Quantitative functional MRI Biomarkers Improved Early Detection of Colorectal Liver Metastases

Yifat Edrei, PhD,1 Moti Freiman, PhD,2 Miri Sklair-Levy, MD,3 Galia Tsarfaty, MD,4 Eitan Gross, MD,5 Leo Joskowicz, PhD,6 and Rinat Abramovitch, PhD1,7*

Purpose: To implement and evaluate the performance of a computerized statistical tool designed for robust and quantitative analysis of hemodynamic response imaging (HRI)-derived maps for the early identification of colorectal liver metastases (CRLM).

Materials and Methods: CRLM-bearing mice were scanned during the early stage of tumor growth and subsequently during the advanced-stage. Three experienced radiologists marked various suspected-foci on the early stage anatomical images and classified each as either highly certain or as suspected tumors. The statistical model construction was based on HRI maps (functional-MRI combined with hypercapnia and hyperoxia) using a supervised learning paradigm which was further trained either with the advanced-stage sets (late training; LT) or with the early stage sets (early training; ET). For each group of foci, the classifier results were compared with the ground-truth.

Results: The ET-based classification significantly improved the manual classification of the highly certain foci (P < 0.05) and was superior compared with the LT-based classification (P < 0.05). Additionally, the ET-based classification, offered high sensitivity (57–63%), accompanied with high positive predictive value (>94%) and high specificity (>98%) for suspected-foci.

Conclusion: The ET-based classifier can strengthen the radiologist’s classification of highly certain foci. Additionally, it can aid in classifying suspected-foci, thus enabling earlier intervention which can often be lifesaving.

Key Words: hemodynamic response imaging; machine learning; SVM; cancer

J. Magn. Reson. Imaging 2013;00:000–000.
© 2013 Wiley Periodicals, Inc.

COLORECTAL CANCER IS the third most common cancer in the United States (1). Hepatic metastases develop in approximately 50% of colorectal cancer cases (2). The development of colorectal liver metastases (CRLM) is the main cause of death in patients presenting advanced stages of the disease (3). CRLM are typically treated by means of partial hepatic resection, and is currently the only curative treatment (2,4). Unfortunately, only a minority of patients with liver metastases is eligible for resection, due to the often late discovery coupled with advanced stages of the metastases (5–7). Thus, early diagnosis is critical for successful resection and for refined treatment selection criteria (8,9). As previously shown, whereas the normal liver is predominantly supplied by the portal vein, in patients with overt hepatic tumor, a higher proportion of liver blood flow is derived from the hepatic artery (10). Additionally, previous studies have shown that small CRLM (<520 µm) are hypovascular compared with the adjacent liver, while only advanced CRLM show an exclusively arterial blood supply (11,12).

Blood oxygenation level-dependent (BOLD) MRI was originally proposed by Ogawa et al (13) to study hemodynamic changes related to neuronal activation. BOLD MRI uses deoxyhemoglobin as an endogenous contrast agent, which enables detection of changes in blood flow, volume, and oxygenation. Increased BOLD signal can occur due to either endogenous effects such as neuronal activity or to exogenous stimuli such as respiratory challenges of hyperoxia or hypercapnia. Hemodynamic response imaging (HRI) is a functional MRI (fMRI) method that involves hypercapnic challenge with brief inhalation of 5% CO2 followed by hyperoxic challenge with brief inhalation of carbogen, which was introduced for monitoring...
changes in liver perfusion and hemodynamics without the need for contrast agent administration (14–16). Recently, the applicability of this method was demonstrated following the hemodynamic changes that occur during CRLM establishment and progression, with particular sensitivity toward the subtle hemodynamic changes induced during the early growth of CRLM (17). These early changes include both tumor hypovascularity and increased arterial blood supply and are fully detectable by HRI. HRI enables noninvasive, nontoxic, three-dimensional analyses of tissue wide hemodynamic changes that occur in the entire liver, without being limited to evaluating only the sampled region, as in traditional biopsy-based techniques.

While the HRI method carries promise for characterization of the vascular profile of CRLM, the currently described analysis is limited to the qualitative assessment of the HRI-derived parameters (i.e., mean ΔSO2 and ΔSCO2), and requires a trained reader. A reliable statistical model designed to quantitatively analyze the early hemodynamic changes in CRLM as detected by HRI is expected to significantly increase the clinical value of HRI collected data. The quantitative analysis must address high intersubject variability in functional activation responses, intertissue variations in MRI signals, and allow for automatic region identification for analysis. A previous study has presented a prototype of an automatic image analysis method that enables the quantitative assessment of suspected CRLMs in their early growth stage, based on reactivity maps which were generated by HRI (18). However, the ultimate added value of such a system can only be assessed by radiologists in a setting mimicking the clinical situation.

Computer-aided diagnosis (CAD) systems are commonly used to extract and display predefined information from hundreds of radiological images. These systems function automatically and generate measures for later comparison by the radiologist to an existent database. However, this approach involves a high frequency of false-positive cases and requires U.S. Food and Drug Administration (FDA) approval. Semi-automatic systems, which require involvement of the radiologist have become quite popular and reduces the number of false-positive cases. As this approach requires deliberate radiologist action, the need for FDA approval is waived.

The aim of the present study was to implement and evaluate the performance of a statistical model designed for the robust and quantitative analysis of HRI-derived maps for the early detection of CLRM in a mouse model. The proposed system couples automatic normalization and automatic clustering of functional hepatic regions, based on the HRI collected data, with a machine learning approach for cluster classification, based on previously collected training data. The automatic normalization step eliminates the high intersubject variability in functional activation responses, while the clustering step objectively defines the regions of interest for analysis based on differences in their hemodynamic properties. The statistical model provides quantitative analysis of the HRI-derived parameters. In this manner, the proposed model equips the radiologist with the ability to identify regions suspected to contain early metastatic lesions and to quantitatively examine the pattern of the hemodynamic changes in these regions against predefined databases of both normal and metastatic hemodynamic patterns.

The added value of the model toward early detection of CLRM was evaluated in a mouse model imitating the clinical situation, wherein the radiologist read only the anatomical images, and, for suspected regions, made use of the functional information provided by the HRI statistical model as a surrogate biomarker. By using the computerized model for classification of suspected foci, we demonstrated that, on average, 60% of the suspected foci were correctly classified as tumors with a high positive predictive value (PPV) rate.

**MATERIALS AND METHODS**

**Mouse Model**

All the experiments were performed in accordance with the guidelines and approval of the Animal Care and Use Committee, which holds National Institutes of Health (NIH) approval (OPRR-A01-5011). CT-26 murine colorectal adenocarcinoma cells (10⁴ cells in 300 µL per mouse) were intrasplenically injected into 7- to 8-week-old, anesthetized male CB6F1 mice (n = 15; Harlan; Ein-Kerem, Israel), as previously described (17,19). In this model, one to five hepatic nodules per mouse were detected by MRI, 13–17 days after cell inoculation.

**MR Imaging**

Anatomical and fMRI scans were performed on a horizontal 4.7 Tesla (T) Biospec spectrometer (Bruker Medical, Ettlingen, Germany) equipped with a 3.5-cm birdcage coil. For the MRI acquisition, mice were anesthetized with pentobarbital (CTS group, Hod-Hasharon, Israel; 30 mg/kg, intraperitoneally injected) and placed in a supine position. Tumors were assessed using axial fast spin echo images (repetition time = 2000 ms; echo time = 37 ms; field of view = 3 cm; in plane resolution = 117 µm; slice thickness = 1 mm). Changes in hepatic hemodynamics were evaluated using the HRI protocol as previously described (17,19). In brief, the images were acquired using T2*-weighted gradient echo images (repetition time = 147 ms; echo time = 10 ms; field of view = 3 cm; in plane resolution = 117 µm; slice thickness = 1 mm; 2 averages; 37 s/image), under normoxic, hypercapnic (5% CO2) and hyperoxic (carbogen; 95% O2 + 5% CO2) conditions. The hypercapnic and hyperoxic reactivity maps are given as the percentage change of the MRI signal intensity (ΔSCO2 and ΔSO2, respectively) (15).

**Data Sets**

The input data set for each mouse consisted of the anatomical images that covered the entire liver and
For each mouse, the liver borders were marked on the anatomical images using Analyze 7.0 (BIR, Mayo Clinic, Rochester, MN). Each mouse was scanned twice: the first scan was acquired during the early stage (ES) of tumor growth and the second acquisition was acquired later during the advanced stage (AS), when tumor presence was clearly visible in the anatomical images. To achieve good correspondence between the ES and AS anatomical images, imaging parameters (including slice thickness, slice orientation, and central positioning) were kept constant, and the imaging procedure was performed by one researcher (Y.E.).

Collecting Foci for Analysis

Three experienced radiologists (Obs), with more than 10 years practice each, participated in the study. Each radiologist reviewed only the anatomical images of each mouse, for all the ES images, and marked several suspected foci on each image. Moreover, the radiologist classified each marked focus as either tumor with high certainty (T) or as a suspected focus (S). In cases that the focus was unmarked by the observer, it was classified as healthy liver (L). The table summarizes the observers' classification of the marked foci in this image and their ground truth.

![Figure 1. Observers' foci marking and classification. a: Representative axial anatomical image of a mouse with four selected foci (1–4). Each focus was classified by each observer as either a tumor with high certainty (T) or as a suspected focus (S). In cases that the focus was unmarked by the observer, it was classified as healthy liver (L). b: The table summarizes the observers' classification of the marked foci in this image and their ground truth.](image)

<table>
<thead>
<tr>
<th>Ground truth</th>
<th>Observer I</th>
<th>Observer II</th>
<th>Observer III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor</td>
<td>19</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Certain</td>
<td>11</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Suspected</td>
<td>28</td>
<td>65</td>
<td>6</td>
</tr>
<tr>
<td>Unmarked</td>
<td>11</td>
<td>65</td>
<td>24</td>
</tr>
<tr>
<td>Unmarked</td>
<td>11</td>
<td>65</td>
<td>60</td>
</tr>
<tr>
<td>Liver</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Unmarked</td>
<td>25</td>
<td>23</td>
<td>85</td>
</tr>
</tbody>
</table>

The total selected foci were separated to either tumor with high certainty or suspected foci according to each of the observer’s decision. Foci that were not marked by the observer were considered as liver under his selection. The ground truth evaluation for each focus was derived from the advanced phase MRI scan.

Statistical Model Construction

The proposed computerized assistance method uses a supervised learning paradigm. The model construction involved four steps: (i) HRI map normalization, (ii) functional clustering, (iii) feature extraction, and (iv) statistical model training. This model is an improved version of a model that was explicitly described in a previous work (18). Each step is briefly described below.

(i) Map Normalization

Normalization of the hemodynamic reactivity maps is necessary to eliminate the high intersubject variability. First, the liver region of interest (ROI) was determined manually and all of the remaining pixels were eliminated from the maps. Next, each HRI map was centered around a mean intensity of zero, with a standard deviation of 1 (18). A truncated mean was used to reduce the sensitivity to noise. The mean and standard deviation were computed for each liver after excluding large vessels.

(ii) Functional Based Clustering

Pixel-wise analysis tends to be highly sensitive to noise resulting from the acquisition process. Thus, we clustered the pixels of the HRI maps into functional partitions, defined by areas containing pixels with similar HRI values for both $\Delta \text{SCO}_2$ and $\Delta \text{SO}_2$ (Fig. 2). The clustering process was done automatically using the mean shift algorithm (20), where both AS maps have the same weight, and the output was a partitioning map which defines the segmentation of the liver to functional-based partitions.

Table 1: All Selected Foci (Filtered) Assignment Versus Ground-Truth per Observer

<table>
<thead>
<tr>
<th>Ground truth</th>
<th>Observer I</th>
<th>Observer II</th>
<th>Observer III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor</td>
<td>19</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Certain</td>
<td>11</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Suspected</td>
<td>28</td>
<td>65</td>
<td>6</td>
</tr>
<tr>
<td>Unmarked</td>
<td>11</td>
<td>65</td>
<td>24</td>
</tr>
<tr>
<td>Liver</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Unmarked</td>
<td>25</td>
<td>23</td>
<td>85</td>
</tr>
</tbody>
</table>

$^a$The total selected foci were separated to either tumor with high certainty or suspected foci according to each of the observer’s decision. Foci that were not marked by the observer were considered as liver under his selection.
(iii) Feature Extraction

A characteristic vector of HRI reactivity features was computed for each cluster in the segmented map. Each vector was comprised of five parameters derived from each DS map: (A) mean, (B) standard deviation, (C) kurtosis, (D) skewness, and (E) interquartile range, resulting in a vector of 10 parameters for each segment (Fig. 3). This representation provides a comprehensive characterization of the hemodynamic reactivity changes unlike the mean value, previously used in Edrei et al (17).

(iv) Statistical Model Training

A support vector machine (SVM) classification model was constructed using the selected sets of tagged feature vectors. The computation was based on the hypercapnic and the hyperoxic values equally. A generalized radial basis function (RBF) kernel with the earth movers distance (EMD) as the affinity measure (21) was used. EMD provides a measure of the distance between two probability distributions, which can be represented by normalized histograms, over a region of interest. The EMD represents the minimal cost required to turn the shape of one probability distribution to the other. Unlike other histogram-based measures, the EMD accounts for cross-bin information, and has been shown to output perceptually natural distances for many applications.

The model was trained using two distinct approaches. In the first: a late training (LT) approach, HRI datasets acquired during AS tumor growth were used for the SVM training. In the second; an early training (ET) approach, HRI datasets acquired during the ES tumor growth were used for the SVM training. In the ET approach; the “leave one out” methodology (22) was used to prevent the model from over adjusting to the tested point. Following this methodology, a single focus from the original sample is used as the test data, and the remaining foci used as the training data for classifier construction. We repeated this process many times such that each focus in the sample is used once as the test data. In both approaches, the ground truth classification was based on the AS status.

Quantitative Foci Analysis

For each selected point, which was marked by the radiologist, the associated feature vector was computed from the corresponding cluster (Fig. 4). The feature-vector was then quantitatively evaluated by the SVM model. The model output is a number representing the likelihood of the selected focus being a metastasis.

Figure 2. Partitioning map created automatically based on the calculated features of HRI reactivity maps. Axial anatomical image (a), the corresponding HRI maps (b, DS\text{CO}_2; c, DS\text{O}_2), and the partitioning map (d) of a representative mouse acquired during the early tumor growth phase. e: Representative healthy liver cluster and its corresponding HRI reactivity. f: Representative CRLM verified cluster and its corresponding HRI reactivity. Note the prominent difference between healthy and CRLM HRI appearance.

Figure 3. Box and Whiskers plots of mean values calculated for each feature. a: Mean. b: Standard deviation. c: Kurtosis. d: Skewness. e: Interquartile range. The values were calculated from clusters in the AS images. The differences between liver (white) and CRLM (gray) were significant only for some of the features (*P < 0.01, **P < 0.0001), however the combination of all the features provides a comprehensive characterization of the hemodynamic reactivity changes.
Model Validation

After we constructed the classifier, we evaluated the improvement of each radiologist’s performance for liver metastasis diagnosis in the early MRI data sets with the assistance of the HRI-based statistical model. We used both LT and ET approaches for model construction and compared their performance.

For each group of foci (highly certain or suspected), the classifier results were compared with the ground-truth status of the focus (determined using the AS images) and the radiologists’ diagnostic performance (either manual or with the model-based computerized assistance, ET or LT) was assessed by descriptive measures: sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). Additionally, we calculated the odds ratio (OR) and its 95% confidence limits (CI). If the resulting 95% CI did not include the value of 1, it was considered statistically significant. The performances of the manual classification and the computerized classifications were compared by the McNemar test. A P value of < 0.05 was considered statistically significant.

RESULTS

Manual Versus Computerized Classification of Foci Assigned as Tumor With High Certainty

Initially, we evaluated the manual versus computer assisted classifications of the foci that were assigned as tumors with high certainty including the ignored foci, for each observer. Table 2 summarizes the decisions of each observer and the corresponding computerized determinations of these foci, compared with their ground truth class as diagnosed from the AS images. The ET-based classifications significantly improved the manual classifications of two of the observers (Obs II and Obs III; P < 0.05; McNemar test), as reflected mainly by improving the sensitivity and the PPV rate (Table 3). The number of foci that were classified as tumors by the ET-based classification, which were later confirmed as tumors, was higher compared with the number manually classified (Table 3; Obs I: 70% versus 63%; Obs II: 65% versus 41%; Obs III: 65% versus 48%), and the number of false-negative cases was reduced by 3–10% (Table 2; Obs I: 9 versus 11; Obs II: 12 versus 20; Obs III: 17 versus 25). In addition, the ET-based classification was significantly better than the LT-based classification (for all the observers; P < 0.05; McNemar test). According to all of the descriptive measures, the ET-based classification was more accurate than the LT-based classification (P < 0.05 for all the observers).

Model-Based Classification of Suspected Foci

Our next goal was to assess the classifier performance on points that were assigned as suspected foci (see examples in Fig. 1a). Only the ET-based classifier was used for these foci classification, because its performance proved more reliable compared with the LT-based classifier on the highly certain foci. By using
the ET-based classifier, 57–63% of the foci that were later confirmed as tumors were classified correctly already at their early growth stage; namely, the ET-based classification offered high sensitivity, accompanied with high PPV rate (> 94%; few false-positive cases) (Tables 4 and 5). Additionally, more than 98% of foci that were later confirmed as healthy liver, were indeed classified correctly. The percentage of false-negative cases was similar to that obtained in the certain foci classification (12% versus 13%, respectively).

**DISCUSSION**

Advanced surgical techniques and loco regional therapies for metastatic colorectal cancer require accurate lesion detection, characterization, and localization (23). In addition, preoperative staging is important for treatment selection to avoid irrelevant procedures (24). Evaluation of tumor resectability requires vascular structure assessment. Yet, along with improved imaging techniques, the need for assistive image analysis tools to support the physician’s decisions has increased. These tools must be highly reliable to service the clinic. Recently, HRI has been demonstrated as a noninvasive imaging technique sensitive to hemodynamic changes in a mouse model of liver metastases (17,19). The utility of BOLD MRI for the assessment of healthy and diseased human livers was recently introduced (25). Yet, interpretation of HRI maps is challenging and highly depends on proper map normalization, appropriate selection of regions of interest and on qualitative interpretation of the image output.

In this study, we propose a machine learning-based method which assists the radiologist to classify suspected foci, based on the functional behavior of the metastatic vasculature. In the mice model, the assisted classification of suspected foci was highly precise (mean accuracy, 87%), where the mean percent of cases correctly diagnosed as tumor was higher than the mean percent of foci correctly diagnosed as healthy liver (mean PPV = 96%; mean NPV = 85%). Although only 60% of the suspected foci were correctly classified as tumors by the computerized tool, nevertheless, no action beyond additional follow-up would be taken without this assistance. The proposed model-based approach yields reliable HRI analysis through automatic map normalization, clustering, and statistical evaluation, and thereby enabling objective and independent analysis.

The sensitivity of ET-based classification was higher than LT-based classification for both highly certain foci and suspected foci, suggesting that the tumoral vascular profile is changing during liver metastasis progression. The use of an ET-based model can probably strengthen the radiologist’s classification regarding highly certain foci and moreover, it can improve his choice concerning the most appropriate therapeutic procedures. In addition, the described model assists in classifying suspected foci at their early growth stage, which probably would enable earlier intervention. It should be mentioned, that in a clinical situation, in cases of suspected foci, no medical procedure would be implemented, excluding recommendation for further follow-up.

While biopsy is considered the gold standard for assessing suspected metastases, this approach often fails to reliably diagnose all the metastases present in the liver and can yield misleading interpretations (26). Functional image guidance based on quantitative HRI analysis using the proposed method, may improve diagnosis accuracy by providing more comprehensive coverage of the entire liver. Today, the use of multiparametric methods to analyze MRI outputs is well accepted (27–29). The presented computational

<table>
<thead>
<tr>
<th>Ground truth</th>
<th>Observer I</th>
<th>Observer II</th>
<th>Observer III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor</td>
<td>16</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>Liver</td>
<td>12</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>ET model decision</td>
<td>Tumor 16</td>
<td>Liver 12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liver 64</td>
<td>Liver 59</td>
<td></td>
</tr>
</tbody>
</table>

Cl, confidence level.
assessed livers of mice that were intrasplenically injected with CT26 colorectal cancer cells; therefore, we cannot conclude whether this computerized method could also discriminate between benign and primary hepatic lesions. However, in previous studies, we showed the different patterns of HRI reactivity in diverse liver pathologies including hepatocellular carcinoma and liver fibrosis (14,17,34); thus, we believe that this method could be more general. Nevertheless, the actual utility has to be evaluated in humans under the clinical conditions. Finally, implementation of HRI in the clinic was met with some difficulties; the main obstacle was that unanesthetized human subjects have increased respiratory rates in response to hypercapnic challenge, in contrast to the current animal model. We hypothesized that use of a rebreathing system (31) can prevent the natural compensation and can raise the PaCO2 in the blood.

In conclusion, we have presented an HRI-based statistical model that improves the overall performance of the radiologist in the early detection of CRLM in a mouse model. The proposed approach enables the radiologist to noninvasively obtain a better assessment of the hemodynamic status of regions of interest within the liver, and thus support his decision-making process. Use of the model in conjunction with radiological evaluation of the anatomical data yields a significant improvement in early detection of CLRM without introducing false-positive recordings. Proper classification of suspected foci is of paramount importance at the early stages of CRLM development and may significantly enhance treatment outcomes.

ACKNOWLEDGMENTS
We thank Mr. Yehonatan Sela for his assistance in data processing. Prof. Norman Grover for his assistance in statistics, and Dr. Yehudit Posen for manuscript editing. R.A. was supported by a grant number 1243/10 from the Israel Science Foundation and M.F was funded by the Hebrew University Hoffman Leadership and Responsibility Fellowship Program.

REFERENCES
4. Simmonds PC, Primrose JN, Colquitt JL, Garden OJ, Poston GJ, Rees M. Surgical resection of hepatic metastases from colorectal...


