## 6.1 Introduction

Colorectal cancer is the third most commonly diagnosed cancer in the United States(Siegel et al., 2019), with 60-70% of patients develop colorectal liver metastasis (CRLM) over the course of their disease(Adloff et al., 2017; Alper Sag et al., 2016; van der Pool et al., 2012). While surgery is the golden standard for curative intent, with an overall 5-year survival rate of ranging from 40% to 60%(Adam et al., 2012; Hackl et al., 2014; House et al., 2010; Nordlinger et al., 2013), it is only available in approximately 20% of patients at the time of diagnosis(Hosokawa et al., 2016).

Percutaneous ablation therapy (PTA), via microwave or radiofrequency ablation, has been shown to have similar overall 5-year survival rates, from 24% to 51%(Gillams & Lees, 2009; Kim et al., 2011; Solbiati et al., 2012), as surgery. Unfortunately, one of the largest challenges with PTA is ensuring sufficient ablation delivery.

Current work on assessing ablation efficacy has focused on 2D margin assessment(Fujioka et al., 2006), 3D assessment with rigid registration(Kaye et al., n.d.), or 3D margin assessment with some form of deformable registration(Sibinga Mulder et al., 2019)[xxx our JVIR paper]. Each of these methods relies on the single metric of minimum delivered margin, without context to the surrounding tissue.

Deep learning classification architectures have been shown to be successful extractors and interpreters of image information beyond what is easily identified by humans. Current deep learning work has produced outcome predictors with medical image in a number of tasks, from detection of pneumonia(Hashmi et al., 2020), to histopathological classification(Iizuka et al., 2020), and identification of radiosensitive populations(Lou et al., 2019).

We hypothesize that a deep learning classification model, with convolutional layers for feature extraction and dense layers for feature interpretation, could provide a useful prediction outcome model of ablation efficacy for the treating interventional radiologist.

## 6.2 Materials and Methods

### Data

#### Selection

We retrospectively acquired 119 patients who had undergone percutaneous ablation therapy (PTA) for metastatic colorectal liver metastasis (CRLM) at our institution. Patient selection criteria included: a minimum follow-up period of 6 months and a maximum of 2 months between pre-treatment and post-treatment contrast enhanced CT imaging. Follow-up data included three possible outcomes: residual disease, local progression, and no progression. For this work, labels 1 and 2 were combined as progression.

#### Segmentation

CRLM were manually segmented on pre-treatment contrast-enhanced CT scans and validated by a radiology trained physician fellow. CRLM segmentations were labeled based on the outcome status of the individual site as progression or non-progression, resulting in a total of 197 sites. An example of a patient with four sites, three of which did not locally progress, Figure 69.

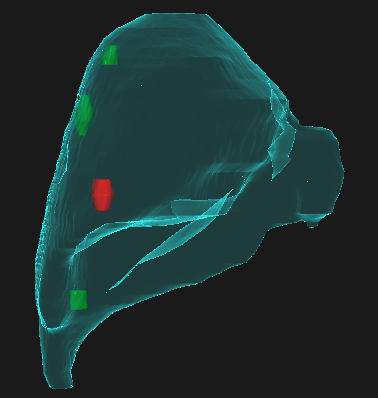


Figure 69: Example of patient with four disease sites. Three of the treated sites did not develop progression (green), and one site did progress (red).

Liver contours for the pre-treatment and post-treatment scans were generated automatically with our in-house liver auto-contouring algorithm(Anderson et al., 2020), manually inspected, and edited as needed.

#### Image Registration

Automatic, intensity based, rigid image registration including rotation was performed focused about the liver within the Raystation Treatment Planning system(Bodensteiner, 2018). Following rigid-registration, biomechanical model-based deformable image registration (Morfeus) was applied within the treatment planning system (RayStation v9B, RaySearch Laboratories, Stockholm, Sweden), using the liver contours as boundary conditions. Morfeus has been extensively validated in the liver (Al-Mayah et al., 2010, 2011; K. K. Brock et al., 2005; Kristy K. Brock et al., 2008; Cazoulat et al., 2016; Velec et al., 2017).

#### Data Extraction

The pre-treatment CT, rigidly registered post-treatment CT, deformable resampled post-treatment CT, and corresponding liver and tumor segmentations were converted into nifti files using our in-house DicomRTTool (Anderson et al., 2021). Using SimpleITK(Beare et al., 2018), all images and masks are resampled to dimensions of 5x1x1mm, with bilinear interpolation and nearest-neighbor resampling, respectively.

#### Presentation to model

Image ‘slabs’ of size 32x64x64 were extracted for each treated site, centered about the segmented disease. Several examples of the pre-treatment image (on the left) and the post-ablation (on the right) are shown in Figure 70.

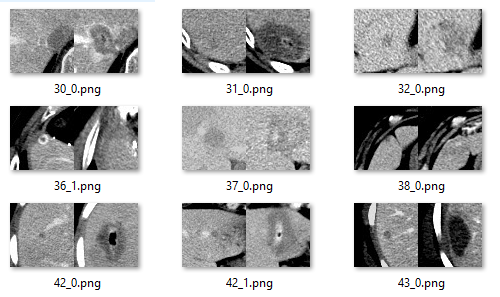


Figure 70: Examples of ‘slabs’ from a variety of patients. Image on the left is pre-treatment and image on the right is post-ablation. Note that the nomenclature is patient ID\_site#, so 42\_0 and 42\_1 are both from the same patient.

#### Distributing the data

Data was selected based on patient ID, not individual site, **ensuring that no patient is distributed between training and validation set if they have multiple treated sites.** One-third of the patients with any progression sites were randomly selected to be part of the validation set. We then randomly selected one-third of the patients who only had non-progression sites to also be in the validation. The final distribution of data can be seen in Table 18.

Table 18: Distribution of patients and lesions between train and validation set. Note that patients were split up based on patient ID, so a single patient will never be in both training and validation groups.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source Information | | | Status | Image Distribution | |
| Source | Patients | Lesions | Non-Progression | Train | Validation |
| Progression |
| MD Anderson  Cancer Center | 119 | 197 | 132 | 94 | 38 |
| 65 | 43 | 22 |

### Architectures

We investigated several alterations to convolutional neural networks for feature extraction, into densely connected networks for outcome. The framework of the feature extraction has three parts: First, convolution blocks are performed in a manner similar to the DenseNet(Huang et al., n.d.) architecture, where previous feature maps are concatenated immediately to deeper layers, enabling efficient backpropagation and reuse of filters. Second, a transition block is optimized to reduce the number of filters by a fraction ranging from 0.5 to 1, where 1 indicates no reduction and 3D max pooling with a stride of 2. Third, feature extraction ends with a global max pooling layer.

The framework of the densely connected network has two parts: First, a set number of dense connections followed by dropout layer. Second, the model predicts two classes, progression and no-progression, with a soft-max activation.

We investigated several parameters in both the feature extractor, densely connected prediction, and parameters outside of the architecture. Each architectural style was run three times to ensure poor initialization was not the root cause of poor performance.

Feature extraction parameters:

* 4, 8, 16, or 32 filters as the number of filters to start the feature extraction
* 4, 8, 16, or 32 as the rate of increase in the number of filters (growth rate)
* 1, 2, 3, 4, or 5 convolution blocks in each dense layers
* 1, 2, 3, or 5 dense convolution blocks followed by transition layers
* 0.5, 0.75, or 1.0 the fraction of filters to decrease by each transition

Densely connected parameters:

* 0, 1, 2, 3, 4, 5, or 7 number of densely connected layer
* 64, 128, 256, or 512 number of connections in each layer
* 0.5 or 0.0 dropout for fraction of connections to drop

Other hyper-parameters:

* Min and max learning rate were created on a model by model basis
* Adam or Stochastic gradient descent optimizers
* Categorical cross entropy, or Cosine loss

### Training the model

We investigated presenting the data in eight different ways, involving combinations of the pre-treatment image, the rigidly registered post-treatment image, the deformably registered post-treatment image, segmentations of the liver, and disease. A breakdown the combinations is shown in Table 19.

Table 19: The methods of presenting data to the model. Note that across all 8 the pre-treatment contrast enhanced CT is presented. Next, either the deformed or rigidly registered post-treatment image is presented. Segmentations of the liver and or disease are also included.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Imaging | | | Contours | |
| Pre-Treatment | Post-Treatment Deformed | Post-Treatment Rigid | Liver | Disease |
| x | x |  |  |  |
| x | x |  | x |  |
| x | x |  |  | x |
| x | x |  | x | x |
| x |  | x |  |  |
| x |  | x | x |  |
| x |  | x |  | x |
| x |  | x | x | x |

To balance the presentation of progression and no-progression, we created two TensorFlow data-generators with interleaving outputs. With a batch size of 32, this ensured each training step had 16 sites with progression and 16 sites without progression. Each generator will reshuffle and repeat, meaning the progression data is sampled more often than the no-progression data. This decision was influenced by recommendations from Tensorflow when working with imbalanced data, <https://www.tensorflow.org/tutorials/structured_data/imbalanced_data>.

#### Cosine Loss

Due to the limited, and biased nature of our dataset (69% of data is no-progression, 31% is progression), we quickly noticed a problem of model overfitting in the loss and AUC curves. Normalization was introduced to the model in the form of L2 normalization and dropout in the dense connections, unfortunately it did not seem to alleviate the overfitting. Dropout introduces sparsity into the dense connections, helping to alleviate overfitting, and L2 normalization attempts push the weights of the connections to be small but not 0.

Recent work has shown that a large part of overfitting due to biased datasets can be related to the usage of softmax + crossentropy loss function(Barz & Denzler, n.d.), and the authors propose the use of a cosine loss function. The cosine loss focuses on increasing the similarity between the L2-norm of the prediction and the ground truth, ‘bounding’ the loss to a unit sphere, Equation (1). Note that we add 1 to scale it from 0-2, rather than -1 to 1. In contrast, the cross entropy and softmax contains exponential and log functions, allowing arbitrarily high values to appear.

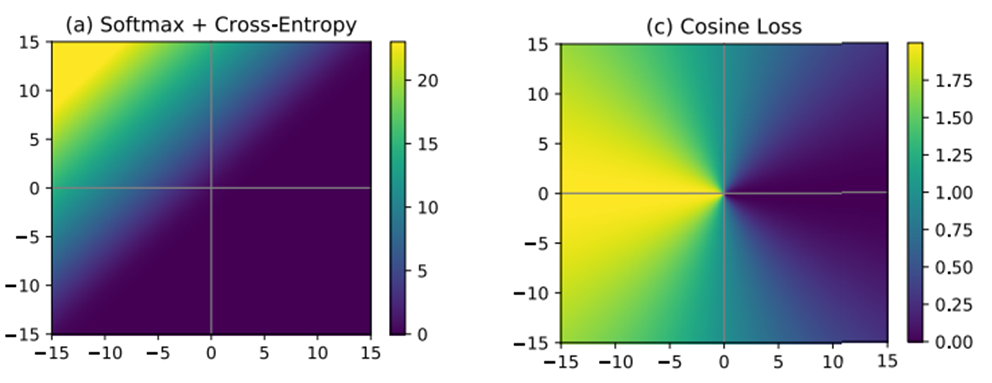


Figure 71: Heat maps of two loss functions in a 2-D feature space. Note that cosine loss is bounded from 0 to 2, while softmax + cross-entropy can take on arbitrarily large values. Figure copied from (Barz & Denzler, n.d.).

|  |  |  |
| --- | --- | --- |
|  |  | Equation (1) |

After adopting the cosine loss function the model loss curves appeared to have less overfitting and was adopted for the rest of the work.

#### Learning Rates

Unique minimum and maximum learning rates were identified using our in-house learning rate finder (GitHub link redacted), adapted from code presented in <https://www.pyimagesearch.com/2019/08/05/keras-learning-rate-finder/>. These minimum and maximum learning rates were used in a Cosine annealing learning rate scheduler (GitHub link redacted). Step-size was varied to be 2000, 5000, or 10,000, with a reduction factor of 0.5.

### Evaluating Training

#### Quantitative

Training was monitored using the Tensorboard module(Abadi et al., n.d.). Area under the curve (AUC) was the metric of choice for overall model performance, as well as sensitivity and specificity at a 50% prediction cutoff. Top performing models, as identified by the AUC, were selected for prediction visualization.

#### Qualitative

To provide human interpretable explanations about the model predictions, we use the integrated gradient (IG) method of identifying which aspect of the image led to model prediction. IG were originally proposed by Sundarajaran et al.(Sundararajan et al., 2017) to identify which aspects of the image were most important in the decision process. Tensorflow offers a tutorial on implementation of the IG method (<https://www.tensorflow.org/tutorials/interpretability/integrated_gradients>) which we adapted from 2D images to 3D images as seen in our data. Each top performing model from the eight possible input data presentations given in Table 19 was investigated qualitatively.

## 6.3 Results

### Quantitative

We present the results between ‘pre-treatment and deformable post-treatment’ or ‘pre-treatment and rigidly registered post-treatment’. The accuracy, sensitivity, specificity, and the AUC for each data presentation method are presented in Table 20. Note that sensitivity and specificity assume a 50% prediction cutoff.

Table 20: Accuracy, percentage of progression sites accurately predicted, percentage of non-progression sites accurately predicted, and the area under the curve (AUC) values for the eight presentation methods of data. Note that accuracy and percentage correct assumes a prediction cut-off of 50%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Accuracy** | **Sensitivity** | **Specificity** | **AUC** |
| Primary + Secondary Deform | 67 | 18 | 95 | 0.66 |
| Primary + Secondary Deform + Liver | 70 | 23 | 97 | 0.68 |
| Primary + Secondary Deform + GTV | 77 | 55 | 89 | 0.81 |
| Primary + Secondary Deform + GTV + Liver | 70 | 36 | 89 | 0.78 |
| Primary + Secondary Rigid | 63 | 0 | 100 | 0.66 |
| Primary + Secondary Rigid + Liver | 67 | 32 | 87 | 0.66 |
| Primary + Secondary Rigid + GTV | 72 | 32 | 95 | 0.72 |
| Primary + Secondary Rigid + GTV + Liver | 67 | 23 | 92 | 0.75 |

With the deformable registered images, increasing the information presented to the model raised the overall performance from just the pre-treatment and post-treatment images (AUC=0.66) to both images and disease segmentation (AUC 0.81). Adding the liver to the deformed images and disease segmentation had a small decrease in AUC. The received operating characteristic (ROC) lines for the individual data presentation methods of pre-treatment and deformed post-treatment images are shown in Figure 72.

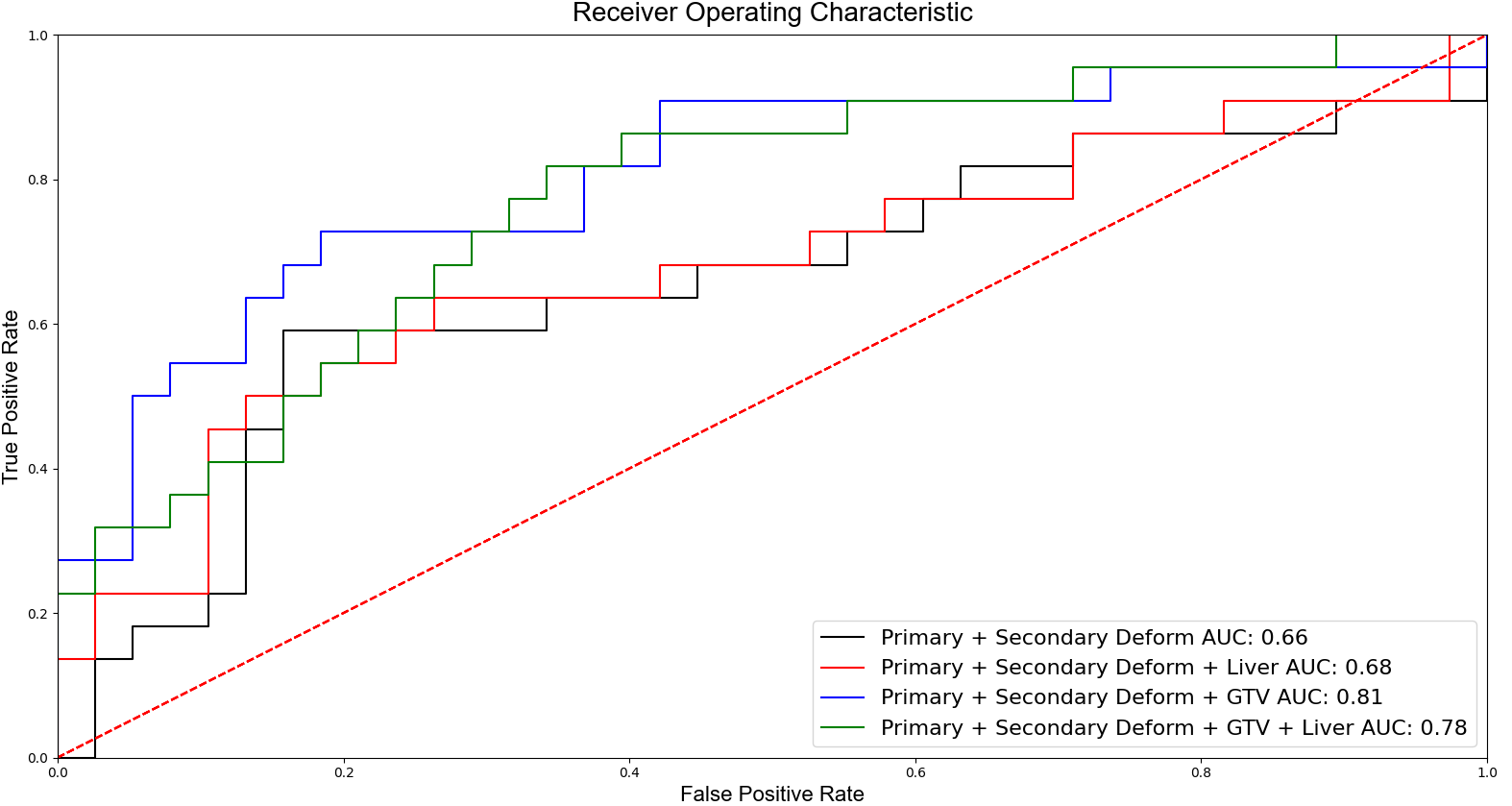


Figure 72: Receiver operating characteristic lines for each presentation of pre-treatment and deformable registered post-treatment data to the model. Area under the curve (AUC) recorded in legend.

With the rigidly registered images, increasing the information presented to the model raised the overall performance from just the pre-treatment and post-treatment images (AUC=0.66) to both images, disease and liver segmentation (AUC 0.75). Adding the liver to the images had no increase in AUC, while adding the disease increased AUC to 0.72, and liver plus disease to 0.75 and disease segmentation had a small decrease in AUC. The received operating characteristic (ROC) lines for the individual data presentation methods of pre-treatment and rigidly registered post-treatment images are shown in Figure 73.

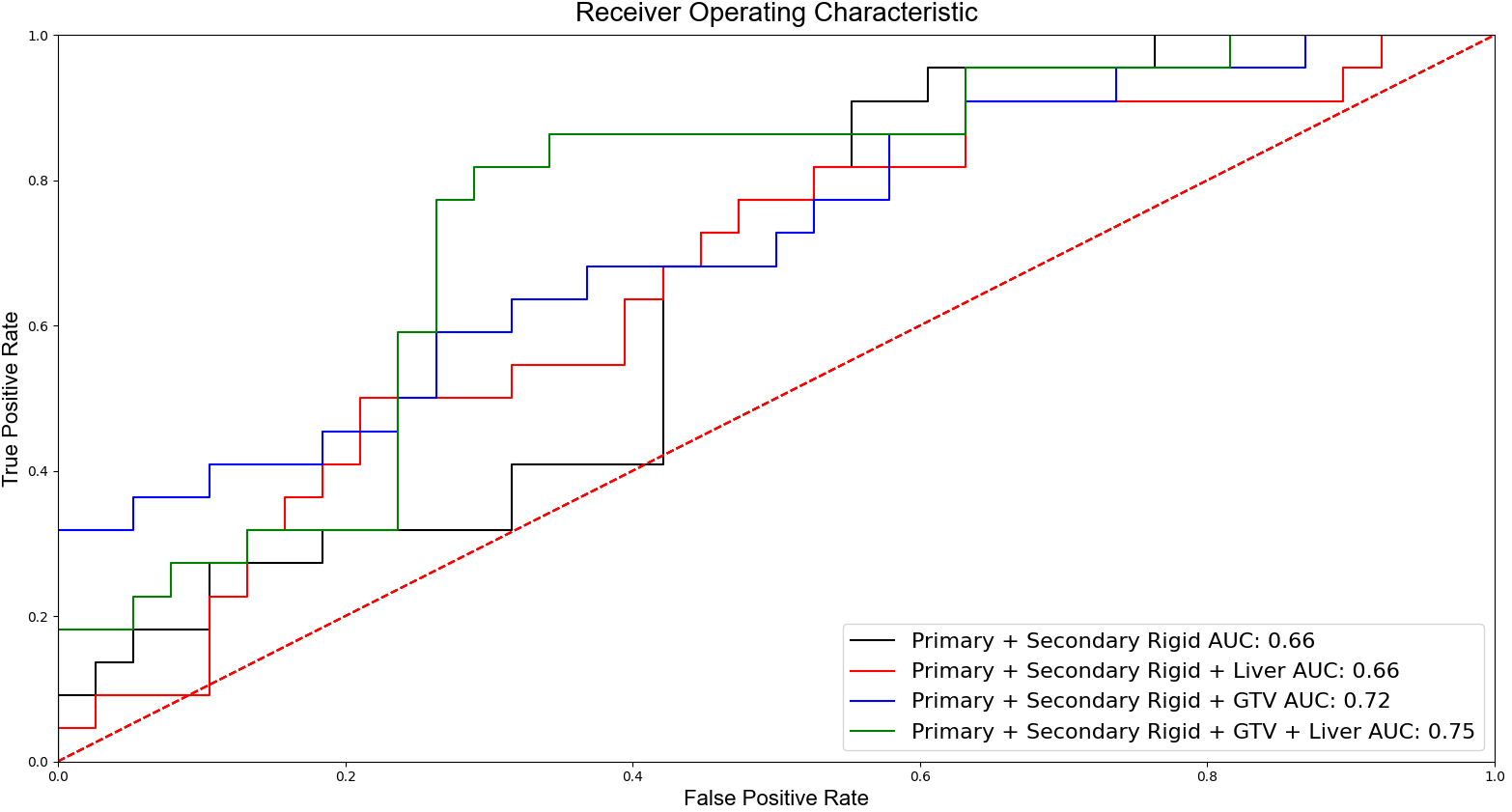


Figure 73: Receiver operating characteristic lines for each presentation of pre-treatment and rigidly registered post-treatment data to the model. Area under the curve (AUC) recorded in legend.

### Qualitative

Qualitative visualization of the integrated gradients (IG) varied across models and sites. We used visualizations of the IG to qualitatively identify if the model was focusing on aspects of the image relevant to the PTA procedure. For example, in Figure 74, the model was presented the pre-treatment image, rigidly registered post-treatment image, and the contours of the liver and disease. Visualizing the IG, the model appears to be most focusing on regions overlapping between the ribs and lung, an artifact of the rigid registration, rather than near the disease site. This would lead us to believe that the model’s prediction were based on aspects outside of what would we consider useful in the decision process for this case.



Figure 74: Overlapping heat map from Integrated Gradients on a progression case. Images (from left to right) are Pre-treatment CT, rigidly registered post-treatment CT, pre-treatment liver, disease, and post-treatment liver. Note that disease is blank, as the disease is not present in these slices. However, the IG show importance was heavily weighted to an area far from the disease.

## Discussion

Across all eight models there appears to be a similar relationship to the AUC as different data is presented. The presentation of the images alone for both rigid and deformable perform equally poorly, with an AUC of 0.66. We believe this is due to the model being presented with a relatively large search space without guidance on the important aspects of the image, and that the limited overall training data does not facilitate sufficient examples before overfitting occurs. Adding the information of the liver segmentation provides minimum improvement to the deformable registration model and none to the rigid registration model, likely due to the fact that the data is centered on the disease within the liver already, simply providing the liver contour doesn’t inform the model much on important aspects. In addition, it may distract the model since the entire liver boundary is not consistently presented in all of the cases, causing additional random variability within the training set.

Both models show a marked improvement when adding the segmentation of the disease information, from 0.66 to 0.81 with deformable registration model, and from 0.66 to 0.72 with rigid registration. We believe that this information helps the model to focus more closely on what is important, being the disease and the corresponding ablation zone between the two images.

With the deformable registration model, adding the liver and disease segmentations caused a slight decrease in performance. We believe that this is due to the fact that the liver contour is not providing substantial new information in the deformed model, as the two livers are already deformed to each other. Visualization of the IG shows that the model sometimes focuses on the boundary edge of the liver contour, rather than the disease segmentation, Figure 75.



Figure 75: Overlay of the integrated gradients onto images presented to the deformable model. From the left, the pre-treatment CT, post-treatment CT, liver contour, and disease. Note that the IG are focusing on the boundary of the liver contour in the top right corner and not the disease.

In the rigid model, adding the liver with the disease contour provided a slight increase in AUC, we speculate this is because the registration is not perfect in the presence of deformation, and the liver contours help orient the model slightly. Despite this increase in AUC, the rigid model’s best AUC is still only 0.75, compared to 0.81 with the deformable registration model. We believe this due to the rigid model’s inability to model deformation present between the pre-treatment and post-treatment images.

It is important to note that the IG method does not tell the user *exactly* why a decision was made. The confidence of a prediction by the model is not directly tied to which aspects of the image would cause a certain output. For example, if the model predicted that the treated site was likely to progress, the output visualizations are a representation of the question ‘which aspects of the image would cause the model predict progression with the highest confidence’. This question can be asked of the model even if the model predicts that no progression is likely to occur. For this reason, it is very important to not only look at the IG heat map, but also look at the model’s prediction confidence. Furthermore, if the model predicts progression, but the IG are not anywhere near the disease/ablation volume, the user should be skeptical as to the benefit of the model’s prediction.

The IG can be useful not only for identifying regions of potentially insufficient ablation, but also for assurances of sufficient ablation. In Figure 76, a case with no local progression, which the model accurately predicted as no progression, we can see that the gradients are focused on the exact center of the disease site and is fully encapsulated by the ablation cavity. We speculate that this is the model identifying that sufficient ablation was delivered around each aspect of the site.

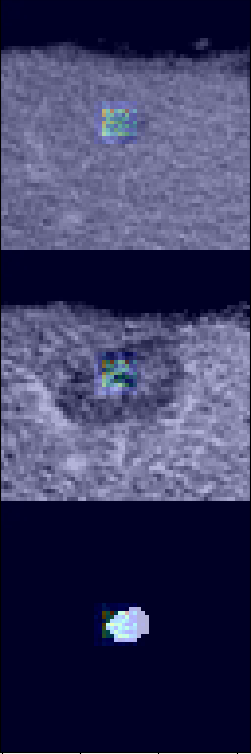


Figure 76: Overlapping heat map from the integrated gradients on a non-progression case. Images (from left to right) are pre-treatment CT, deformable registered post-treatment CT, and disease segmentation. The model’s prediction of ‘non-progression’ appears to have focused entirely on the disease site encapsulated within the ablation volume.

Exact metrics of what corresponds to a useful AUC can vary based on application. A score of 0.5 suggests the model has no ability to discriminate between two outcomes, equivalent to the flip of a coin. Values from 0.7 to 0.8 are generally acceptable, 0.8-0.9 are excellent, and 0.9-1.0 are outstanding(Mandrekar, 2010).

Depending on the conservativeness of the user, they might prefer a higher specificity to sensitivity, or vice-versa. For this work, we believe that a higher sensitivity to local progression is most important, hopefully identifying regions of insufficient ablation with the combination of prediction and IG map. The desire to be more conservative means the user should lean towards a potential addition ablation if the prediction for local progression is slightly less than 50%. This would lead to an increased sensitivity in identifying patients with potentially insufficient ablation, but would also lead to an increase in the number of patients receiving potentially unnecessary additional ablation. The desired balance between these two potential outcomes will be determined by the individual user, and their knowledge of the particular case being treated.

## Conclusion

We have created an outcome prediction model which assess the likelihood of outcome based on pre-treatment contrast enhanced CT, a deformed post-treatment contrast enhanced CT, and segmentations of the disease site with an AUC of 0.81. Visualization of the integrated gradients for a given prediction should be used to assess model efficacy, and can help identify potential regions of interest in the treatment images. We believe that the prediction output of the model along with the integrated gradients can help provide meaningful information to the interventional radiologist.

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Research, G. (n.d.). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. Retrieved July 28, 2020, from www.tensorflow.org.

Adam, R., De Gramont, A., Figueras, J., Guthrie, A., Kokudo, N., Kunstlinger, F., Loyer, E., Poston, G., Rougier, P., Rubbia‐Brandt, L., Sobrero, A., Tabernero, J., Teh, C., & Van Cutsem, E. (2012). The Oncosurgery Approach to Managing Liver Metastases from Colorectal Cancer: A Multidisciplinary International Consensus. *The Oncologist*, *17*(10), 1225–1239. https://doi.org/10.1634/theoncologist.2012-0121

Adloff, M., Arnaud, J. P., Thebault, Y., Ollier, J. C., & Schloegel, M. (2017). Hepatic metastases from colorectal cancers. *Chirurgie - Memoires de l’Academie de Chirurgie*, *116*(2), 144–149. https://doi.org/10.5005/jp-journals-10018-1241

Al-Mayah, A., Moseley, J., Hunter, S., Velec, M., Chau, L., Breen, S., & Brock, K. (2010). Biomechanical-based image registration for head and neck radiation treatment. *Physics in Medicine and Biology*, *55*(21), 6491–6500. https://doi.org/10.1088/0031-9155/55/21/010

Al-Mayah, A., Moseley, J., Velec, M., & Brock, K. (2011). Toward efficient biomechanical-based deformable image registration of lungs for image-guided radiotherapy. *Physics in Medicine and Biology*, *56*(15), 4701–4713. https://doi.org/10.1088/0031-9155/56/15/005

Alper Sag, A., Selcukbiricik, F., & Mandel, N. M. (2016). Evidence-based medical oncology and interventional radiology paradigms for liver-dominant colorectal cancer metastases. In *World Journal of Gastroenterology* (Vol. 22, Issue 11, pp. 3127–3149). Baishideng Publishing Group Co., Limited. https://doi.org/10.3748/wjg.v22.i11.3127

Anderson, B. M., Lin, E. Y., Cardenas, C. E., Gress, D. A., Erwin, W. D., Odisio, B. C., Koay, E. J., & Brock, K. K. (2020). Automated Contouring of Contrast and Noncontrast Computed Tomography Liver Images With Fully Convolutional Networks. *Advancesradonc*. https://doi.org/10.1016/j.adro.2020.04.023

Anderson, B. M., Wahid, K. A., & Brock, K. K. (2021). Simple Python Module for Conversions between DICOM Images and Radiation Therapy Structures, Masks, and Prediction Arrays. *Practical Radiation Oncology*. https://doi.org/10.1016/j.prro.2021.02.003

Barz, B., & Denzler, J. (n.d.). *Deep Learning on Small Datasets without Pre-Training using Cosine Loss*.

Beare, R., Lowekamp, B., & Yaniv, Z. (2018). Image segmentation, registration and characterization in R with simpleITK. *Journal of Statistical Software*, *86*(1), 1–35. https://doi.org/10.18637/jss.v086.i08

Bodensteiner, D. (2018). RayStation: External beam treatment planning system. *Medical Dosimetry*, *43*(2), 168–176. https://doi.org/10.1016/j.meddos.2018.02.013

Brock, K. K., Sharpe, M. B., Dawson, L. A., Kim, S. M., & Jaffray, D. A. (2005). Accuracy of finite element model-based multi-organ deformable image registration. *Medical Physics*, *32*(6Part1), 1647–1659. https://doi.org/10.1118/1.1915012

Brock, Kristy K., Hawkins, M., Eccles, C., Moseley, J. L., Moseley, D. J., Jaffray, D. A., & Dawson, L. A. (2008). Improving image-guided target localization through deformable registration. *Acta Oncologica*, *47*(7), 1279–1285. https://doi.org/10.1080/02841860802256491

Cazoulat, G., Owen, D., Matuszak, M. M., Balter, J. M., & Brock, K. K. (2016). Biomechanical deformable image registration of longitudinal lung CT images using vessel information. *Physics in Medicine and Biology*, *61*(13), 4826–4839. https://doi.org/10.1088/0031-9155/61/13/4826

Fujioka, C., Horiguchi, J., Ishifuro, M., Kakizawa, H., Kiguchi, M., Matsuura, N., Hieda, M., Tachikake, T., Alam, F., Furukawa, T., & Ito, K. (2006). A Feasibility Study. Evaluation of Radiofrequency Ablation Therapy to Hepatocellular Carcinoma Using Image Registration of Preoperative and Postoperative CT. *Academic Radiology*, *13*(8), 986–994. https://doi.org/10.1016/j.acra.2006.05.011

Gillams, A. R., & Lees, W. R. (2009). Five-year survival in 309 patients with colorectal liver metastases treated with radiofrequency ablation. *Eur Radiol*, *19*, 1206–1213. https://doi.org/10.1007/s00330-008-1258-5

Hackl, C., Neumann, P., Gerken, M., Loss, M., Klinkhammer-Schalke, M., & Schlitt, H. J. (2014). Treatment of colorectal liver metastases in Germany: A ten-year population-based analysis of 5772 cases of primary colorectal adenocarcinoma. *BMC Cancer*, *14*(1). https://doi.org/10.1186/1471-2407-14-810

Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics*, *10*(6). https://doi.org/10.3390/diagnostics10060417

Hosokawa, I., Allard, M. A., Gelli, M., Ciacio, O., Vibert, E., Cherqui, D., Sa Cunha, A., Castaing, D., Miyazaki, M., & Adam, R. (2016). Long-Term Survival Benefit and Potential for Cure after R1 Resection for Colorectal Liver Metastases. *Annals of Surgical Oncology*, *23*(6), 1897–1905. https://doi.org/10.1245/s10434-015-5060-8

House, M. G., Ito, H., Gönen, M., Fong, Y., Allen, P. J., DeMatteo, R. P., Brennan, M. F., Blumgart, L. H., Jarnagin, W. R., & D’Angelica, M. I. (2010). Survival after Hepatic Resection for Metastatic Colorectal Cancer: Trends in Outcomes for 1,600 Patients during Two Decades at a Single Institution. *Journal of the American College of Surgeons*, *210*(5), 744–752. https://doi.org/10.1016/j.jamcollsurg.2009.12.040

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (n.d.). *Densely Connected Convolutional Networks*. Retrieved July 28, 2020, from https://github.com/liuzhuang13/DenseNet.

Iizuka, O., Kanavati, F., Kato, K., Rambeau, M., Arihiro, K., & Tsuneki, M. (2020). Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-58467-9

Kaye, E. A., Cornelis, F. H., Petre, E. N., Tyagi, N., Shady, W., Shi, W., Zhang, Z., Solomon, S. B., Sofocleous, C. T., & Durack, J. C. (n.d.). *Volumetric 3D assessment of ablation zones after thermal ablation of colorectal liver metastases to improve prediction of local tumor progression*. https://doi.org/10.1007/s00330-018-5809-0

Kim, K. H., Yoon, Y. S., Yu, C. S., Kim, T. W., Kim, H. J., Kim, P. N., Ha, H. K., & Kim, J. C. (2011). Comparative analysis of radiofrequency ablation and surgical resection for colorectal liver metastases. *Journal of the Korean Surgical Society*, *81*(1), 25. https://doi.org/10.4174/jkss.2011.81.1.25

Lou, B., Doken, S., Zhuang, T., Wingerter, D., Gidwani, M., Mistry, N., Ladic, L., Kamen, A., & Abazeed, M. E. (2019). An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. *The Lancet Digital Health*, *1*(3), e136–e147. https://doi.org/10.1016/S2589-7500(19)30058-5

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Nordlinger, B., Sorbye, H., Glimelius, B., Poston, G. J., Schlag, P. M., Rougier, P., Bechstein, W. O., Primrose, J. N., Walpole, E. T., Finch-Jones, M., Jaeck, D., Mirza, D., Parks, R. W., Mauer, M., Tanis, E., Van Cutsem, E., Scheithauer, W., & Gruenberger, T. (2013). Perioperative FOLFOX4 chemotherapy and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC 40983): Long-term results of a randomised, controlled, phase 3 trial. *The Lancet Oncology*, *14*(12), 1208–1215. https://doi.org/10.1016/S1470-2045(13)70447-9

Sibinga Mulder, B. G., Hendriks, P., Baetens, T. R., Van Erkel, A. R., Van Rijswijk, C. S. P., Van Der Meer, R. W., Van De Velde, C. J. H., Vahrmeijer, A. L., Mieog, J. S. D., & Burgmans, M. C. (2019). Quantitative margin assessment of radiofrequency ablation of a solitary colorectal hepatic metastasis using MIRADA RTx on CT scans: A feasibility study. *BMC Medical Imaging*, *19*(1), 71. https://doi.org/10.1186/s12880-019-0360-2

Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, *69*(1), 7–34. https://doi.org/10.3322/caac.21551

Solbiati, L., Ahmed, M., Cova, L., Ierace, T., Brioschi, M., & Goldberg, S. N. (2012). Small liver colorectal metastases treated with percutaneous radiofrequency ablation: Local response rate and long-term survival with up to 10-year follow-up. *Radiology*, *265*(3), 958–968. https://doi.org/10.1148/radiol.12111851

Sundararajan, M., Taly, A., & Yan, Q. (2017). *Axiomatic Attribution for Deep Networks*.

van der Pool, A. E. M., Damhuis, R. A., Ijzermans, J. N. M., de Wilt, J. H. W., Eggermont, A. M. M., Kranse, R., & Verhoef, C. (2012). Trends in incidence, treatment and survival of patients with stage IV colorectal cancer: A population-based series. *Colorectal Disease*, *14*(1), 56–61. https://doi.org/10.1111/j.1463-1318.2010.02539.x

Velec, M., Moseley, J. L., Svensson, S., Hårdemark, B., Jaffray, D. A., & Brock, K. K. (2017). Validation of biomechanical deformable image registration in the abdomen, thorax, and pelvis in a commercial radiotherapy treatment planning system. *Medical Physics*, *44*(7), 3407–3417. https://doi.org/10.1002/mp.12307