, 85.2% for hemangiomas, and 96.0% for metastases. The PPV for retrieving lesions with similar histology (i.e., similarly benign or malignant lesions) was 91.5%, and the NPV was 92.2%.

As a further step to improve the system’s performance and underlying algorithms, we qualitatively evaluated the false-positive and false-negative classifications to recognize possible patterns of error (Fig 5). Perhaps the most salient of our findings was that 66% of the metastases that were misclassified (false-negatives) occurred in only 2% of the patients, all with an extensive disease and at least 10 liver lesions present. Moreover, 60% of the metastases that were misclassified as hemangioma were adjacent to liver veins, which might have misled the system’s assessment. Additionally, 75% of the hemangiomas that were misclassified as malignant exhibited atypical appearance in the given examinations, showing either late or early filling.

The exploited random forest-based similarity framework is very fast. The system’s run time for the core feature computation and similarity calculation on standard hardware was...
only 220 ± 44 ms on average per lesion; the retrieval-based
decision support in the present system is thus real-time.

Radiologists

The most experienced radiologist (reader 1, 16 years of
work experience) showed a diagnostic accuracy of 74% for
lesion histology and 51% for lesion type. The second most
experienced radiologist (reader 2, 3 years of work experience)
showed a diagnostic accuracy of 68% (histology) and 46%
(lesion type). The least experienced radiologist (reader
3, 2 years of work experience) showed a diagnostic accuracy
of 52% for lesion histology and 22% for lesion type.

When examining the percentage of interreader agreement,
we found reader 1 and reader 2 to agree on the same diagnosis
more often (71% for histology, 63% for type) than reader 1 and
reader 3 (58% for histology, 43% for type). The interreader
agreement is slightly higher when comparing the two less-
experienced radiologists (62% for histology, 53% for type).
Both trends are further reinforced by Cohen’s $\kappa$ values. For
histology we found a $\kappa$ value of 0.42 for the two more
experienced radiologist and 0.15 when comparing the most
experienced and the least experienced radiologists, whereas
we found a $\kappa$ value of 0.26 for the two least-experienced
radiologists.

When separately evaluating the percentage of interreader
agreement according to the diagnosis of lesion type, heman-
giomas were found to be most reliably diagnosed with 70% 
agreement, HCCs with 53%, cysts with 52%, metastases 
with 41%, and FNHs with 22%.

In all the tasks, the performance of each radiologist was 
significantly inferior with respect to that of the CBIR system,
confirmed with McNemar’s test for significance of the differ-
ence between two proportions ($P < .05$).

DISCUSSION AND CONCLUSION

Automatic characterization of liver lesions from CT scans 
has been under active study in applied radiology for over a
decade. Chen et al(16) were among the first to introduce a
system for liver lesion type classification. The neural
network–based system was able to discriminate hemangioma 
from HCC on unenhanced CT with an accuracy of 83%, vali-
dated on a set of 20 HCCs and 10 hemangiomas. Since
then, related studies (12,17–22) have focused largely on
lesion type discrimination for a selected subset of lesion

types but not on malignancy prediction.

Many of these studies similar to that of Chen et al (16) work 
with a nonmetastatic liver and try to discriminate HCC from
benign lesions, such as hemangiomas and cysts (17,19–21).
All the related studies use certain descriptors of lesion 
texture and a certain machine-learning algorithm; the
reported accuracies range from 83% in Chen et al (16) to
100% in Gletsos et al (18), although the samples studied are
likely nonrepresentative, with the largest study (22) including 87 HCCs, 62 cysts, and 60 hemangiomas.

The metastatic liver and the discrimination of metastases from benign lesions, similar to our work, were studied (12,17,19). While Bilello et al (19) use low-level lesion texture descriptors (17) and study the use of clinically motivated semantic features representing enhancement patterns, Napel et al (12) propose and study the combination of both texture descriptors and high-level semantic features. Napel et al (12) reach a competitive performance with the mean precision surpassing 90% at all levels of recall; however, all the reported sample sizes remain small, ranging from 10 metastases in Napel et al (12) to 53 in Nino-Murcia et al (17). An interesting observation made by Bilello et al (19) is the difficulty to automatically discriminate metastases from hemangiomas, as opposed to discriminating other types of lesions. Moreover, their reported AUC ROC is 0.887, compared to an AUC ROC of 0.955 for our system.

Similar to our work and in addition to lesion type prediction considered in the aforementioned research, Seltzer et al (23) study malignancy characterization on a metastatic liver. Both the performance of radiologists and the automatic characterization are analyzed, and it is concluded that the latter might be superior (93.6% AUC ROC) to the former (92.9% AUC ROC). It is important to note that this performance is achieved with a combination of information from two modalities: contrast-enhanced CT and contrast-enhanced MRI. In our studies, we are able to achieve similar or better performance with CT volumes only. Seltzer et al (23) exploit high-level semantic annotations as features only, and their sample size is also limited to a total of 146 lesions. The most remarkable contribution of this study is the proof that the performance of an automatic system (using the high-level descriptors only) may be better than that of an experienced radiologist, as far as malignancy characterization is concerned.

Most related studies in automatic lesion characterization consider “black-box” classification models, such as a neural networks or a support vector machine (SVM). Contrary to this, we (12) and Yang et al (22) propose a retrieval-based classification model in which the ultimate system decision regarding the lesion class is based on the classes of the retrieved lesions, adding transparency to the decision-making process and opening additional possibilities in decision support, based on information from the retrieved patient records.

Similar to our work, Yang et al (22) consider a distance function learning framework, although they apply it to a non-metastatic liver for discrimination of HCC from benign lesions. Only texture features are used as descriptors in their work, and their rather strict lesion selection criterion (only one, the most dominant lesion per patient is chosen) may

<table>
<thead>
<tr>
<th>True positives</th>
<th>False positives</th>
<th>False negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>

*Figure 5. Example of misclassifications. On the first row, input image of a metastatic lesion, true positive retrieval of metastasis (a), false-positive retrieval of hemangioma that was misclassified as metastasis (b), false-negative metastasis that was misclassified as hemangioma (c), note the adjacent liver vein right next to the metastasis, a possible reason for misclassification that was detected in 60% of the metastasis cases that were misclassified as hemangioma. On the second row, input image of a hemangioma, true-positive retrieval of hemangioma (d), false-positive retrieval of metastasis that was misclassified as hemangioma (e), again depicting an adjacent liver vein. On the third row, input image of a cyst, true-positive retrieval of cyst (f), false-positive retrieval of a metastasis that was misclassified as cyst (g), false-negative hemangioma that was misclassified as malignant (h). On the third row, input image of a cyst, true-positive retrieval of cyst (i), false-negative cyst that was misclassified as metastasis (j).*
clearly bias the results, which reach a mean average precision above 90%. Additionally, the combination of information from multiple contrast-agent phases is studied in their work, which has proved to increase the mean average precision to 94.5%. The combination of information from multiple phases is out of the scope of our study but remains a very interesting direction for future work.

Two important differences between our work and that of Napel et al (12) are the distance function used for retrieval and the number of retrieved cases necessary for an accurate prediction. While Napel et al (12) use the simple weighted sum of differences in corresponding feature values as the distance function (which is also known as the taxicab, or \( L_1 \) metric), we exploit here the more sophisticated intrinsic random forest similarity, which has proved to result in a better predictive performance (10). In addition to predictive performance, another benefit of our distance function is the smaller number of nearest neighbors required for an acceptable performance in decision making. With our system, an acceptable performance is achieved with a few neighbors, and no significant improvements are observed when considering more than seven neighbors. Napel et al (12) demonstrate that the taxicab metric may require 20 or more nearest neighbors, and the accuracy differs significantly (more than 10%) when compared to the accuracy achieved with only a few retrieved cases. Achieving acceptable predictive performance with only one or a few retrieved cases is a desirable property, since it means that the retrieved result list is short and pertinent and the radiologist will need less time to review it, understand the suggestions regarding lesion malignancy and type, and review the patient records to make a better decision regarding further therapy or patient management.

We demonstrated that the performance of our system is promising even without the additional input of information (high-level semantic features). However, it was also demonstrated that this additional high-level information is likely to increase system performance, a finding consistent with Napel et al (12).

We believe that being able to solely input an ROI corresponding to an indeterminate liver lesion into the CADx system is a far more practical clinical workflow than having to provide additional, compulsory information in the form of semantic features. Especially, as van Ginneken et al (24), in accordance with others (10,25), stated that seamless workflow integration is inevitable for CADe/CADx to be successful not only in a laboratory environment but also in the clinical radiological routine. According to the framework suggested by Depeursinge et al (25), we present a system that can initiate the image search process from the ROI drawn around an indeterminate liver lesion and provide reliable decision support; which is a straightforward handling that is easy to integrate into the clinical workflow. Doi (26) demands that the millions of images stored in a clinical picture archiving and communicating system be capitalized on by future decision support systems. Our CADx system requires a short computation time for a similar lesion list to be retrieved (and therefore a suggested categorization) and is easy to integrate into the hospital radiology information system/picture archiving and communicating system as it does not demand special hardware. Akgül et al (27) state that some subtle image features that might not be discernible by humans still might be distinguished through automatically computed image descriptors. Our study is another confirmation of this. As demonstrated in this study, when characterizing CT liver lesions with image information limited to the liver, and without additional patient-specific information (e.g., previous scans, additional examinations like MRI, laboratory values like tumor markers, etc), the performance of radiologists was significantly inferior to that of the CADx system. Consequently, we suggest that future decision support systems be provided with all available patient-specific information (e.g., laboratory values, physical examination results, patient history, current medication, etc), which is normally taken into consideration by radiologists. Furthermore, we propose that future decision support systems be able to analyze all images of an examination and take all image information into consideration, again similar to what is done by a radiologist. This could certainly have a positive impact on the predictive performance of CADx.

Moreover, our liver lesion characterization module may serve as a dedicated component in a multipurpose CAD system. As outlined by Krupinski (28), there is a clear need for a multipurpose CAD available for all types of images and lesions, as opposed to having a single CAD application for each particular lesion type. Such a system should point out abnormalities, provide a likelihood of malignancy and a probability of other characteristics such as lesion type, point out traces of the lesion on previous examinations, track lesion progress over time, promptly detect therapy response, provide examples of similar lesions from other cases, and even propose a therapy.

Similar to CADe systems, the CADx system can be particularly helpful for inexperienced readers (24,29) and will not have intrarad and inter-reader variance, typical for a human reader. Meanwhile, van Ginneken et al (24) suggested that computer systems aiding radiologists in detecting lesions may not be the most efficient use of computer assistance. They suggest that computer assistance for interpretation and diagnosis instead can be of more importance, for example, by supplying a computer-generated ranking of lesions suspicious for malignancy. This, among other features, is what our presented CADx is potentially able to provide, using the same method and slightly adjusted workflow. The study of this feature is an interesting direction for future work. In addition to ranking the suspicious image regions, it will display lesions of similar histology or type for each input lesion in question and supply additional information concerning the retrieved lesions, such as clinical information accessible from the related patient record and information from the previous radiological reports.

Various authors have confirmed that small, subcentimeter CT liver lesions are difficult to distinguish (30,31). We
demonstrated that the presented software could distinguish smaller liver lesions with comparable accuracy as studies distinguishing small liver lesions using MRI (32,33).

We demonstrated in this study that the interreader agreement and accuracy of interpretation of radiological findings based on lesion appearance alone increase with experience. Furthermore, we found a slightly higher interreader agreement for the two less-experienced radiologists in comparison to the agreement of the most- and less-experienced radiologist. Moreover, we demonstrated that when using CT image features alone, hemangiomas were the lesions most reliably diagnosed, while cysts and metastases remain challenging.

One limitation of our study was that we evaluated only three different types of liver lesions. This was due to the fact that for the machine learning and training process of the CBIR-based CADx system, a large number of lesions of the same type are required, and at the time of the study we had limited numbers of other lesions (e.g., HCCs, CCCs, FNHs, adenomas, etc.). However, the integration of HCCs and FNHs into the CBIR-based CADx system is currently in progress.

To summarize, we presented a CADx system relying on CBIR that distinguishes liver lesions robustly, which has been proved to provide promising accuracy even for hard-to-interpret, subcentimeter liver lesions. While emphasizing radiological usability, the presented system can be seamlessly integrated into the daily routine. We have also proved that this software application outperformed radiologists in diagnosing liver lesions when both system and radiologists relied on image information alone. Furthermore, we demonstrated that the human assessment of liver lesions (based on images) requires experience and that the classification remains difficult.

To conclude, applying state of the art CADx methods for liver lesion characterization in daily routine is practical, reliable, and accurate and will improve patient care and outcome.

SUPPLEMENTARY DATA
Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.acra.2013.09.001.

REFERENCES