

# Automatic vessel attenuation measurement for quality control of contrast-enhanced CT: Validation on the portal vein

Kevin McCoy<sup>1,2</sup> | Sujay Marisetty<sup>3</sup> | Dominique Tan<sup>4</sup> | Corey T. Jensen<sup>5</sup> | Jeffrey H. Siewerdsen<sup>6</sup> | Christine B. Peterson<sup>2</sup> | Moiz Ahmad<sup>6</sup>

<sup>1</sup>Department of Statistics, Rice University, Houston, Texas, USA

<sup>2</sup>Department of Biostatistics, Division of Basic Science Research, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>3</sup>Rice University, Houston, Texas, USA

<sup>4</sup>The University of Texas, Austin, Texas, USA

<sup>5</sup>Department of Abdominal Imaging, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>6</sup>Department of Imaging Physics, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

## Correspondence

Moiz Ahmad, Department of Imaging Physics, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

Email: [MAhmad@mdanderson.org](mailto:MAhmad@mdanderson.org)

## Funding information

National Cancer Institute, Grant/Award Number: P30CA016672; National Science Foundation, Grant/Award Number: 1842494; Cancer Prevention and Research Institute of Texas, Grant/Award Number: RP210028

## Abstract

**Background:** Adequate image enhancement of organs and blood vessels of interest is an important aspect of image quality in contrast-enhanced computed tomography (CT). There is a need for an objective method for evaluation of vessel contrast that can be automatically and systematically applied to large sets of CT exams.

**Purpose:** The purpose of this work was to develop a method to automatically segment and measure attenuation Hounsfield Unit (HU) in the portal vein (PV) in contrast-enhanced abdomen CT examinations.

**Methods:** Input CT images were processed by a vessel enhancing filter to determine candidate PV segmentations. Multiple machine learning (ML) classifiers were evaluated for classifying a segmentation as corresponding to the PV based on segmentation shape, location, and intensity features. A public data set of 82 contrast-enhanced abdomen CT examinations was used to train the method. An optimal ML classifier was selected by training and tuning on 66 out of the 82 exams (80% training split) in the public data set. The method was evaluated in terms of segmentation classification accuracy and PV attenuation measurement accuracy, compared to manually determined ground truth, on a test set of the remaining 16 exams (20% test split) held out from public data set. The method was further evaluated on a separate, independently collected test set of 21 examinations.

**Results:** The best classifier was found to be a random forest, with a precision of 0.892 in the held-out test set to correctly identify the PV from among the input candidate segmentations. The mean absolute error of the measured PV attenuation relative to ground truth manual measurement was 13.4 HU. On the independent test set, the overall precision decreased to 0.684. However, the PV attenuation measurement remained relatively accurate with a mean absolute error of 15.2 HU.

**Conclusions:** The method was shown to accurately measure PV attenuation over a large range of attenuation values, and was validated in an independently collected dataset. The method did not require time-consuming manual contouring to supervise training. The method may be applied to systematic quality control of contrast-enhanced CT examinations.

## KEYWORDS

computed tomography, contrast enhancement, image quality, machine learning, quality control, random forest

## 1 | INTRODUCTION

Computed tomography (CT) is an essential diagnostic modality in the modern health care enterprise. A large percentage of all CT imaging examinations are performed with contrast enhancement to aid the detection or characterization of disease, using an intravenous administration of iodinated contrast media. Contrast enhancement is essential in the detection of many cancers, for example, liver metastases of colorectal cancer, that are otherwise undetectable without contrast media.<sup>1</sup> Even with the use of contrast media and state-of-the-art CT imaging technology, subtle lesions are often difficult to detect.<sup>2,3</sup> Suboptimal imaging is produced if insufficient contrast agent is delivered to the anatomy of interest or if the scan is mistimed relative to the vascular dispersion of contrast agent.<sup>1</sup> Potential consequences are misdiagnosis or the need to repeat the scan. Both situations are harmful to the patient and a burden on the health care system. Therefore, quality control (QC) has a critical role in maximizing the diagnostic accuracy of contrast-enhanced CT imaging. Although the incidence of poor contrast enhancement in CT imaging may be limited to a low percentage of outliers, the exact incidence is unknown, and methods are needed to detect and track incidence of these outliers.

An optimum contrast enhancement depends on several factors: an appropriate contrast administration protocol (i.e., the selected contrast agent, amount, injection pressure, and flow rate), appropriate synchronization between intravenous injection and CT scan, and optimal scan parameters to maximize enhancement. Although the medical physics field has provided mature QC methodologies ranging from CT system performance evaluation<sup>4</sup> to observer models of detection,<sup>5</sup> the literature on methods for QC of contrast enhancement is sparse and only generic contrast administration guidelines are available.<sup>6,7</sup> QC of contrast enhancement is understandably challenging: many QC methods rely on physical phantom measurements as a convenient means of testing imaging performance without experimentation on living subjects, but these phantoms lack physiological realism. Computational human phantoms provide more realistic physiology,<sup>8,9</sup> and may have a role in commissioning a new CT scan or contrast protocol. However, it is difficult to anticipate patient physiological variation, contrast mis-administration events, and patient dependent image acquisition and reconstruction parameters in phantom tests. Direct QC enhancement assessment in clinical images may therefore be a useful adjunct to phantom testing in a QC program, and may provide information on the quality of contrast administration and scan timing.<sup>10</sup> Contrast quality assessment can be done through extensive manual data collection,<sup>11</sup> but the size of such studies is limited.

Automated measurements could be used in a systematic review of large numbers of CT exams. Abadi et al. presented a method of automatically measuring attenuation of the abdominal aorta in chest CT based on the assumption of a circular vessel cross-section in axial format images, and used this method to quantify aortic attenuation in a large patient cohort.<sup>12</sup> Machine learning (ML) and deep learning (DL) image segmentation are promising methods that may aid in the more general vessel measurement problem. Although DL methods have been successfully applied to organ segmentation, they have generally not been widely explored for vessel segmentation. One challenge is the high variation in vessel enhancement across CT exams (relative to more consistent organ parenchymal enhancement), making it difficult to apply a DL method, which typically require a large amount of diverse training data with manual ground truth segmentations for supervised training. Furthermore, the present demonstrations of DL vessel segmentation<sup>13–17</sup> are intended for specific applications like surgical guidance, where there is an implicit assumption of good vessel enhancement and image quality.

Rather than using a DL segmentation on *unstructured* image data, we propose instead using a Random Forest classifier<sup>18</sup> operating on *structured* image feature data. The classifier's objective is to correctly select the segmentation corresponding to a specific vessel from among a set of input candidate segmentations, which are generated by a classical imaging processing method. Segmentation features describing the location, shape, and pixel value distribution are provided to the classifier as structured input features. This approach differs from DL where relevant features are learned from unstructured image data to correctly segment target anatomy. Random Forest methods tend to outperform DL methods on structured input data and offer better interpretability.<sup>19</sup> This motivates the development of a simple ML method for segmenting a blood vessel based on input candidate segmentations.

In this work, we present a method of automatically segmenting a blood vessel so that a quantitative measurement of its attenuation can be reported. We chose the portal vein (PV) in routine contrast-enhanced abdomen CT as a target vessel. The enhancement of the PV is related to liver parenchymal enhancement, which is relevant to the difficult task of detecting subtle liver metastases. Furthermore the complex structure of the PV and the existence of many normal anatomical variants<sup>20</sup> presents a challenge to accurate segmentation. The method presented here solves this problem by first applying classical image processing to enhance and segment tube-like structures in the image volume, and second, using a ML method to determine whether or not a segmentation corresponds to the PV. Several ML algorithms were tested to identify a best algorithm

for this application. The proposed method was designed to meet the following criteria. First, it must have high *precision* in classification of PV segmentations. Also called positive predictive value, precision is defined as  $\frac{TP}{TP+FP}$ , where  $TP$  is the number of true positive and  $FP$  is the number of FP classifications. Note that high precision is a more stringent requirement than high *specificity* in this application, since false segmentation classification affects PV attenuation measurement accuracy. Second, the automated PV intensity measurement must be accurate over a broad range of possible PV intensity levels: ( $\sim 100$ – $300$  Hounsfield unit [HU]).

## 2 | METHODS

### 2.1 | Training and validation data

We identified a publicly available Pancreas-CT image data set<sup>21</sup> in The Cancer Imaging Archive.<sup>22</sup> This dataset includes 82 contrast enhanced abdominal CT scans and was originally compiled for testing automatic pancreas segmentation. However, the CT scans cover the entire abdomen and are suitable for analysis of PV attenuation. A previous review identified a large variation in image noise in this data set.<sup>23</sup> This public dataset was selected for training and testing our method because of the diverse PV attenuation. The slice thickness in this data set ranged from 1.5 to 2.5 mm, and the CT scans were acquired on two different scanners (Philips and Siemens models) with 120 kVp tube voltage. Additional image metadata are documented on the Pancreas-CT Collection site.<sup>21</sup>

### 2.2 | Feature extraction

In the first stage of the method, we extract shape, location, and intensity features for candidate PV segmentations. All image preprocessing and feature extraction were performed using ImageJ.<sup>24</sup> The image is first smoothed with a 2D Gaussian filter with  $\sigma = 3$  mm. The  $\sigma$  parameter value was chosen in consideration of the population average PV diameter. For other target vessels, the value should be set proportional to vessel diameter. Next, we applied the 3D *tubeness* filter (also called *vesselness*), which uses the Hessian spatial gradients method.<sup>25</sup> This filter results in enhancement of all tubular structures in the image, including blood vessels, bones, organ boundaries, and other curvilinear objects with tube radius close to the  $\sigma$  of the 2D Gaussian filter. The *vesselness* filter sparsified the image, making it more amenable to segmentation. The threshold used for segmentation is determined by finding the 10th percentile of intensity of the *vesselness* image. The resulting binary mask image is then divided into separate segmentations based on 2D connectedness

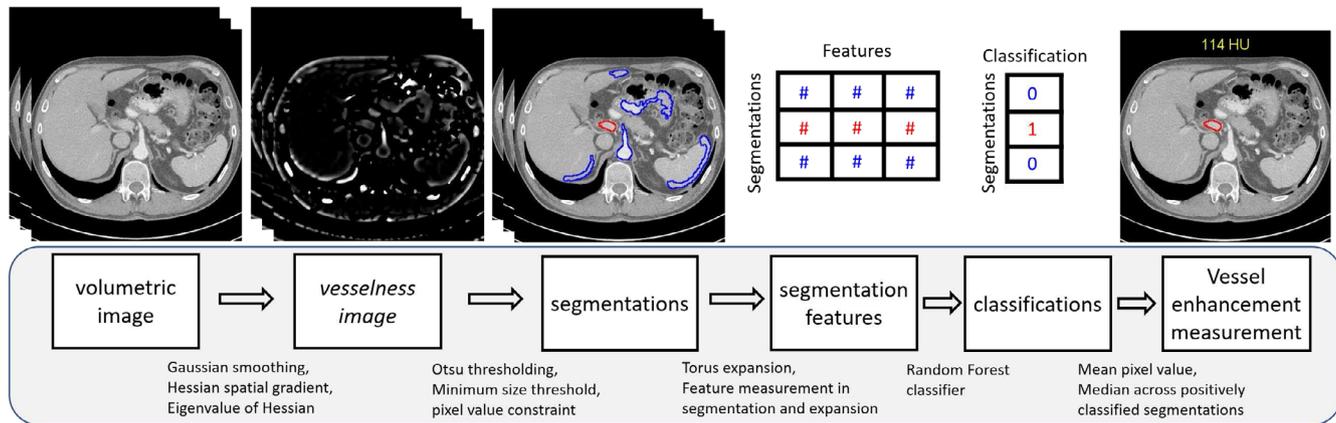
of pixels in the mask. Although the *vesselness* operation is performed in 3D, the image segmentation was performed in 2D in the axial plane, rather than in 3D. The reason for this choice is two-fold: First, 3D segmentation was highly variable from patient to patient, especially with regard to the inclusion of the network of inter-connected blood vessels. 3D segmentation features such as volume, surface area, and so forth tend to be inconsistent. Second, 3D segmentations tended to erroneously include structures outside of blood vessels through erroneous connections distributed throughout the 3D volume. In comparison, 2D axial segmentations are much more local and avoid these two problems.

Our method yields an ensemble of candidate 2D segmentations. We quantified shape, location, and intensity features of the candidate segmentations to be used as inputs to the classifier. We used the built-in library of feature definitions in ImageJ.<sup>26</sup> Most of these features (e.g., mean pixel value, segmentation area) are axiomatic in definition. The min and max “Feret” features represent the minimum and maximum distance between two parallel tangents of the segmentation at any arbitrary angle.<sup>27</sup> In addition to these default features, we defined new features aimed at characterizing the region immediately surrounding each candidate segmentation. We created a 5 mm expansion of each candidate segmentation to define a new segmentation: a “rind” around the original segmentation. The expansion is implemented using the “Make Band” function in ImageJ, which uses morphological dilation to expand the segmentation by  $n$  pixels; we set  $n$  to correspond to 5 mm. We extracted the pixel mean and standard deviation within the rind, labeled “Exp\_Mean” and “Exp\_StdDev” respectively. Location features are relative to the patient. In total, we extracted 28 unique features for 126 216 candidate segmentations across the Pancreas-CT image data set.

The second stage of our method used a ML classifier, described in a following section, to determine which of the candidate segmentations correspond to the PV. Next, the mean pixel value was calculated for each segmentation. Finally, the median value across positively classified segmentations was taken as the output attenuation measurement. A flowchart diagram depicting the vessel attenuation measurement is provided in Figure 1.

### 2.3 | Ground truth annotation

Each candidate segmentation was manually annotated with a binary label (negative or positive) corresponding to whether or not the segmentation was the PV. These annotations were performed by a board certified diagnostic medical physicist and two trained research assistants. Each observer annotated all segmentations, resulting in three annotations per segmentation. The majority consensus was taken as the ground truth



**FIGURE 1** Flowchart diagram of the automated vessel attenuation measurement method. The first stage of the method produces candidate vessel segmentations and their quantified features. The red segmentation is the ground truth PV definition, while the blue segmentations are ground truth negatives. The second stage identifies uses ML to classify the segmentations based on their features, and outputs the aggregated pixel value in the positively classified segmentations. ML, machine learning; PV, portal vein.

classification of the segmentation. The annotation guidance defined the PV as the length of the PV from the junction of superior mesenteric vein with the splenic vein to the bifurcation of the PV into right-anterior and right-posterior PVs. A segmentation was labeled positive if any portion of the segmentation included any part of the PV as defined above, provided that the segmentation did not erroneously include unconnected structures outside of the vessel. In addition to this manual classification, each observer also measured the PV attenuation using an oval region-of-interest with width and height between 4 and 5 mm, placed in the PV immediately proximal to the right PV branch. A single 2D Region of interest (ROI) measurement was taken on the slice chosen by the observer to most prominently display the PV. The mean measurement across observers was taken as the ground truth PV attenuation.

## 2.4 | Data summarization and cleaning

Exploratory data analysis was performed to understand the distribution, correlations, and redundancies of features in the training dataset. First, principal component analysis (PCA) was performed for dimensionality reduction and visualization. Next, feature correlations were computed to identify highly-correlated (redundant) features, which were excluded from subsequent analysis. These analyses were conducted on the 80% training set to avoid peeking at the hold-out validation set, described in more detail in the following section.

## 2.5 | Classifier development

Next, we developed a ML classifier to classify PV segmentations on the basis of extracted segmentation features. We first split the Pancreas-CT collection into “Training” and “Hold-out Test” sets, with a ran-

dom 80%/20% split. To avoid training bias, we split the data across patients, rather than segmentations, since segmentations for the same patient are potentially correlated. This split resulted in 66 and 16 exams in the Training and Hold-out Test sets, respectively. The Hold-out Test data were not used until a final model was chosen; that is, the presented model was not adjusted based on findings on the Hold-out Test set. The 66 exams in the Training set were further split into five groups for the purpose of five-fold cross-validation for parameter tuning.

After extracting the features for each candidate segmentation, we provided these features to a ML classifier for PV classification. The data were normalized such that each feature had a mean of 0 and a standard deviation of 1 across the Training set. The resulting normalization was applied to the test and validation data sets. We tested nine ML methods: Naïve Bayes (NB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR), K Nearest Neighbors (KNN),<sup>28</sup> Support Vector Machine (SVM),<sup>29</sup> Random Forest,<sup>18</sup> Gradient Boosting, and Neural Network.<sup>30,31</sup> Each model, with the exception of Gradient Boosting, used the standard *scikit-learn*<sup>32</sup> Python library implementation. XGBoost<sup>33</sup> was used for gradient boosting for its computational speed over the *scikit-learn* implementation of the algorithm. All models used default model hyperparameters except for class weights; given the class imbalance in our problem (negative class outnumbering positive class by an approximate 75:1 ratio), we used “balanced” class weights proportional to class size if class weights were available in that classifier model.

We performed hyperparameter tuning using only the Training data to further optimize the performance of the default *scikit-learn* Random Forest classifier implementation (default values of the most relevant parameters - *n\_estimators*:100, *criterion*: “gini”, *class\_weight*: “None”, *max\_depth*: “None”, *max\_features*: “auto”). The tuned

classifier is called “RF Base Model”, and the only parameter value different from default was:  $n\_estimators:1000$ .

The Random Forest makes a positive classification if the calculated classification probability  $p > 0.5$  for a segmentation. Based on the observation the Random Forest (RF) Base Model made zero PV positive classifications in a few patient image sets, the RF Base Model was further modified (“RF Model 2”) such that it returns *at least* the top three most likely PV segmentations, even if  $p < 0.5$ .

To evaluate classifier performance, we computed the precision, recall (sensitivity), and area under the Receiver operating characteristic (ROC) curve (AUC). In our application, false negative classifications of segmentations are less concerning than false positives (FP), particularly since there are multiple image slices in which a PV can be identified. On the other hand, a FP vessel classification results in an inaccurate attenuation measurement. Lastly, the accuracy of automatically measured PV attenuation was taken as an end-to-end figure of merit of the overall method.

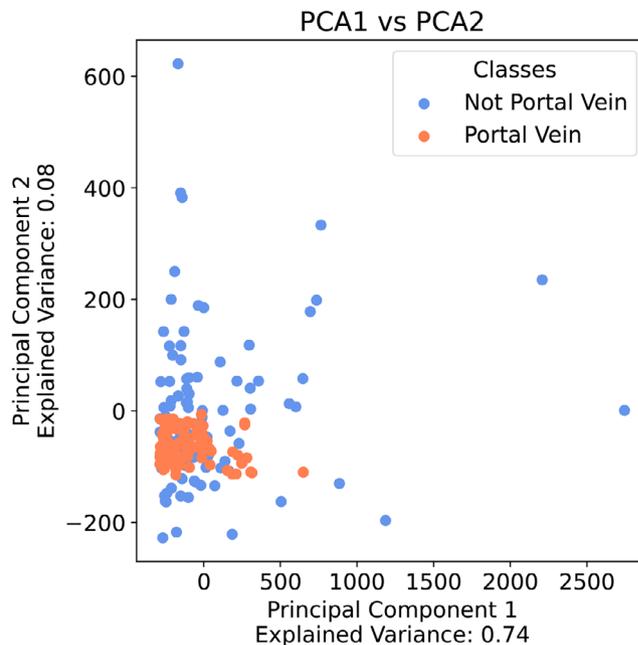
## 2.6 | Independent test set

The method was tested on a separate data set of 21 consecutive contrast-enhanced abdomen CT scans performed at one of our institutions (“Independent Test Set”). The image data were accessed retrospectively under a protocol approved by the Institutional Review Board and informed patient consent was waived. All scans used 120 kV tube voltage and were timed at the portal venous phase at a delay of 50 s after opacification reached +100 HU in the descending aorta at the level of the celiac trunk. A patient size-based contrast bolus of 125 or 150 mL at an injection flow rate of 3 or 4 mL/s was administered. Patients with an abdominal effective diameter<sup>34</sup> of 30 cm or greater received the higher volume and injection rate. The contrast agent was either 350 mg I/mL concentration Omnipaque (GE Healthcare, Waukesha, WI, USA) or 320 mg I/mL concentration Visipaque (GE Healthcare), depending on the patient’s renal function; the bottle concentrations of these two agents were considered to be practically equivalent.

The precision, sensitivity, and AUC of our method in discriminating automated segmentations were assessed; the accuracy of automated attenuation measurement in the PV was also assessed. The same model evaluated on the Hold-out Test set was also applied to the Test set with no further model tuning.

## 2.7 | Application to scan timing assessment

To motivate application to QC, we assessed whether the PV measurement may be a quantitative marker of mistiming on a few example CT exams. We found four examples of timing reported as early or late as deter-



**FIGURE 2** PCA projection plot – PV segmentations tend to be highly clustered in the first two principal components of features, but show substantial overlap with non-PV segmentations. Plotted data were randomly sub-sampled to 100 points in both classes for visual clarity. PCA, principal component analysis; PV, portal vein.

mined by interpreting radiologist. These observations were recorded as free text notes in an electronic communication system used by radiologists, technologists, and medical physicists to report and troubleshoot technical imaging issues. Some of these examples also had available comparisons of normally-timed CT scans of the same patient performed on a different date. The attenuation of the PV and the inferior vena cava (IVC), and the ratio of PV-to-IVC for these CT scans were measured manually, and the PV attenuation was also measured automatically.

## 3 | RESULTS

A PCA<sup>35</sup> plot of segmentation features in the Training and Hold-out Test sets revealed that the PV segmentations were less heterogeneous as a group than the non-PV segmentations (Figure 2). The tight clustering of the PV class suggests a classifier based on the unsupervised linear projection created by PCA may achieve high sensitivity. However, the plot also shows substantial overlap of the two classes, suggesting that a linear classifier would achieve poor precision.

A feature correlation heat-map (Figure 3) shows that a few strongly correlated features can be excluded. The horizontal location features X, XM, BX, FeretX were redundant; similarly, the vertical location features Y, YM, BY, and FeretY were redundant, based on correlation coefficient  $\rho > 0.93$ . All of these features except X and Y (centroid coordinates) were dropped in

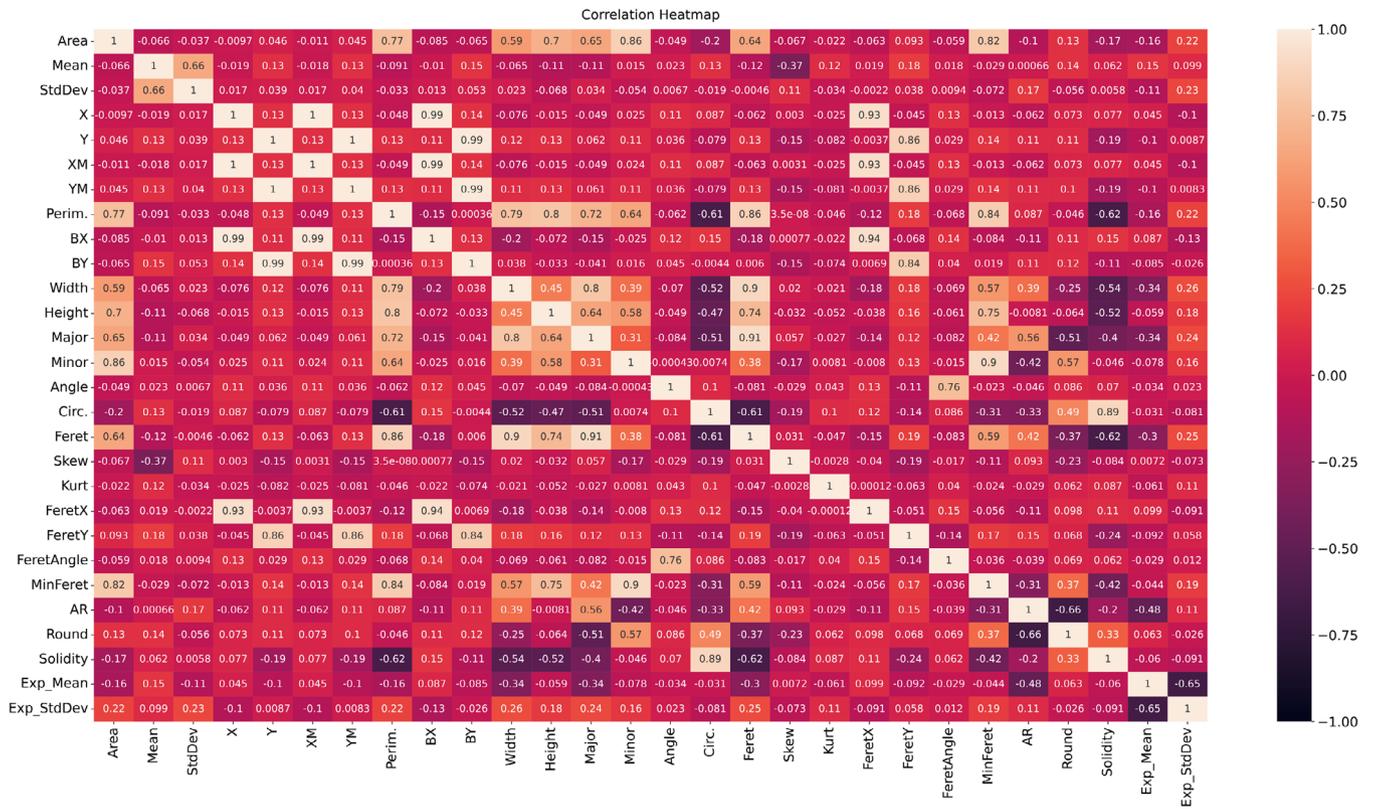


FIGURE 3 Heat map of feature correlations – Different position features were highly correlated and deemed to be redundant.

TABLE 1 Precision, Recall, and AUC of different ML classifiers in the Training Set without any tuning.

	NB	LDA	QDA	LR	KNN
Precision	0.072	0.288	0.103	0.099	0.519
Recall / Sensitivity	0.854	0.020	0.919	0.899	0.397
AUC	0.835	0.510	0.893	0.880	0.695
	SVM	Random Forest	Gradient boosting	Neural network	
Precision	0.202	<b>0.745</b>	0.733	0.698	
Recall / Sensitivity	0.921	<b>0.254</b>	0.394	0.531	
AUC	0.930	<b>0.626</b>	0.696	0.764	

Abbreviations: AUC, area under the ROC curve; KNN, K nearest neighbors; LDA, linear discriminant analysis; LR, logistic regression; ML, machine learning; NB, naïve bayes; QDA, quadratic discriminant analysis; SVM, support vector machine.

subsequent analysis. The Mean feature (mean pixel value of segmentation) was not used as this is the objective quantity of interest. In all, 21 features remained after feature reduction.

### 3.1 | Comparison of ML methods

The average precision, recall, and AUC over the five-fold cross-validation split in the Training Set are shown in Table 1. Random Forest was found to be the best off-the-shelf ML method according to precision, with gradient boosting and neural network close behind. Sev-

eral other methods achieved a higher AUC; however, the AUC is based on the ROC function of sensitivity and specificity, and does not describe the precision-sensitivity trade-off. The Random Forest classifier was selected for further optimization.

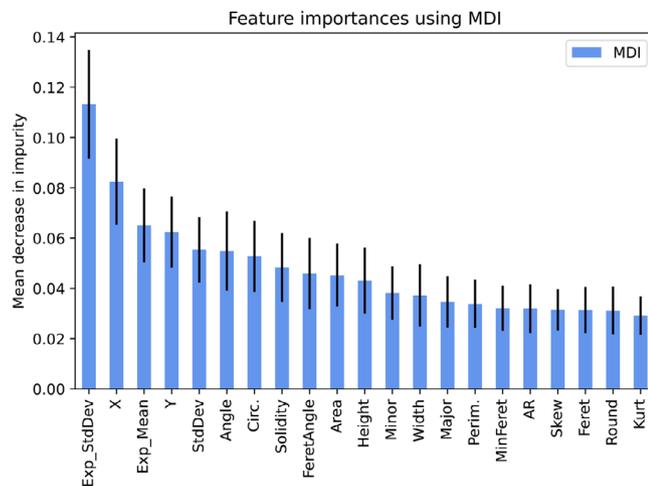
### 3.2 | Training and hold-out test performance

The performance of the RF Base Model and RF Model 2 over five-fold cross-validation in the Training set and in the Hold-out Test set is presented in Table 2. Overall,

**TABLE 2** Random Forest classifier performance after hyperparameter tuning.

	RF Base Model	RF Model 2
<b>Training set (five fold CV)</b>		
Precision	0.892	0.845
Recall / Sensitivity	0.196	0.227
AUC	0.598	0.613
<b>Hold-out Test set</b>		
Precision	0.892	0.788
Recall / Sensitivity	0.223	0.235
AUC	0.611	0.617

Abbreviation: AUC, area under the ROC curve.

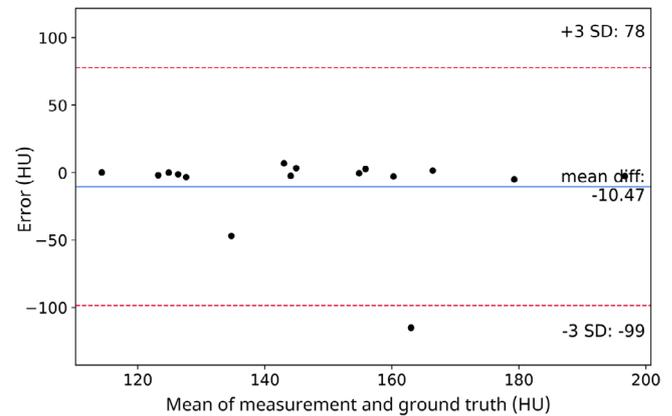


**FIGURE 4** Feature importance scores - MDI value indicates relative importance. Black bars represent 1 standard deviation of MDI values collected across each tree in the forest. MDI, mean decrease in impurity.

the method accurately identifies the PV segmentations: 0.892 and 0.788 precision in the Hold-out Test set for RF Base Model and RF Model 2, respectively. The modification in RF Model 2 trades precision for a small gain in sensitivity. Although small, this gain is helpful in those individual exams where the RF Base Model makes zero positive classifications. RF Model 2 was used in subsequent analyses.

An illustrative example of the model robustness is found in Figure 1. The depicted CT exam had one of the lowest PV attenuation in the Hold-out Test set (114 HU) because the scan was timed relatively early in the late-arterial phase rather than in the intended portal venous phase. Note how the arteries show greater enhancement than the PV. Despite the low attenuation, the PV was correctly identified in this case.

Feature importance scores were also extracted from the final RF model and are shown in Figure 4. Importance scores are reported as mean decrease in impurity (MDI),<sup>36</sup> a measure of how much each feature helps the model split the data by class across the leaf nodes



**FIGURE 5** PV measurement accuracy in the Hold-out Test Set – A Bland-Altman plot of the measurement errors, automated minus ground truth, as a function of varying PV attenuation. PV, portal vein.

**TABLE 3** Classification performance in the Independent Test Set.

	RF Base Model	RF Model 2
Precision	0.684	0.649
Recall / Sensitivity	0.164	0.302
AUC	0.581	0.650

Abbreviation: AUC, area under the ROC curve.

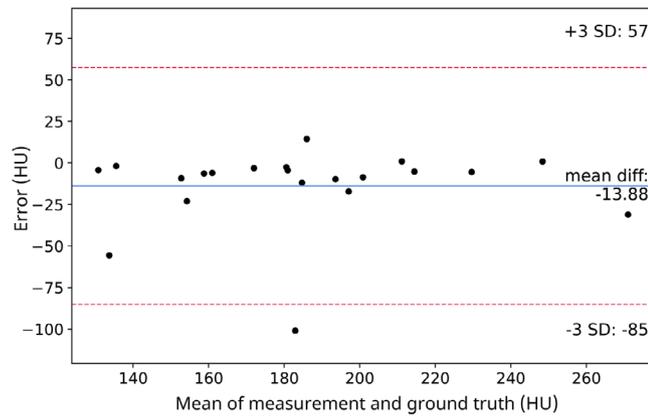
of the random forest trees. Exp\_StdDev and Exp\_Mean were among the most important features, demonstrating that pixels immediately surrounding a segmentation are helpful in classification. The next most important features were: X, Y, Angle, FeretAngle, and Circularity.

### 3.3 | Accuracy of PV attenuation measurement

The automated PV attenuation measurement (RF Model 2) and ground truth across individual cases in the Hold-out Test set are presented in Figure 5. Errors were within 10 HU in 14 out of 16 cases. Across the Hold-out Test set, the average absolute error was 13.4 HU, and the root mean-squared error (RMSE) was 30.1 HU. We found inconsistent vessel measurement across positively classified segmentations (high standard deviation across positive segmentations) in two cases. Nevertheless, the automated measurement was accurate in these two cases, indicating that taking the median across segmentations affords partial robustness against misclassification.

### 3.4 | Independent test set

The performance of the RF classifier in the Independent Test set is summarized in Table 3. The classification precision and sensitivity were 0.68 and 0.18, respectively.



**FIGURE 6** PV measurement accuracy in the Independent Test Set – A Bland-Altman plot of the measurement errors, automated minus ground truth, as a function of varying PV attenuation. PV, portal vein.

Although the performance decreased in comparison to the Hold-out Test set, attenuation measurement did not suffer substantially. Across the Independent Test set, the average absolute error was 15.2 HU, and the RMSE was 27.3 HU. Overall, the model is able to measure PV attenuation within 20 HU in 17 out of 21 cases, and within 10 HU in 15 out of 21 cases. Importantly, such accuracy was found over a wide range of PV attenuation values: 125–250 HU (Figure 6). Notably, cases where attenuation was greater than 200 HU in the Independent Test set were measured accurately, even though the training data did not include such cases.

We found a consistent under-estimation bias in the Independent Test set. The large negative errors in four cases were attributed to FP classifications. The errors were negative in sign because non-PV segmentations have a lower mean pixel value than the PV, on average.

### 3.5 | Application to scan timing assessment

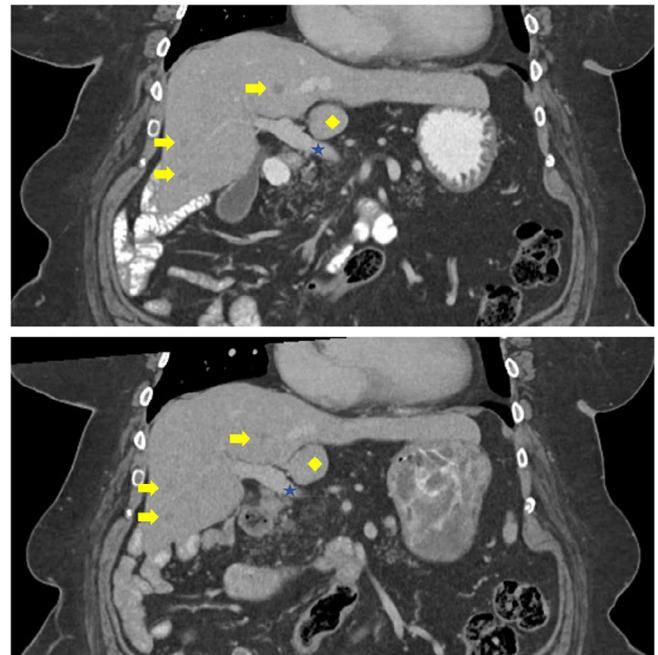
The attenuation of the PV and the IVC, and the ratio of PV-to-IVC for these CT scans are presented Table 4. In these few examples, we see the PV/IVC ratio is low (1.06–1.10) in exams with late timing, moderate (1.23–1.41) in normally timed scans, and high (1.62–2.03) in exams with early timing. PV attenuation is lower in scans with either early or late timing by at least 30 HU, in comparison to the paired control cases with normal timing. Images from these examples are shown in Figures 7, 8, and 9.

Based on these examples, measurements of attenuation in major blood vessels may be used to detect scan mis-timing or contrast mis-administration events, which could be tracked longitudinally as a QC measure in a radiology department. These data could be used, for example, to determine the volume of exams

**TABLE 4** PV and IVC attenuation measurements in selected abdomen CT scans with known scan mistiming.

Patient index	PV (HU)	IVC (HU)	PV/IVC	Timing
1	129	117	1.10	Late
1	151	107	1.41	Normal
2	141	133	1.06	Late
2	188	153	1.23	Normal
3	195	121	1.62	Early
3	232	185	1.26	Normal
4	243	120	2.03	Early

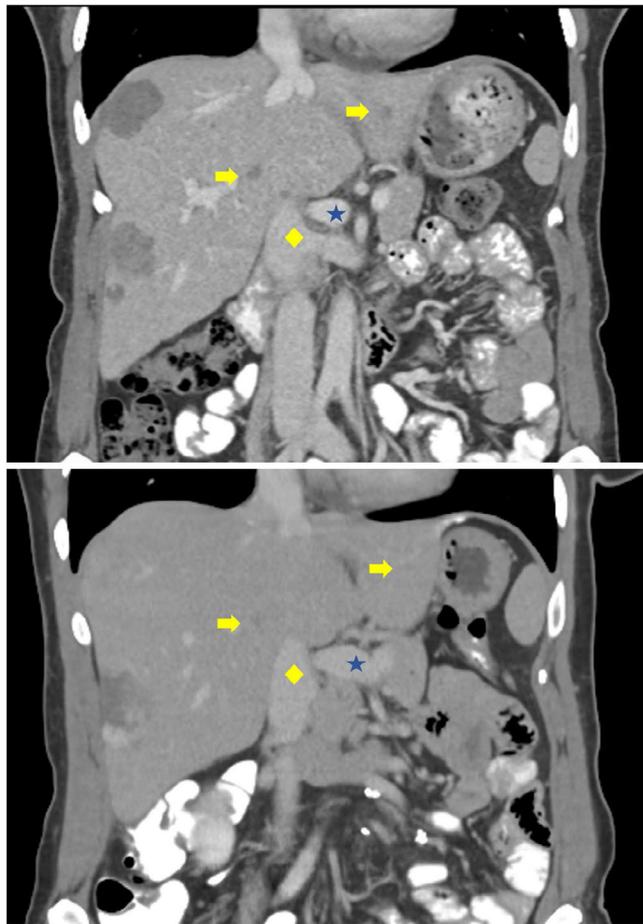
*Note:* Data for a normally timed scan for the same patient is also provided, if available. Scans from the same patient are labeled with identifying patient index. Abbreviations: CT, computed tomography; HU, Hounsfield units; IVC, inferior vena cava; PV, portal vein.



**FIGURE 7** **Top:** Normally timed abdomen CT scan. The PV (blue star) has a moderately higher attenuation (151 HU) than the IVC (yellow diamond - 107 HU). PV/IVC ratio is 1.41. Multiple liver metastases (yellow arrows) are conspicuous. **Bottom:** Abdomen CT of the same patient with *late* scan timing. The PV (blue star) has a approximately the same attenuation (129 HU) as the IVC (yellow diamond - 117 HU). PV/IVC ratio is 1.10. Metastases (yellow arrows) are less conspicuous. Window level and width: 550/50 HU. CT, computed tomography; HU, Hounsfield units; IVC, inferior vena cava; PV, portal vein.

affected by sub-optimal scan timing due to congestive heart failure.<sup>37</sup>

These examples also provide preliminary evidence of generalization of the method to CT exams using different tube voltages other than 120 kV. Automatic PV intensity measurement was approximately as accurate in these examples with tube voltage ranging from 100 to 140 kV as compared to the Independent Test set. The automatic PV measurement accuracy results on these

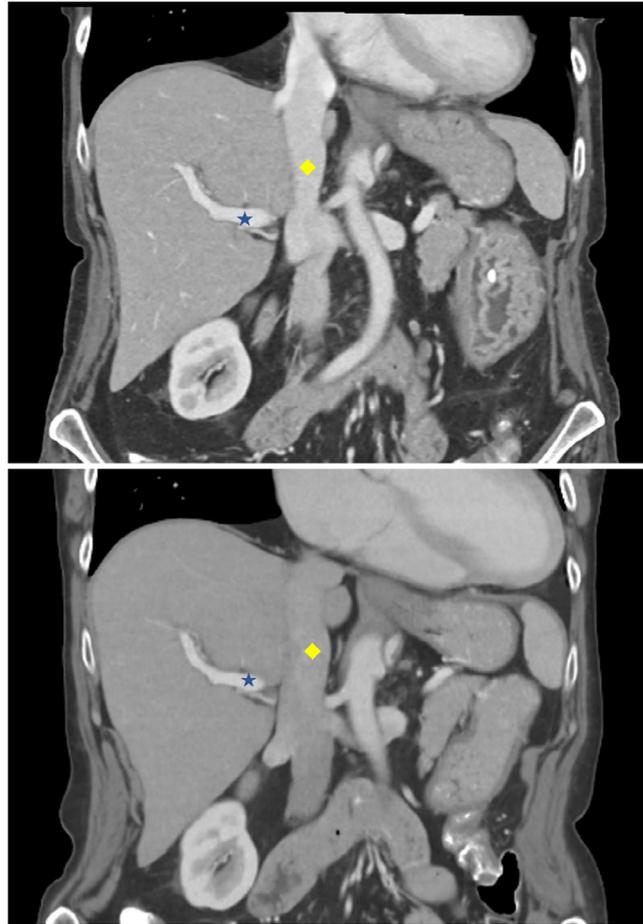


**FIGURE 8** **Top:** Normally timed abdomen CT scan. The PV (blue star) has a moderately higher attenuation (188 HU) than the IVC (yellow diamond - 153 HU). PV/IVC ratio is 1.23. Multiple liver metastases (yellow arrows) are conspicuous. **Bottom:** Abdomen CT of the same patient with *late* scan timing. The PV (blue star) has a approximately the same attenuation (141 HU) as the IVC (yellow diamond - 133 HU). PV/IVC ratio is 1.06. Metastases (yellow arrows) are not visualized. Window level and width: 550/50 HU. CT, computed tomography; HU, Hounsfield unit; IVC, inferior vena cava; PV, portal vein.

examples is provided in Table 5. Though these examples are limited in number, tube voltage does not appear to be an impeding factor to generalization. Tube voltage mainly influences PV attenuation, and the method was accurate over a broad range of PV attenuation encountered in the Hold-out Test and Independent Test Sets. Tube voltage also affects image noise. Further work is needed to validate the method's generalization to varying tube voltage, as well as varying spatial resolution and image noise.

#### 4 | DISCUSSION

We presented a method for automatic pixel value measurement in a blood vessel of interest in CT images. This automated PV intensity measurement may be used in a



**FIGURE 9** **Top:** Normally timed abdomen CT scan. The PV (blue star) has a moderately higher attenuation (232 HU) than the IVC (yellow diamond - 185 HU). PV/IVC ratio is 1.26. **Bottom:** Abdomen CT of the same patient with *early* scan timing. The PV (blue star) has much greater attenuation (195 HU) than the IVC (yellow diamond - 121 HU). PV/IVC ratio is 1.62. Window level and width: 550/50 HU. CT, computed tomography; HU, Hounsfield unit; IVC, inferior vena cava; PV, portal vein.

QC program to track contrast enhancement in key blood vessels. Future work will generalize our method to other major blood vessels, for example, carotid, pulmonary, renal arteries, for application to specialized contrast-enhanced CT exams in these organs. The method would be re-trained to other vessels in a similar fashion as shown here. Since contrast-enhanced CT exams constitute a large portion of all CT exams, we expect our method to have broad application to CT QC.

We emphasize our following key findings: First, we have identified the combination of *vesselness* image processing and Random Forest ML as a tractable method for the problem of (exclusively) segmenting a vessel with complex curvature and widely varying level of contrast enhancement. Second, we have explained the relative importance of different segmentation features. We found that those features related to the vessel location and the pixel values in its immediate external vicinity are especially important. Third, we found good

**TABLE 5** Accuracy of PV attenuation measurement in selected abdomen CT scans with varying tube voltage.

Pat index	Tube voltage (kV)	Auto PV (HU)	Ground truth PV (HU)	Error (HU)
1	130	102	129	-27
1	140	145	151	-6
2	100	141	141	0
2	100	182	188	-6
3	120	191	195	-4
3	100	214	227	-13
4	120	242	243	-1

Note: Automated PV attenuation measurement is labeled "Auto PV". Manual PV measurement is labeled "Ground truth PV".

Abbreviations: HU, Hounsfield unit; PV, portal vein.

overall automatic measurement accuracy ( $< 10$  HU error in 15/21 exams), despite only 79% classifier precision in distinguishing PV segmentations. This is explained by the observation that the method has multiple chances of identifying a PV across multiple slices. Some robustness to misclassification is afforded by aggregation across multiple positively classified segmentations. Lastly, and crucially, we have validated the method's performance in an independent data set.

Classification performance was substantially lower in the Independent Test set compared to the Hold-out Test set, and further method refinement and increased classification accuracy is desired. In the application to assessment of scan timing, an error of 10 HU may affect the assessment, based on the examples above. Therefore, a 95% success rate in measurement of PV intensity with  $< 10$  HU error may be a useful objective. Our method's performance approached but does not yet meet this objective.

We discuss potential refinements to improve method performance. The failure mode seen in the high-error cases related to the RF classifier and low sensitivity in classifying the PV segmentation, even when a candidate PV segmentation was present. The modification in RF Model 2 partially mitigated this failure mode by increasing RF classifier sensitivity at the expense of precision. Local re-training of the RF and customization of the *vesselness* threshold to our center's exam population is expected to improve the model performance. Another failure mode, seen in 4 out of 21 cases in the Independent Test set, was that fewer than four 2D candidate segmentations corresponding to the PV were produced by image processing; that is, the image processing did not segment the PV in all slices. To remedy this failure, future refinement could incorporate multi-scale smoothing in the *vesselness* processing to capture a wider range of PV diameters, and thus increase the sensitivity in segmenting edge-case PV

As noted, DL methods have been very successful in medical image segmentation tasks. Given sufficient and diverse training data, a DL method may outper-

form our method in the present application, but it is unclear how much training data would be needed. This work demonstrated vessel segmentation with a relatively small training data set compared to those typically used to train DL algorithms, and without manually defined contours for supervised learning.

Finally, we note that a future study demonstrating correlation between automated measurement and subjective assessment of enhancement quality would provide definitive evidence of the utility of the proposed method.

## 5 | CONCLUSIONS

A method of automated measurement of PV pixel intensity in contrast-enhanced abdomen CT was demonstrated. The method was accurate over a wide range of PV intensities, and this result held in an independent test dataset. The method does not require expensive manual vessel contouring for supervised training.

## ACKNOWLEDGMENTS

Kevin McCoy is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1842494, and by the Ken Kennedy Institute Computational Science and Engineering Recruiting Fellowship, funded by the Energy HPC Conference. D.T. and S.M. are supported by the CPRIT Research Training Award, funded by the CPRIT Training Program (RP210028). This work was also supported by the institutional CCSG (cancer center support grant) from the NIH/National Cancer Institute under award number P30CA016672.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## REFERENCES

1. Baron RL. Understanding and optimizing use of contrast material for CT of the liver. *AJR Am J Roentgenol*. 1994;163(2):323-331.
2. Jensen CT, Wagner-Bartak NA, Vu LN, et al. Detection of colorectal hepatic metastases is superior at standard radiation dose CT versus reduced dose CT. *Radiology*. 2019;290(2):400-409.
3. Fletcher JG, Yu L, Li Z, et al. Observer performance in the detection and classification of malignant hepatic nodules and masses with CT image-space denoising and iterative reconstruction. *Radiology*. 2015;276(2):465-478.
4. Månsson L. Methods for the evaluation of image quality: a review. *Radiat Prot Dosim*. 2000;90:89-99.
5. Verdun F, Racine D, Ott JG, et al. Image quality in CT: from physical measurements to model observers. *Physica Med*. 2015;31:823-843.
6. *ACR Manual on Contrast Media*. American College of Radiology; 2015. Accessed June 19, 2023. [https://www.acr.org/-/media/ACR/Files/Clinical-Resources/Contrast\\_Media.pdf](https://www.acr.org/-/media/ACR/Files/Clinical-Resources/Contrast_Media.pdf)
7. Bae KT. Intravenous contrast medium administration and scan timing at CT: considerations and approaches. *Radiology*. 2010;256:32-61.
8. Kainz W, Neufeld E, Bolch WE, et al. Advances in computational human phantoms and their applications in biomedical

- engineering – A topical review. *IEEE Trans Radiat Plasma Med Sci.* 2018;3:1-23.
9. Whitehead JF, Laeseke PF, Periyasamy S, Speidel MA, Wagner MG. In silico simulation of hepatic arteries: an open-source algorithm for efficient synthetic data generation. *Med Phys.* 2023;50(9):5505-5517.
  10. Szczykutowicz TP, Viggiano B, Rose S, Pickhardt PJ, Lubner MG. A metric for quantification of iodine contrast enhancement (Q-ICE) in computed tomography. *J Comput Assist Tomogr.* 2021;45:870-876.
  11. Jensen CT, Blair KJ, Wagner-Bartak NA, et al. Comparison of abdominal CT enhancement and organ lesion depiction between weight-based scanner software contrast dosing and a fixed-dose protocol in a tertiary care oncologic center. *J Comput Assist Tomogr.* 2019;43:155-162.
  12. Abadi E, Sanders J, Samei E. Patient-specific quantification of image quality: an automated technique for measuring the distribution of organ hounsfield units in clinical chest CT images. *Med Phys.* 2017;44:4736-4746.
  13. Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms – review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed.* 2018;158:71-91.
  14. Ciecholewski M, Kassjański M. Computational methods for liver vessel segmentation in medical imaging: a review. *Sensors.* 2021;21:2027.
  15. Wu M, Qian Y, Liao X, Wang Q, Heng P-A. Hepatic vessel segmentation based on 3D swin-transformer with inductive biased multi-head self-attention. *BMC Med Imaging.* 2023;23:1-14.
  16. Gao Z, Zong Q, Wang Y, et al. Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation. *IEEE Trans Med Imaging.* 2023;42(10):3059-3068.
  17. Li R, Huang Y-J, Chen H, et al. 3D graph-connectivity constrained network for hepatic vessel segmentation. *IEEE J Biomed Health Inform.* 2021;26:1251-1262.
  18. Breiman L. Random forests. *Mach Learn.* 2001;45:5-32.
  19. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell.* 2020;2:56-67.
  20. Sureka B, Patidar Y, Bansal K, Rajesh S, Agrawal N, Arora A. Portal vein variations in 1000 patients: surgical and radiological importance. *Br J Radiol.* 2015;88:20150326.
  21. Roth H, Farag A, Turkbey EB, Lu L, Liu J, Summers RM. Data from Pancreas-CT (Version 2) [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU>; 2016.
  22. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* 2013;26:1045-1057.
  23. Ahmad M, Jacobsen MC, Thomas MA, Chen HS, Layman RR, Jones AK. A benchmark for automatic noise measurement in clinical computed tomography. *Med Phys.* 2021;48:640-647.
  24. Abramoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. *Biophotonics Intern.* 2004;11:36-42.
  25. Sato Y, Nakajima S, Atsumi H, et al. 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.* Springer; 1997:213-222.
  26. National Institutes of Health, ImageJ Analyze Menu, Accessed June 19, 2023. <https://imagej.nih.gov/ij/docs/menus/analyze.html>; 2018.
  27. Walton W. Feret's statistical diameter as a measure of particle size. *Nature.* 1948;162:329-330.
  28. Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: consistency properties. *Int Stat Rev.* 1989;57:238-247.
  29. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell. Syst.* 1998;13:18-28.
  30. Rosenblatt F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Vol 55. Spartan Books; 1962.
  31. Gallant SI. Perceptron-based learning algorithms. *IEEE Trans Neural Netw.* 1990;1:179-191.
  32. F. Pedregosa, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825-2830.
  33. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM; 2016:785-794.
  34. McCollough C, Bakalyar DM, Bostani M, et al. Use of water equivalent diameter for calculating patient size and size-specific dose estimates (SSDE) in CT: the report of AAPM task group 220. *AAPM report.* 2014;2014:6-23.
  35. Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol.* 1933;24:498-520.
  36. Louppe G, Wehenkel L, Suter A, Geurts P. Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems.* C.J. Burges and L. Bottou and M. Welling and Z. Ghahramani and K.Q. Weinberger. Curran Associates, Inc. 2013;26. [https://proceedings.neurips.cc/paper\\_files/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2013/file/e3796ae838835da0b6f6ea37bcf8bcb7-Paper.pdf)
  37. Jensen CT, Khetan R, Adkins J, et al. Delayed bolus-tracking trigger at CT correlates with cardiac dysfunction and suboptimal portovenous contrast phase. *Abdom Radiol.* 2021;46:826-835.

**How to cite this article:** McCoy K, Marisette S, Tan D, et al. Automatic vessel attenuation measurement for quality control of contrast-enhanced CT: Validation on the portal vein. *Med Phys.* 2024;51:5954–5964. <https://doi.org/10.1002/mp.17267>