Evaluating performance of a user-trained MR lung tumor autocontouring algorithm in the context of intra- and interobserver variations

Eugene Yip and Jihyun Yun
Department of Medical Physics, Cross Cancer Institute, 11560 University Avenue, Edmonton, AB T6G 1Z2, Canada
Medical Physics Division, Department of Oncology, University of Alberta, 11560 University Avenue, Edmonton, AB T6G 1Z2, Canada

Zsolt Gabos, Sarah Baker, and Don Yee
Department of Radiation Oncology, Cross Cancer Institute 11560 University Avenue, Edmonton, AB T6G 1Z2, Canada
Radiation Oncology Division, Department of Oncology, University of Alberta, 11560 University Avenue, Edmonton, AB T6G 1Z2, Canada

Keith Wachowicz, Satyapal Rathee, and B. Gino Fallone
Department of Medical Physics, Cross Cancer Institute, 11560 University Avenue, Edmonton, AB T6G 1Z2, Canada
Medical Physics Division, Department of Oncology, University of Alberta, 11560 University Avenue, Edmonton, AB T6G 1Z2, Canada

(Received 17 August 2017; revised 17 October 2017; accepted for publication 10 November 2017; published 15 December 2017)

Purpose: Real-time tracking of lung tumors using magnetic resonance imaging (MRI) has been proposed as a potential strategy to mitigate the ill-effects of breathing motion in radiation therapy. Several autocontouring methods have been evaluated against a “gold standard” of a single human expert user. However, contours drawn by experts have inherent intra- and interobserver variations. In this study, we aim to evaluate our user-trained autocontouring algorithm with manually drawn contours from multiple expert users, and to contextualize the accuracy of these autocontours within intra- and interobserver variations.

Methods: Six nonsmall cell lung cancer patients were recruited, with institutional ethics approval. Patients were imaged with a clinical 3 T Philips MR scanner using a dynamic 2D balanced SSFP sequence under free breathing. Three radiation oncology experts, each in two separate sessions, contoured 130 dynamic images for each patient. For autocontouring, the first 30 images were used for algorithm training, and the remaining 100 images were autocontoured and evaluated. Autocontours were compared against manual contours in terms of Dice’s coefficient (DC) and Hausdorff distances ($d_H$). Intra- and interobserver variations of the manual contours were also evaluated.

Results: When compared with the manual contours of the expert user who trained it, the algorithm generates autocontours whose evaluation metrics (same session: $DC = 0.90(0.03)$, $d_H = 3.8(1.6)$ mm; different session $DC = 0.88(0.04)$, $d_H = 4.3(1.5)$ mm) are similar to or better than intraobserver variations ($DC = 0.88(0.04)$, and $d_H = 4.3(1.7)$ mm) between two sessions. The algorithm’s autocontours are also compared to the manual contours from different expert users with evaluation metrics ($DC = 0.87(0.04)$, $d_H = 4.8(1.7)$ mm) similar to interobserver variations ($DC = 0.87(0.04)$, $d_H = 4.7(1.6)$ mm).

Conclusions: Our autocontouring algorithm delineates tumor contours (<20 ms per contour), in dynamic MRI of lung, that are comparable to multiple human experts (several seconds per contour), but at a much faster speed. At the same time, the agreement between autocontours and manual contours is comparable to the intra- and interobserver variations. This algorithm may be a key component of the real time tumor tracking workflow for our hybrid Linac-MR device in the future.

Key words: autocontouring, linac-MRI hybrid, MRI guidance, observer variability, respiratory motion management, tumor tracking

1. INTRODUCTION

Intrafractional tumor motion due to respiration poses a significant challenge to lung cancer radiotherapy. Recently, real-time, magnetic resonance image (MRI) guided radiation therapy has been made possible by the development of novel hybrid devices that combine the imaging capability of a MR imager and a radiation therapy device. These systems are capable of real-time tumor imaging for tracking, in which the radiation beam follows a moving tumor in real-time. Numerous groups have proposed schemes to acquire dynamic 2D MR images to track tumor motion during radiation delivery. Ideally, the tumor (i.e., gross tumor volume [GTV]) would be contoured in these dynamic images by an expert radiation oncologist such that the radiation field is conformed using dynamic multi-leaf collimator (MLC) to the known shape...
and location of tumor. The AAPM Task Group 76\textsuperscript{1} report recommends a maximum delay time of 0.5 s for real time tumor tracking. This delay includes times for image acquisition, reconstruction, contouring, and MLC repositioning. It is unfeasible for a human expert to continually contour images at a rate of lower than 0.5 s per image for the duration of treatment. To address this, our group\textsuperscript{6,39} and others\textsuperscript{8} have previously proposed real time autocontouring algorithms. A key feature to imaging and contouring in real time is that it does not rely on regular, predictable motion of lung tumor. Some of these algorithms have been validated using in-vivo images\textsuperscript{6,8} against the “gold standard” contours drawn by a single human expert. However, with a single expert, the impact of the expert’s individual preferences and the inherent uncertainties (i.e., intraobserver and inter-observer variations) in the gold standard contour set were not assessed. In this work, 3 human experts in radiation oncology performed manual contouring of the tumor in dynamic lung MR images using identical viewing/display conditions, in two different sessions. A subset of these contours is used to train the algorithm. The two objectives of this study are (a) To evaluate the difference of automatically determined contours from the manual ones drawn by different human experts, and (b) To contextualize the autocontouring algorithm’s agreement metrics with intra- and interobserver metrics, which represent inherent uncertainties in manual contours. We hypothesize that our algorithm can generate autocontours, with agreement metrics against standard manual contours that are comparable to intra- and interobserver variation metrics.

2. MATERIALS AND METHODS

2.A. Patient selection and MR image acquisition

In this institutional ethics approved pilot study, patients with nonsmall cell lung cancer (NSCLC) undergoing stereotactic radiotherapy are recruited for this study based on selection by a radiation oncologist. Six patients with stage 1 lung cancer, tumor diameter <5 cm were selected. Early stage lung cancer patients are most likely to benefit from the highly conformal, curative radiation treatment provided by MR based tumor tracking. The patients are scanned using a 3 T MRI (Philips Achieva, The Netherlands) using a dynamic, 2D balanced steady state free precession (bSSFP) sequence (Matrix size, 128 × 128, Voxel size = 3 mm × 3 mm × 20 mm, TE = 1.1 ms, TR = 2.2 ms, Imaging time per frame = 270 ms). Each imaging study was performed with the patients undergoing free breathing.

2.B. Manual contouring

Three experts (2 radiation oncologists with 10+ years of experience and a radiation oncology resident with 5 years of post graduate experience) manually delineated the tumor contours in the image series of each patient, both to train the algorithm and to serve as a reference for comparison to automatically drawn contours. To minimize bias arising from contouring hardware (i.e., monitor contrast), a common computer terminal, with the freely available CERR\textsuperscript{11} (Computation Environment for Radiation Therapy), a MATLAB based software, is used to perform all manual contours. As the CERR platform is not the standard contouring software at the clinic, the contouring experts are given training to familiarize with the software for the given tasks. For each of the 6 patients, each of the 3 experts is asked to manually contour the GTV in 130 dynamic images of the dataset (Session 1). Identical window and level settings (the default CERR setting for the MR images) are used in all contouring sessions to minimize variation; this process mimics the practice of applying standardized CT window/levels for the manual lung tumor contouring at our institute. To ensure independence, each of the contouring experts is blinded to the contours drawn by the other 2 experts. For intraobserver variation assessment, the experts, blinded to their previous contours, are asked to recontour the images at least 1 week after the first round of contouring (Session 2). In summary, a total of 6 patients × 3 oncologists × 2 sessions × 130 images are manually contoured.

2.C. Autocontouring algorithm

The autocontouring algorithm has been discussed in detail previously,\textsuperscript{6,12} and it is briefly explained here. Eckhorn\textsuperscript{13} introduced a type of neural networks to model the electrochemical mechanism of a cat’s visual cortex. A modified model, known as the Pulse Coupled Neural Networks (PCNN),\textsuperscript{14} has shown success in image processing applications such as image enhancement, segmentation and classification.\textsuperscript{15–17} In this model, a network comprises of a 2D matrix of individual neurons. Each individual neuron receives continuous input signals, but its output is binary (firing/not firing). The neurons, in close proximity to each other, are interlinked such that the output of a neuron depends not only on its own input but also those of its interlinked neighbors, thus, the neurons tend to fire together in clusters. In our MR images, a tumor consists of heterogeneous cluster of hyperintense pixels in comparison to surrounding healthy tissues. Each neuron in the model can emulate an image pixel; its input being the pixel’s value in the image. As the output of this neuron is influenced both by its pixel value and the pixel values of its neighbors, the neural network model serves to group heterogeneous tumor pixels together. Applying this algorithm iteratively enhances the contrast of the tumor compared to the surrounding healthy tissues.\textsuperscript{5}

The current clinical standard for segmenting tumor in an image is the manually drawn contour by the radiation oncologist. The algorithm design accounts for the individual preferences of clinicians, thus, it requires a small number of preparatory images for training. In a prospective implementation of this software, it is anticipated this training process will take place prior to treatment (i.e., during RT simulation), requiring an additional MRI scan. However, for this retrospective validation study, the first 30 images, out of the dataset of 130 images, are used for training (i.e., the training
images). The remaining 100 images (i.e., the tracking images) are used for evaluating the algorithm. The manually drawn contours (ROI_{STD}) on the training images are used to optimize the parameters of the algorithm for an individual patient. An example training image with a contour for each of the 6 patients is shown in Fig. 1.

2.C.1. Main contouring algorithm

The main algorithm can be summarized in the following steps: (a) Normalized cross correlation determines the approximate location of the tumor and yields [Fig. 2(a)], (b) Contrast enhancement, performed using the PCNN algorithm, yields [Fig. 2(b)], (c) Thresholding of the image via the Otsu’s Method yields [Fig. 2(c)], (d) Removal of small islands yields [Fig. 2(d)], (e) Smoothing of the tumor shape yields [Fig. 2(e)], (f) Morphology/Shifts, which includes erosion, dilation and translation, yields the final autocontour (ROI_{AUTO}) shown in [Fig. 2(f)]. The entire process takes <20 ms (Windows 7, Intel i7-2600K, 4GB RAM) per contour.

As the algorithm design adapts to the individual expert user’s contouring preferences, step 2 (contrast enhancement by neural network) and step 6 (morphological/linear shift) are governed by the parameters optimized using the individual user’s own training contour set. Due to this feature, the algorithm generates autocontours that reflect the individual user’s contouring preferences.

2.C.2. Parameter optimization via algorithm training

The algorithm requires manually drawn ROI_{STD} on the training images, which are used to find optimized parameters for each individual patient. As described by Yun et al., the adaptive particle swarm optimization determines the optimized parameters in 2–3 hr offline, prior to tracking. During the optimization process, test parameters are entered into the main algorithm which creates autocontours (ROI_{AUTO}). The

![Fig. 1. A sample dynamic lung sagittal MR image of each of the 6 patients, with ROI_{STD} contoured on the image.](image-url)

![Fig. 2. A summary of the autocontouring process, a normalized cross correlation gives the approximate location of the tumor (a), PCNN based contrast enhancement gives (b), thresholding leads to (c), removal of small islands leads to (d), smoothing leads to (e), and morphology/translation leads to final result in(f).](image-url)
optimal set of parameters yield the contours that are best matched to the training ROI\textsubscript{STD} in terms of the Dice Coefficient (DC), define as follows.

\[
DC = 2 \frac{\text{Area}(\text{ROI}\textsubscript{STD} \cap \text{ROI}\textsubscript{AUTO})}{\text{Area}(\text{ROI}\textsubscript{STD}) + \text{Area}(\text{ROI}\textsubscript{AUTO})}
\]

2.D. Evaluation of contours

For evaluation of the contours in the tracking images, we performed Automatic vs. Manual, and Manual vs. Manual contour comparisons. For automatic vs. manual comparisons, we evaluate the contour agreement between (a) trained automatic contours with its trainer’s contours in the same session, (b) trained automatic contours and its trainer’s contours in a different session, and (c) trained automatic contours against the contours of a different user. To present the results, the following notation is used: A\textsubscript{12} represents autocontours trained by expert 1 in session 2; whereas M\textsubscript{11} denotes manual contours drawn by expert 3 in session 1. The automatic contouring algorithm, trained using the contours from 30 training images (images 1–30) from each of the 6 manual contouring sessions (3 contourer x 2 sessions) is used to generate 6 sets of 100 automatic contours (ROI\textsubscript{AUTO}) on the 100 tracking images (images 31–130), labeled as A\textsubscript{11}, A\textsubscript{12}, A\textsubscript{21}, A\textsubscript{22}, A\textsubscript{31}, A\textsubscript{32}. These are compared against the 6 sets of manual contours (ROI\textsubscript{STD}) drawn on the same tracking images, namely, M\textsubscript{11}, M\textsubscript{12}, M\textsubscript{21}, M\textsubscript{22}, M\textsubscript{31}, M\textsubscript{32}. We group the data into 3 distinct types of comparisons: same user same session (SUSS) match, same user different session (SUDS) match, and different user (DU) match. As an example, A\textsubscript{11}, 100 ROI\textsubscript{AUTO} generated with training contours from user 1, session 1, can be compared against M\textsubscript{11}, which are 100 ROI\textsubscript{STD} generated by user 1 in session 1 (i.e., SUSS), or M\textsubscript{12}, which are ROI\textsubscript{STD} generated by the same user at a different session, (i.e., SUDS), or M\textsubscript{21}, M\textsubscript{22}, M\textsubscript{31}, M\textsubscript{32}, which are ROI\textsubscript{STD} generated by different users (i.e., DU).

For manual vs. manual comparisons M\textsubscript{11}, M\textsubscript{12}, M\textsubscript{21}, M\textsubscript{22}, M\textsubscript{31}, M\textsubscript{32} are compared against M\textsubscript{11}, M\textsubscript{12}, M\textsubscript{21}, M\textsubscript{22}, M\textsubscript{31}, M\textsubscript{32}. The SUSS comparison (i.e., M\textsubscript{11} vs. M\textsubscript{11}, etc.) is trivial, as they are identical contours. The SUDS manual vs. manual match (i.e., M\textsubscript{11} vs. M\textsubscript{31}, etc.) represents interobserver variations.

Two metrics for contour agreements are used: Firstly, DC, introduced in Section 2.C.2, evaluated the agreement between two sets of contours, (i.e., ROI\textsubscript{AUTO} vs. ROI\textsubscript{STD} for automatic vs. manual comparisons) in the 100 tracking images. It should be noted that our automatic algorithm has no prior knowledge of the 100 manually drawn ROI\textsubscript{STD} on the tracking images. These 100 ROI\textsubscript{STD} are compared against corresponding ROI\textsubscript{AUTO} to validate the algorithm performance only after the autocontouring session is completed. In addition to DC, the Hausdorff distance\textsuperscript{21} (d\textsubscript{H}) is also used as an alternative metric for comparing contours. To calculate d\textsubscript{H} between contours A and B, the following steps are taken. For every point on contour A, the shortest distance to any point in contour B is determined. The largest of these distances is denoted as d(A, B). Conversely, from every point on contour B, the shortest distance to any point in contour A is determined. The largest of these distances is denoted as d(B, A).

\[
d_H = \max(d(A, B), d(B, A))
\]

Unlike the area based DC, d\textsubscript{H} is quite sensitive to small discrepancies in the contours, even though those discrepancies have a small impact in the overall area of the contour and the DC. Thus, d\textsubscript{H} provides a useful alternative metric for our contour comparisons.

3. RESULTS

3.A. Automatic vs. manual comparisons

The six autocontour datasets, A\textsubscript{11}, A\textsubscript{12}, A\textsubscript{21}, A\textsubscript{22}, A\textsubscript{31}, A\textsubscript{32} are compared against the 6 sets of manual contours drawn on the same tracking images, namely, M\textsubscript{11}, M\textsubscript{12}, M\textsubscript{21}, M\textsubscript{22}, M\textsubscript{31}, M\textsubscript{32}. DC is shown in Table I, with the diagonal element representing the automatic vs. manual SUSS DC, bolded elements representing automatic vs. manual SUDS DC, and underlined elements representing automatic vs. manual DU DC. The overall mean and standard deviation (i.e., mean (SD)) values for automatic vs. manual SUSS DC is 0.90 (0.03), SUDS DC is 0.88(0.04), DU DC is 0.87(0.04).

The equivalent analysis for automatic vs. manual Hausdorff distance (d\textsubscript{H}) is shown in Table II. Overall, the mean (SD) values for SUSS d\textsubscript{H} is 3.8(1.6) mm, the SUDS d\textsubscript{H} is 4.3 (1.5) mm, and the DU match d\textsubscript{H} is 4.8(1.7) mm.

<table>
<thead>
<tr>
<th></th>
<th>M\textsubscript{11}</th>
<th>M\textsubscript{12}</th>
<th>M\textsubscript{21}</th>
<th>M\textsubscript{22}</th>
<th>M\textsubscript{31}</th>
<th>M\textsubscript{32}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A\textsubscript{11}</td>
<td>0.90(0.03)</td>
<td><strong>0.90(0.03)</strong></td>
<td>0.86(0.04)</td>
<td>0.87(0.05)</td>
<td>0.87(0.04)</td>
<td>0.87(0.04)</td>
</tr>
<tr>
<td>A\textsubscript{12}</td>
<td><strong>0.89(0.03)</strong></td>
<td>0.91(0.03)</td>
<td>0.84(0.04)</td>
<td>0.83(0.05)</td>
<td>0.85(0.04)</td>
<td>0.85(0.04)</td>
</tr>
<tr>
<td>A\textsubscript{21}</td>
<td>0.89(0.04)</td>
<td>0.90(0.04)</td>
<td>0.89(0.04)</td>
<td><strong>0.86(0.05)</strong></td>
<td>0.87(0.04)</td>
<td>0.87(0.04)</td>
</tr>
<tr>
<td>A\textsubscript{22}</td>
<td>0.86(0.05)</td>
<td>0.87(0.04)</td>
<td><strong>0.85(0.05)</strong></td>
<td>0.89(0.04)</td>
<td>0.88(0.04)</td>
<td>0.88(0.04)</td>
</tr>
<tr>
<td>A\textsubscript{31}</td>
<td>0.87(0.04)</td>
<td>0.88(0.04)</td>
<td>0.87(0.05)</td>
<td>0.88(0.04)</td>
<td>0.90(0.03)</td>
<td><strong>0.89(0.04)</strong></td>
</tr>
<tr>
<td>A\textsubscript{32}</td>
<td>0.88(0.04)</td>
<td>0.88(0.03)</td>
<td>0.88(0.04)</td>
<td>0.86(0.05)</td>
<td><strong>0.89(0.03)</strong></td>
<td>0.90(0.03)</td>
</tr>
</tbody>
</table>

Medical Physics, 45 (1), January 2018
3.B. Manual vs. manual comparisons (Intra- and Interobserver variations)

The manual contours for the 100 tracking images (i.e., images 31-130 in each data set), labelled M11, M12, M21, M22, M31, M32, are compared against each other. DC is shown in Table III. The diagonal column represents the SUSS manual vs. manual comparisons, which always returns “1” as they are identical contours. The bolded elements represent manual vs. manual SUDS match (i.e., intraobserver variations). Underlined elements represent the manual vs. manual DU match (i.e., interobserver variations). The overall mean(SD) values for SUDS/interobserver DC is 0.88(0.04), and the DU/interobserver DC is 0.87(0.04).

The manual to manual Hausdorff distances (dH, in mm) are shown in Table IV. The diagonal elements represent comparisons of identical contours (SUSS) thus always yields “0”. The bolded elements represent SUDS (i.e., intraobserver) dH, while underlined elements represent DU (i.e., interobserver) dH. The overall mean(SD) values for SUDS (i.e., intraobserver) dH is 4.3(1.7) mm and the DU/interobserver dH is 4.7 (1.6) mm.

A summary of the SUSS, SUDS and DC metrics for manual vs. manual and manual vs. automatic comparisons is shown in Table V.

4. DISCUSSION

Intra- and interobserver variations for manual contours of lung tumors are well documented in the literature, in CT, PET/CT fusion, and in 4DCT images. In terms of MRI of lung cancer, intra/inter observer metrics have been evaluated for breath held lung volumes in MR/CT images, as well as diffusion metrics. However, in terms of lung tumor target delineation, to the best knowledge of the authors (i.e., PubMed/Google Scholar), this is the first assessment of intraobserver and interobserver variations in the dynamic lung MR images for monitoring the changes in tumor shape and location in real-time. This is also the first study in which a user-trained autocontouring algorithm has been benchmarked against variation metrics of multiple expert users, as suggested by Valentini et al. on dynamic lung MR images.

The key objective of this study is to validate the autocontouring algorithm in the context of the variations in contours drawn by experts. If one simply assumed a single expert’s manual contours to be a “gold standard”, an approach
previously used by our group,\textsuperscript{6,12} and others,\textsuperscript{8} then the accuracy of algorithm is reflected in the same user same session match (SUSS) DC of 0.90(0.03) and d\textsubscript{H} of 3.8(1.6) mm. Clearly, the algorithm does not generate a perfect match (DC = 1, d\textsubscript{H} = 0) with the gold standard. However, this error may be attributed to the inherent variations in manual contouring process. If the same expert on a different session, or another expert, creates another set of manual contours on the same set of images that is compared against the original “gold standard”, significant variation can be observed, (i.e., intraobserver DC = 0.88(0.04), d\textsubscript{H} = 4.3(1.7) mm, interobserver DC = 0.87(0.04), d\textsubscript{H} = 4.7(1.6) mm). Since there is no way to determine which set of manual contours is the actual “truth”, the “gold standard” is only as accurate as these intra- and interobserver variations indicate. Hence, we argue that, if a user’s manually drawn contours are used as the standard for comparison, an autocontour is considered to be accurate if it has same-user automatic vs. manual agreement metrics (SUSS/SUDS) comparable to intraobserver variations of that individual user, and a different-user automatic vs. manual agreement metrics (DU) comparable to interobserver variations between that user and other users. Table V gives a side by side comparison of equivalent metrics for manual vs. manual and automatic vs. manual comparisons. In our limited study of a modest number of patients (6), observers (3), and sessions (2), these results indicate that our user trained automatic algorithm is indeed as accurate as the human experts, in both the same-user and different-user scenarios.

Since the patients in this study were imaged on a 3 T MRI system, a few comments on the general applicability of this study are warranted. In particular, the contrast with noise ratio of tumor depends on the main magnetic field strength\textsuperscript{29} of the MRI and the Linac-MR hybrids typically operate at lower field strengths\textsuperscript{2–5} than 3 T. Therefore, the quality of the acquired images and contouring performance may be different for the actual Linac-MR systems. Additionally, other factors, such as choice of MRI sequence, scan parameters (i.e., TE/TR, voxel size, etc.), and presence of image artifacts from various sources (including bias field artifacts, which are not corrected for in this study) may all affect image quality and contouring performance. However, in our previous work\textsuperscript{30–32} we have tested this autocontouring algorithm with images degraded retrospectively by additional noise and k-space undersampling, and has shown that the algorithm works reasonably well with lower image quality scenarios, albeit with slightly poorer performance metrics.

Our intra/interobserver agreement (mean DC: 0.88/0.87) are higher compared to the some previous lung tumor contouring studies on CT (mean DC:0.51/0.51) vs. 4DCT (mean DC: 0.80/0.80) images,\textsuperscript{24} as well as on nonregistered PET-CT (median DC: 0.58/0.61) vs. registered PET-CT (median DC: 0.71/0.70) images.\textsuperscript{23} However, there are major differences in the study protocols (i.e., 2D contouring vs. 3D contouring) and in imaging modality (dynamic MRI vs. CT/4DCT/PET-CT). Therefore, a direct comparison of the results in this study to the previous studies is not advised.

In summary, our trained algorithm’s ability to match its expert user’s contours (SUSS, SUDS) is comparable to the uncertainties from intraobserver variations from those experts, while our algorithm’s ability to match to a different expert’s contours (DU) is comparable to the uncertainties from interobserver variations. These results show that the autocontouring algorithm can, in less than 20 ms, produce contours on dynamic lung MR images that are comparable, in terms of accuracy, to that of human experts who generally takes several seconds to produce a manual contour. However, our interpretations of these results are limited to that of lung tumors, where there is a naturally a large amount of contrast between tumor and background. The ability of our algorithm to perform in other tumor sites (i.e., liver, pancreas) would require further investigations. In terms of implementation of the autocontouring algorithm on Linac-MR hybrids, we have integrated a prior version of this software with a preclinical 0.2 T Linac-MR prototype and demonstrated its capabilities to deliver conformal radiation to a moving phantom by imaging/autocontouring/MLC reshaping in real time.\textsuperscript{10} For real time tumor tracking radiation delivery in-vivo, further studies are needed to validate the algorithm in a larger cohort of patients, with the ultimate aim of integrating this software to perform real time tumor tracking with a 0.5 T clinical Linac-MR system\textsuperscript{5} (Aurora–RT, MagnetT\textsubscript{x} Oncology Solutions).

5. CONCLUSIONS

In this study, we evaluated our individual user-trained autocontouring algorithm on dynamic lung MR images for 6 NSCLC patients with multiple users. Autocontours from the trained algorithm for a particular expert agrees with the manual contours of the same user within the intraobserver variations. Autocontours from the trained algorithm by a particular expert agrees with the manual contours of different experts within the interobserver variations. These results suggest that the user trained autocontouring algorithm is capable of tracking a moving tumor on MR images as accurately as human experts, but with much faster speed (<20 ms). This algorithm may be a key component of the real time tumor tracking workflow for our hybrid Linac-MR device in the future.

ACKNOWLEDGMENTS

This work is supported by Alberta Innovates: Health Solutions, CRI0 Team grant.

CONFLICTS OF INTEREST

Fallone is a cofounder and CEO of MagnetT\textsubscript{x} Oncology Solutions.

\textsuperscript{a}Author to whom correspondence should be addressed. Electronic mail: eyip@ualberta.ca.
REFERENCES