Automatic Vessel Intensity Measurement for Quality Control of Contrast-enhanced CT

Kevin McCoy[1,2] Dominique Tan[3] Sujay Marisetty[4] Christine B. Peterson[2] Moiz Ahmad[2]

[1]Department of Statistics, Rice University [2]The University of Texas, MD Anderson Cancer Center [3]The University of Texas, Austin [4]Rice University

Background

- Adequate image enhancement of organs and blood vessels of interest is an important aspect of image quality in contrast-enhanced computed tomography (CT).
- Current methods of image evaluation entail subjective review.
- There is a need for an objective method for evaluation of vessel contrast that can be automatically applied to large sets of CT exams.
- In this work, we present a method of automatically segmenting the portal vein so that its mean pixel value can be reported which serves as a proxy for image quality.

Methods

- 82 contrast enhanced abdominal CT scans[1] were obtained from The Cancer Imaging Archive, which were originally compiled for testing automatic pancreas segmentation. However, the CT scans cover the entire abdomen and are suitable for analysis of portal vein enhancement. A previous review identified a large variation in image noise in this data set[2]; therefore, this was selected as a training data set with potentially diverse portal vein enhancement.
- In the first stage of the method, images were smoothed, filtered (for vessel / tube-ness)[3], and finally pixel thresholded[4]. We then extracted shape and intensity features from candidate portal vein segmentations using ImageJ.
- Final dataset included 126,216 rows (candidate segmentations) and 21 columns (features).
- The second stage uses machine learning algorithms to classify the segmentations based on their features, and outputs the aggregated pixel value in the positively classified segmentations.
- A random 80/20 train:test split was used. Importantly, we split the data across patients, rather than segmentations, to avoid introducing bias due to potential within-subject correlations.
- Various ML models were tested, including Naive Bayes, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Logistic Regression, K Nearest Neighbors, Support Vector Machine, Random Forest, gradient boosting, and Neural Networks.
- A modification of the RF base model was introduced (RF Model 2) such that this model returns *at least* the top three most likely PV segmentations, even if $p < 0.5$.
- All computations carried out using python and standard sklearn / xgboost implementations.

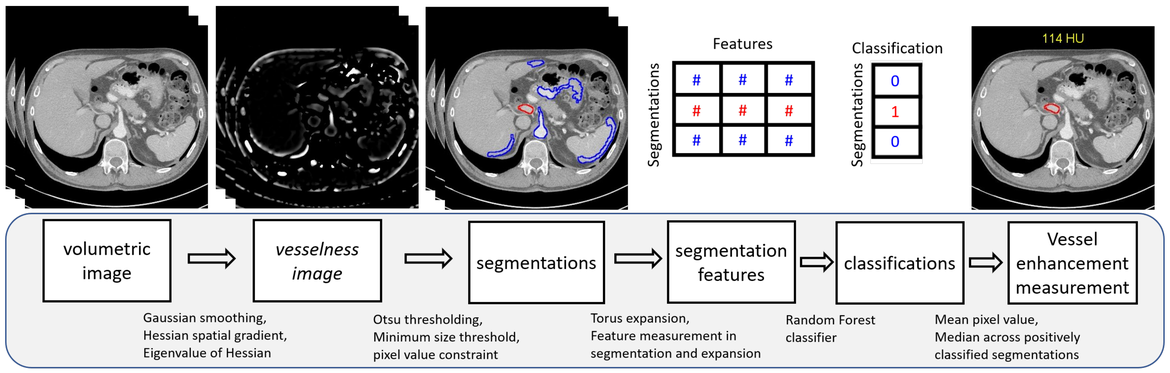

Figure 1. Flowchart diagram illustrating the automated vessel enhancement measurement method.

Results

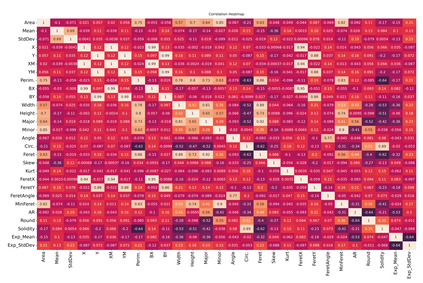

Figure 2. Feature Correlation Heatmap.

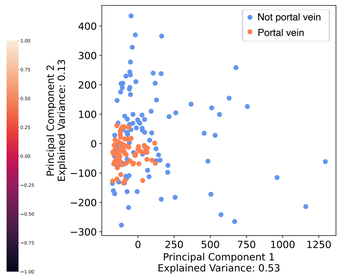

Figure 3. Principle Component Analysis (PCA) of raw data.

	NB	LDA	QDA	LR	KNN
Precision	0.072	0.288	0.103	0.099	0.519
Recall	0.854	0.020	0.919	0.899	0.397
AUC	0.835	0.510	0.893	0.880	0.695

	SVM	**Random Forest**	Boosting	Neural Network
Precision	0.202	**0.745**	0.733	0.698
Recall	0.921	**0.254**	0.394	0.531
AUC	0.930	**0.626**	0.696	0.764

Table 1. Precision, Recall, and AUC of different machine learning classifiers.

	RF Base Model	RF Model 2
Train set (5 fold CV)		
Precision	0.892	0.845
Recall	0.196	0.227
AUC	0.598	0.613
Test set		
Precision	0.986	0.902
Recall	0.425	0.446
AUC	0.712	0.723

Table 2. Random Forest classifier performance after hyperparameter tuning.

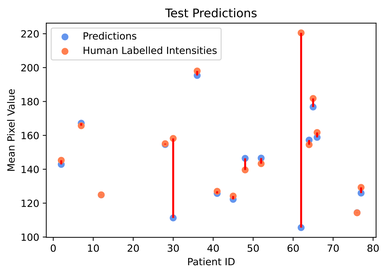

Figure 4. Test Set PV measurement accuracy.

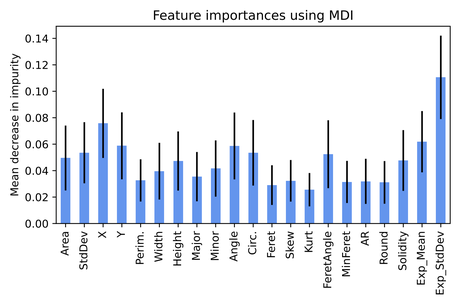

Figure 5. Feature Importances.

Results

	RF Base Model	RF Model 2
Ext. Val Data		
Precision	0.676	0.630
Recall	0.157	0.289
AUC	0.578	0.644

Table 3. Performance on external validation data.

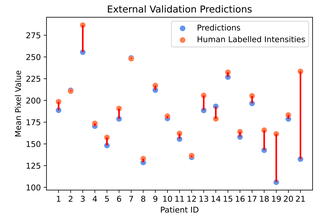

Figure 6. Validation Set PV measurement accuracy.

Discussion

- The proposed ML pipeline may detect scan timing or contrast administration problems. If tabulated over a review period, these could be tracked longitudinally as a quality control measure. The method may also be generalized to other major blood vessels.
- First, we found that good overall automatic measurement accuracy is achieved, despite imperfect precision of the classifier in distinguishing PV segmentations. Second, we have identified Random Forest as a suitable ML method for the present problem, and boosting and neural networks are other potential candidates. Third, we have explained the relative importance of different segmentation features. We found the features related to the immediate vessel vicinity are especially important.
- Classification performance was substantially lower in the Validation set compared to the Test set, and further method refinement and increased classification accuracy is desired. In particular, intra-slice and intra-scan correlations could be used to improve the classification accuracy.
- Our ML method treats candidate segmentation as observations individually. However, there are two levels of grouping that we ignore: multiple segmentations within a slice and within a volumetric image set. Additionally, the presence of a portal vein in a particular slice may influence the prediction confidence of a segmentation in an adjacent slice at a similar (x,y) location. Furthermore, there is typically one portal vein segmentation per slice.

References

[1] H. Roth, A. Farag, E. B. Turkbey, L. Lu, J. Liu, and R. M. Summers, "Data from Pancreas-CT (Version 2) [Data set]." The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU, 2016.

[2] M. Ahmad, M. C. Jacobsen, M. A. Thomas, H. S. Chen, R. R. Layman, and A. K. Jones, "A benchmark for automatic noise measurement in clinical computed tomography," *Medical Physics*, vol. 48, no. 2, pp. 640–647, 2021.

[3] Y. Sato, S. Nakajima, H. Atsumi, T. Koller, G. Gerig, S. Yoshida, and R. Kikinis, "3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," in *International Conference on Computer Vision, Virtual Reality, and Robotics in Medicine*, pp. 213–222, Springer, 1997.

[4] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1842494, as well as The Ken Kennedy Institute Computational Science and Engineering Recruiting Fellowship, funded by the Energy HPC Conference.