Computerized method for evaluating diagnostic image quality of calcified plaque images in cardiac CT: Validation on a physical dynamic cardiac phantom

Martin King,^{a)} Zachary Rodgers, and Maryellen L. Giger^{b)}

Department of Radiology, Committee on Medical Physics, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Dianna M. E. Bardo

Department of Diagnostic Radiology, Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239

Amit R. Patel

Department of Medicine, The University of Chicago, 5841 South Maryland Avenue, MC 5084, Chicago, Illinois 60637

(Received 27 April 2010; revised 6 August 2010; accepted for publication 3 September 2010; published 14 October 2010)

Purpose: In cardiac computed tomography (CT), important clinical indices, such as the coronary calcium score and the percentage of coronary artery stenosis, are often adversely affected by motion artifacts. As a result, the expert observer must decide whether or not to use these indices during image interpretation. Computerized methods potentially can be used to assist in these decisions. In a previous study, an artificial neural network (ANN) regression model provided assessability (image quality) indices of calcified plaque images from the software NCAT phantom that were highly agreeable with those provided by expert observers. The method predicted assessability indices based on computer-extracted features of the plaque. In the current study, the ANN-predicted assessability indices were used to identify calcified plaque images with diagnostic calcium scores (based on mass) from a physical dynamic cardiac phantom. The basic assumption was that better quality images were associated with more accurate calcium scores.

Methods: A 64-channel CT scanner was used to obtain 500 calcified plaque images from a physical dynamic cardiac phantom at different heart rates, cardiac phases, and plaque locations. Two expert observers independently provided separate sets of assessability indices for each of these images. Separate sets of ANN-predicted assessability indices tailored to each observer were then generated within the framework of a bootstrap resampling scheme. For each resampling iteration, the absolute calcium score error between the calcium scores of the motion-contaminated plaque image and its corresponding stationary image served as the ground truth in terms of indicating images with diagnostic calcium scores. The performances of the ANN-predicted and observer-assigned indices in identifying images with diagnostic calcium scores were then evaluated using ROC analysis.

Results: Assessability indices provided by the first observer and the corresponding ANN performed similarly ($AUC_{OBS1}=0.80$ [0.73,0.86] vs $AUC_{ANN1}=0.88$ [0.82,0.92]) as that of the second observer and the corresponding ANN ($AUC_{OBS2}=0.87$ [0.83,0.91] vs $AUC_{ANN2}=0.90$ [0.85,0.94]). Moreover, the ANN-predicted indices were generated in a fraction of the time required to obtain the observer-assigned indices.

Conclusions: ANN-predicted assessability indices performed similar to observer-assigned assessability indices in identifying images with diagnostic calcium scores from the physical dynamic cardiac phantom. The results of this study demonstrate the potential of using computerized methods for identifying images with diagnostic clinical indices in cardiac CT images. © 2010 American Association of Physicists in Medicine. [DOI: [10.1118/1.3495684](https://doi.org/10.1118/1.3495684)]

Key words: cardiac CT, computed tomography, computer-aided-diagnosis, phantom, calcium score

I. INTRODUCTION

Within the past several years, continuing advances in computed tomography (CT) technology have allowed for improved visualization of coronary anatomy and better detection of stenotic lesions.^{1,2} However, even with these technological advances, motion artifacts continue to remain

an important issue in the interpretation of cardiac CT images. Motion artifacts can adversely affect the accurate delineation of important coronary structures. Furthermore, motion artifacts have been shown to increase the variability of coronary calcium scores,³⁻⁶ reduce the diagnostic performance of cardiac CT for the detection of stenotic coronary lesions,^{1,7,8} and

complicate efforts to characterize the composition of coronary plaques.⁹ As a result, physicians must often evaluate the quality of an image in order to determine whether meaningful information can be obtained.

The task of evaluating a motion-contaminated image can be challenging and potentially time-consuming because of its subjective nature. Since different physicians have varying perceptions on how motion artifacts influence the overall image quality, physicians may come to differing conclusions on the diagnostic quality of a given image. Although image quality rating scales (e.g., 1=no motion artifacts; 5=severe motion artifacts) are available for interpreting motion-contaminated images,^{6,10–13} different physicians may continue to arrive at different ratings depending on how they categorize motion artifacts.

Recently, a computerized method for characterizing the image quality of motion-contaminated calcified plaques in noncontrast-enhanced CT scans was developed. This method was developed using simulated calcified plaques from the software dynamic NCAT phantom.^{14–17} In this method, morphological, intensity-based, and dynamic features were extracted from these plaques.¹⁸ These quantitative features were then inputted into an artificial neural network (ANN) regression model,¹⁹ which provided assessability indices on a scale from 1 (excellent image quality) to 5 (very poor image quality) for each individual plaque image. The assessability indices provided by this computerized method were shown to be highly agreeable with those provided by expert physician observers. Assessability indices were also capable of finding optimal phases for interpreting calcified plaque images in an extremely efficient and potentially more consistent manner.²⁰ Furthermore, these findings were replicated in a small pilot study involving a physical dynamic cardiac phantom^{21,22} scanned with a clinical 64 channel CT scanner.²³

However, the overall purpose of a noncontrast-enhanced CT scan is not to characterize the quality of a calcified plaque image but to obtain an accurate calcium score. Since calcified plaque images with more motion artifacts have been associated with less reproducible calcium scores,⁶ images with more accurate calcium scores may be identified by stratifying images based on assessability indices. The inclusion of calcium scores from better quality images may provide physicians with a more meaningful impression of a patient's overall calcified plaque burden.

The purpose of this study is to determine whether ANN-predicted assessability indices perform similar to observer-assigned indices for the task of identifying images with diagnostic calcium scores. A calcium score (CS) is diagnostic if its absolute error, which is defined as the absolute difference between the calcium score of a motion-contaminated image with that of a stationary image, is below a predetermined threshold value. The absolute calcium score error due to the presence of object motion can be accurately quantified under the precisely controlled conditions offered by the phantom, which was previously used in the pilot study.²³ This metric can then be used as a gold standard for evaluat-

FIG. 1. Dynamic cardiac phantom. (a) A view of the entire apparatus immediately surrounding the elastic cylinder, which corresponds to the left ventricle of the heart. The cardiac and respiratory motions of the cylinder are driven by two pistons, which are connected to a pneumatic system (not shown). (b) A close-up view of the elastic cylinder with embedded tubes representing coronary arteries.

ing the performances of ANN-predicted and observer-assigned assessability indices in identifying images with diagnostic calcium scores.

II. MATERIALS AND METHODS

II.A. Dynamic cardiac phantom and calcified plaque model

The physical dynamic cardiac phantom^{21,22} was used to simulate cardiac motion. This device consisted of an elastic cylinder (100 mm in diameter $\times$ 100 mm in length) that was positioned between two polymethyl methacrylate plates, as shown in Fig. 1(a). Two pneumatic pistons controlled the actual motion of the cylinder. One piston drove the elastic cylinder along a curved translational trajectory that was representative of respiratory motion. The other piston provided the elastic cylinder with the contractile and torsional movements that mimicked cardiac motion. This piston effectively compressed the elastic cylinder by driving the movable polymethyl methacrylate plate against the stationary plate. The rates and amplitudes of the phantom's cardiac and respiratory motions were specified by the user on a dedicated computer. The main control unit also provided an artificial ECG signal, which allowed for gated reconstructions.

Hydroxyapatite was used to physically model calcified coronary plaques. Although the hydroxyapatite disks used in this study had a much higher attenuation (~ 1600 HU) than typical coronary calcium, the elastic cylinder of the phantom also had a greater attenuation (400 HU) than that of the human heart (~ 0 HU). Multiple hydroxyapatite disks were crushed into numerous small plaques with random morphologies. A subset of 100 plaques with a mean volume of 17.0 mm^3 and a median volume of 10.0 mm^3 was selected. This size distribution was chosen based on a study reporting the volumes of individual calcified plaques in CT images of human subjects.²⁴

A 2D grid consisting of 100 points (5 rows and 20 columns) was mapped onto the elastic cylinder. Plaques were randomly assigned to a unique grid location and divided into 4 groups of 25. Only one group of plaques was attached onto

the phantom and scanned at a particular time. For each group, the 25 plaques were glued onto 5 separate rubber bands evenly spaced along the longitudinal axis of the phantom, as shown in Fig. 1(b). Because of their elasticity, the rubber bands remained firmly attached to the phantom during cardiac motion. Since their attenuations were similar to that of the phantom, the rubber bands were not visible on the reconstructed CT images.

II.B. Image acquisition

A 64-channel Philips Brilliance CT scanner was used to acquire scans of the dynamic cardiac phantom. All scans were acquired using a retrospectively gated helical acquisition protocol. A retrospective acquisition protocol was implemented in order to obtain calcified plaque images over a wide range of cardiac phases. The following parameters were used: 120 kVp, 238 mA, 0.42 s gantry rotation time, and 0.2 helical pitch. Each group of 25 plaques was scanned once with cardiac motion at a heart rate of either 60, 70, 80, or 90 bpm. A stationary scan (0 bpm) was also obtained for each group. Respiratory motion was not activated during the cardiac motion or stationary scans.

Images of the moving plaques were reconstructed at 50%, 60%, 70%, 80%, and 90% phases of the R-R interval using commercial reconstruction software from the Philips Brilliance CT system. Images were only reconstructed during these phases because these phases corresponded to diastolic phases of lower cardiac motion. Each image was reconstructed with a pixel size of 0.488 mm in the axial plane and 0.4 mm in the z direction. Center points on each of these plaques were then chosen manually. Separate 20 mm $\times$ 20 mm $\times$ 20 mm region-of-interest (ROI) images positioned around these center points were then extracted. A total of 500 ROI images (25 plaques $\times$ 4 scans $\times$ 5 phases) was obtained.

II.C. Ground truths

II.C.1. Assessability indices

Two expert physician observers with multiple years of experience in analyzing cardiac CT images (DB, AP) provided separate sets of assessability indices for all 500 ROI images. As shown in Table I, these ratings were provided on an ordinal scale from 1 (excellent image quality) to 5 (very poor image quality). Each observer made their ratings independently on a computerized workstation (see Fig. 2). This workstation allowed the observer to view the ROI images of the plaques in a random sequence. The observer could visualize the image under different window and level conditions as well as varying magnifications. In addition, the observer was able to extract attenuation values in Hounsfield units (HU) from any voxel within the image. However, no other information was provided. Prior to rating the 500 ROI images, the observer visualized plaque images from the initial pilot study²³ in order to become familiar with evaluating images of motion-contaminated calcified plaques on the dy-

TABLE I. Assessability indices and representative images ($L:400/W:800$).

Image	Assessability Index	Description
1 (Excellent)	Calcium sharply defined. No motion artifacts with well-defined plaque margins.
2 (Good)	Calcium well-defined. Minimal motion artifacts with minor blurring.
3 (Moderate)	Calcium visible. Tail-shaped artifacts of moderate extent.
4 (Poor)	Calcium poorly defined. More severe tail-shaped artifacts.
5 (Very Poor)	Calcium indeterminable.

dynamic cardiac phantom. For purposes of discussion, the sets of assessability indices provided by the two observers are designated as OBS1 and OBS2.

II.C.2. Coronary calcium score

A coronary calcium score was calculated for each plaque in the ROI image. This score was calculated by first thresholding the plaque using an intensity of 530 HU. A threshold value of 530 HU was used instead of the typical value of 130 HU because the intensity of the phantom varied around 400 HU. The intensity of the human heart muscle (myocardium) on a noncontrast-enhanced CT scan, on the other hand, often varies around 0 HU (see Table II).

The volume and mean intensity of this thresholded region were then calculated and multiplied together to obtain a “proportional mass” calcium score. The proportional mass calcium score was directly proportional to the calcium mass

FIG. 2. Workstation interface used by observers to assign assessability indices.

TABLE II. Typical intensity ranges for the physical dynamic cardiac phantom and the human heart. Intensity ranges are provided for the heart muscle (myocardium) and calcified plaques. The lower limits of the intensity ranges for the plaques are defined by the intensity thresholds used to calculate the calcium score.

Image	Dynamic cardiac phantom (HU)	Human heart (HU)
Myocardium	380 to 420	−20 to 20
Calcified plaques	530 to 1600	130 to 1200

score commonly used for quantifying coronary calcium²⁵ and automatically corrected for the effects of partial volume averaging. The actual coronary mass score was not used because the scanner-specific calibration factor required for calculating this entity was not available.

In order to obtain the absolute calcium score error for a given plaque at a specific cardiac phase, the difference between the calcium score from the motion-contaminated image and the calcium score from the stationary image was calculated. The absolute value was then taken in order to obtain the absolute error. The reproducibility of calcium scores between successive scans was not used in this study because this metric was dependent on, not one, but two images of the same plaque.⁶ This reproducibility index may be misleading in the case that one image is heavily motion-contaminated and the second is not, even though both images correspond to the same cardiac phase. Our experiences with simulated data have shown that this situation occasionally arises for plaques undergoing a great deal of motion.

II.D. Computerized method for assigning assessability indices

The computerized method for assigning assessability indices involved three key steps: Image segmentation, feature extraction, and prediction of assessability indices. This method was based on previous studies involving the NCAT and dynamic cardiac phantoms,^{18,20,23} and is outlined in Fig. 3. Brief summaries of the three steps are provided below.

FIG. 3. Flowchart detailing computerized method for acquiring ANN-predicted assessability indices.

II.D.1. Segmentation

The calcified plaques were segmented in a manner similar to that shown in Fig. 4 of Ref. 18. First, the same ROI image was processed independently with threshold, Laplacian, and Sobel filters. The threshold filter was applied with a threshold value of 530 HU to obtain the first binary mask. The Laplacian filter was applied to a smoothed version of the ROI image and then thresholded with a value of −100 to create a second binary mask. The Sobel filter was applied to the original ROI image and then thresholded with a value of 3000 to obtain a third binary mask. Regions of this mask were labeled based on a 26-point connectivity, and fragments with a volume less than 10 pixels or mean intensity less than 150 HU were eliminated. In order to fill in holes within the Sobel-filtered mask, a dilation operator with a cross-shaped structuring element was then applied. A voxel-by-voxel product of this binary mask with the Laplacian-filtered binary mask was then obtained to obtain a fourth mask. A voxel-by-voxel sum of this fourth mask with the first binary mask from the threshold filter was then performed to make a fifth mask. Finally, connected component labeling based on the 26-point connectivity was performed to identify all regions of the mask. Only regions with a mean intensity greater than 150 HU, a volume greater than 15 pixels, and a distance from the manually chosen center of the ROI less than 15 pixels were retained within the segmentation result. All numerical parameters in this algorithm were chosen empirically based on the seven calcified plaques analyzed in the initial pilot study.²³

A segmentation result was obtained for 490 out of 500 (98%) motion-contaminated images and all 100 stationary images. Calcified plaques were barely discernible in the ten images without segmentation results, and these images were automatically assigned assessability indices of 5. Segmentations for two plaques at multiple cardiac phases are shown in Fig. 4.

II.D.2. Feature extraction

Second, the following six phase-correlated features were extracted from each image: 3D velocity, 3D acceleration, edge-based volume, threshold-based volume, sphericity, and standard deviation of intensity. The 3D velocity and 3D acceleration metrics were calculated using calcified plaque positions over multiple cardiac phases. These features were calculated using the same methods described in Sec. 2.3 of Ref. 18. Furthermore, since each feature was sampled over multiple cardiac phases for a given plaque, the mean values of each of the six phase-correlated features for a given plaque were also calculated. Feature values are also included for the two plaques shown in Fig. 4.

Feature selection was not performed for this particular study due to the desire to include the entire plaque database in the training and testing of the ANN regression model. The six features chosen for inclusion in this study were the same six features identified through stepwise regression in the previous NCAT study.¹⁸ Furthermore, these features were used

FIG. 4. Phase-correlated images, segmentations, and feature values of two different plaques at 70 bpm from a single iteration of the bootstrap resampling scheme. The features are 3D velocity (VEL), 3D acceleration (ACC), edge-based volume (VOL-E), threshold-based volume (VOL-T), sphericity (SPHER), and standard deviation of intensity (STD INT).

successfully in the previously conducted pilot study involving the dynamic cardiac phantom.²⁶

II.D.3. ANN regression model

Finally, a linear-output ANN regression model¹⁹ was used to predict assessability indices. This regression model was the same as that used in the previous NCAT and dynamic cardiac phantom pilot study. The model consisted of a three-layer backpropagation network with a sigmoidal activation function in the hidden layer and a linear activation function in the output layer. The network was cast in a 13-10-5 structure. The 13 inputs for a given image included the six feature values, the corresponding six mean feature values unique to a given plaque, and the heart rate. The five binary outputs represented the predicted assessability index in “thermometer” code.²⁷ The ANN regression model was trained over 1000 training epochs using the set of assessability indices provided by a single observer, and thus provided assessability indices reflective of that observer’s inherent biases. ANN1 and ANN2 represent the sets of ANN-predicted assessability indices corresponding to the observer-assigned sets, OBS1 and OBS2, respectively.

The previous NCAT and dynamic cardiac phantom pilot studies also included assessability indices predicted from ordinal least-squares regression models (LR). These methods were not included in this study because the mean performance of the LR methods in assigning agreeable assessability indices never surpassed that of the ANN model in both studies.

FIG. 5. Flowchart detailing the evaluation of ANN-predicted and observer-assigned assessability indices. Note that ROC analysis was performed on each set of assessability indices, separately. CCC represents the repeated measures concordance correlation coefficient. AUC represents the area under the curve.

II.E. Performance evaluation

II.E.1. Performance metrics

Three metrics were used to compare the performance of the ANN regression model to that of the observer. Two of these metrics were comparisons between the ANN-predicted and observer-assigned indices. As shown in Fig. 5, the first metric was the mean difference between the ANN-predicted and the observer-assigned assessability indices.²⁸ The second metric was the repeated measures concordance correlation coefficient (CCC), which analyzed the agreeability of the ANN-predicted and observer-assigned indices.^{29–31} The original concordance correlation coefficient was not used because this metric was not able to take into account the inherent correlations resulting from the multiple phase-correlated images of each plaque within the same dataset.

ROC analysis was used to evaluate, in an independent manner, the performances of the ANN-predicted and the observer-assigned indices in identifying images with diagnostic calcium scores (see Fig. 5). An absolute cut-off value was used to partition images into lower and higher absolute calcium score errors. This value was established as half of the median absolute calcium score error for the entire 500 image database. Images with absolute errors lower than this cutoff were defined as having diagnostic calcium scores. The rationale for using this cut-off value was that 25% of the 500 images in the database would have a diagnostic calcium score if the distribution of absolute calcium score errors were normally distributed. Even if the distribution of the sampled absolute calcium score errors were not normally distributed, only a minority of images with the lowest absolute calcium score errors would be labeled as diagnostic. This cut-off value should be easily generalizable to other datasets since it is dependent on the median absolute calcium score for any given dataset.

An empirical ROC curve was generated by varying the assessability index threshold below which images were placed into the diagnostic calcium score category. The PROPROC program then produced a fitted ROC curve. The area under the curve (AUC) value was used as the final performance metric.^{32,33} AUC values were generated for both

observer-assigned (OBS1 and OBS2) and ANN-predicted (ANN1 and ANN2) assessability indices.

II.E.2. Resampling scheme

In order to generate mean and confidence intervals for the three performance metrics, a 0.632 nonoverlapping block-bootstrap resampling scheme was implemented.^{34–36} This scheme was performed on the entire 500 image database, which consisted of the 100 plaques reconstructed over the 5 cardiac phases. For each of the 1000 iterations, 100 plaques were sampled with replacement. The images associated with the sampled plaques were placed into the training group, and the images associated with the unsampled plaques were allocated into the testing group. The ANN regression model was then parametrized using feature values and observer-assigned assessability indices from the training group. This model then predicted assessability indices for the testing group based on their extracted feature values. The three performance metrics discussed above were calculated. Afterward, mean values of all bootstrapped samples were obtained for the three performance metrics. Confidence intervals at the 95% level were also generated for each metric using the percentile method.^{37,38} The entire resampling scheme was performed twice due to the presence of two sets of observer-assigned indices, OBS1 and OBS2. The training and testing groups for both repetitions of the resampling scheme were the same over all iterations.

II.E.3. Subgroup analysis

The performances of the observer-assigned and ANN-predicted indices in identifying images with diagnostic calcium scores were also evaluated for plaque images corresponding to different heart rates, cardiac phases, and plaque locations. For heart rate, bootstrapped AUC values for OBS1, OBS2, ANN1, and ANN2 assessability indices were calculated for four groups of plaques with heart rates of 60, 70, 80, and 90 bpm. In terms of cardiac phase, AUC values were obtained for two groups of plaque images obtained at 50%–60% phases vs 70%–90% phases. Plaque images obtained at 50%–60% cardiac phases generally corresponded to phases of higher cardiac motion than images obtained at 70%–90% cardiac phases for this physical phantom. For plaque location, AUC values were calculated for two groups of plaques located closer to and further from the movable polymethyl methacrylate plate. The closer group consisted of images of ten plaques obtained from the two rubber bands closer to the moving plate. The further group consisted of images of 15 plaques obtained from the three rubber bands further from the moving plate. In general, plaques closer to the movable plate experienced greater cardiac motion.

III. RESULTS

The median calcium score for the entire 500 image database was 16.8. Images with absolute calcium score errors below 8.4, which was half of the median calcium score, were defined as having diagnostic calcium scores. The frequencies of assessability indices assigned by both observers are shown

FIG. 6. Histogram of observer-assigned assessability indices. The lighter and darker bars represent assessability indices assigned by OBS1 and OBS2, respectively.

in Fig. 6. The average and median assessability indices for OBS1 were 1.9 ± 1.2 and 1, respectively. The average and median assessability indices for OBS2 were 2.3 ± 1.3 and 2, respectively. The mean difference between the two sets of assessability indices was -0.4 ± 0.8 , and the CCC between the two sets was 0.48.

III.A. Performance metrics

For both bootstrap resampling schemes, the average number of images with an absolute calcium score error lower than the threshold value of 8.4 was 140.9. In the first bootstrap resampling scheme, the mean difference between the OBS1 and ANN1 assessability indices was $-0.13 [-0.29, 0.02]$. The CCC was 0.48 [0.35, 0.61]. The AUC_{OBS1} measuring the performance of the first observer's ratings in identifying images with diagnostic calcium scores was 0.80 [0.73, 0.86]. The AUC_{ANN1} was 0.88 [0.82, 0.92]. Although the mean AUC_{ANN1} was greater compared to AUC_{OBS1} , these AUC values were considered similar based on their overlapping bootstrap confidence intervals. For the second bootstrap resampling scheme, the mean difference between the OBS2 and ANN2 assessability indices was $-0.06 [-0.23, 0.10]$. The CCC was 0.59 [0.48, 0.68]. The AUC_{OBS2} of 0.87 [0.83, 0.91] was similar to the AUC_{ANN2} of 0.90 [0.85, 0.94] based on their overlapping confidence intervals.

Table III contains a summary of AUC values for observer-assigned and ANN-predicted assessability indices for the entire database and subgroups based on heart rate, cardiac phase, and plaque location. For each of the individual subgroups, the AUC_{OBS1} and AUC_{ANN1} values were similar to overlapping confidence intervals. However, the mean diagnostic performances of ANN1 for the 70 bpm, 90 bpm, and 50%–60% subgroups were noticeably improved from those of OBS1. The AUC_{OBS2} and AUC_{ANN2} values were also similar for all subgroups. However, the mean diagnostic performances of ANN2 for the 90 bpm and 70%–90% subgroups were noticeably improved from those of OBS2.

III.B. Analysis of assessability indices for two phase-correlated image sets

Figure 7 shows two sets of phase-correlated images of different calcified plaques with observer-assigned and ANN-

TABLE III. Performances of observer-assigned and ANN-predicted indices in identifying plaque images with diagnostic calcium scores. Performances for subgroups based on heart rate, cardiac phase, and plaque location are also included. Mean number of diagnostic images represents the mean number of images with diagnostic calcium scores selected over all bootstrap iterations.

Category	Subcategory	Total number of images	Mean number of diagnostic images	OBS1 AUC	ANN1 AUC	OBS2 AUC	ANN AUC
Complete database	—	500	123	0.80 [0.73, 0.86]	0.88 [0.82, 0.92]	0.87 [0.83, 0.91]	0.90 [0.85, 0.94]
Heart rate	60 bpm	125	39	0.86 [0.72, 0.95]	0.89 [0.78, 0.97]	0.90 [0.82, 0.97]	0.90 [0.82, 0.96]
	70 bpm	125	33	0.80 [0.67, 0.92]	0.92 [0.84, 0.98]	0.90 [0.83, 0.96]	0.93 [0.87, 0.98]
	80 bpm	125	41	0.85 [0.70, 0.97]	0.91 [0.78, 1.00]	0.90 [0.83, 0.99]	0.94 [0.87, 1.00]
	90 bpm	125	32	0.72 [0.56, 0.89]	0.82 [0.65, 1.00]	0.79 [0.65, 0.91]	0.88 [0.73, 0.97]
Cardiac phase	50%–60%	200	42	0.60 [0.50, 0.70]	0.71 [0.60, 0.81]	0.70 [0.63, 0.77]	0.72 [0.62, 0.81]
	70%–90%	300	105	0.81 [0.58, 0.98]	0.85 [0.60, 1.00]	0.86 [0.71, 0.96]	0.95 [0.84, 1.00]
Plaque location	Closer	200	36	0.80 [0.70, 0.89]	0.89 [0.81, 0.95]	0.86 [0.80, 0.92]	0.90 [0.85, 0.95]
	Farther	300	90	0.80 [0.71, 0.89]	0.88 [0.79, 0.94]	0.88 [0.82, 0.94]	0.90 [0.83, 0.95]

predicted assessability indices from the same bootstrap iteration. The plaque shown in panel (a) was located closer to the moving plate and was scanned at 70 bpm. For this plaque, four out of five images had a diagnostic coronary calcium score (based on an absolute calcium score error <8.4). Assuming that images with assessability indices less than or equal to 2 had diagnostic image quality, ANN1 and ANN2 both were able to identify three out of four images with diagnostic coronary calcium scores. On the other hand, OBS1 and OBS2 were only able to identify one out of four and two out of four images with diagnostic scores, respectively. The one image with the nondiagnostic coronary calcium score was correctly identified by both observer-assigned and ANN-predicted indices. For the plaque in panel (b), which was located further from the moving plate and was scanned at 70 bpm, three out of five images had a diagnostic coronary calcium score. The ANN1 and ANN2 indices both were able to identify all three images with diagnostic calcium scores, as well as the two images with nondiagnostic

calcium scores. However, OBS1 and OBS2 each were only able to correctly identify two out of three images with diagnostic calcium scores. Furthermore, OBS1 incorrectly identified one out of two images with nondiagnostic calcium scores as diagnostic.

Figure 8 consists of bubble plots depicting the agreeability of the OBS1-ANN1 and OBS2-ANN2 assessability pairings for all plaques analyzed in this bootstrap iteration. For both pairings, most plaque images were located on or within one unit from the diagonal of this plot. The CCCs for the OBS1-ANN1 as well as OBS2-ANN2 pairings were 0.52 and 0.61, respectively. An ROC curve depicting the performances of the observer-assigned and ANN-predicted indices in identifying images with diagnostic calcium scores is presented in Fig. 8(c). AUC values for OBS1, ANN1, OBS2, and ANN2 were 0.81, 0.92, 0.87, and 0.88, respectively. An extremely marked improvement in AUC was noted with the ANN1 assessability indices compared to the OBS1 indices. Although the difference in AUC between the ANN2 and OBS2 indices was much less prominent, a smaller improvement was expected due to the higher baseline AUC for the OBS2 indices.

(a)					
Phase	50%	60%	70%	80%	90%
Obs1	5	5	4	1	4
ANN1	4	4	2	2	2
OBS2	5	5	2	3	1
ANN2	5	5	2	1	2
Abs CS Error	286.3	2.0	1.8	0.1	0.2
(b)					
Phase	50%	60%	70%	80%	90%
Obs1	3	2	3	1	2
ANN1	4	4	2	1	1
OBS2	5	3	3	1	1
ANN2	5	4	2	2	2
Abs CS Error	35.6	60.7	3.0	0.4	0.1

FIG. 7. Phase-correlated images and assessability indices of the same two plaques shown in Fig. 4. Observer-assigned assessability indices (OBS1 and OBS2), corresponding ANN-predicted assessability indices (ANN1 and ANN2), and absolute calcium score (CS) errors are listed within each panel. Assessability indices indicative of better image quality (1 and 2) and lower absolute CS errors (less than the cut-off value of 8.4) are shaded.

IV. DISCUSSION

The results of this study demonstrate that automated methods for evaluating cardiac CT image quality have the potential for providing information of clinical significance. In particular, this study showed that ANN-predicted assessability indices performed similar to observer-assigned indices for the task of identifying images with diagnostic coronary calcium scores. Their similar performances persisted for ratings provided by two different observers and under different conditions of heart rate, cardiac phase, and plaque location. Furthermore, despite the overlapping confidence intervals, the noticeable disparity between the mean AUC_{OBS1} of 0.80 and the mean AUC_{OBS2} of 0.87 was effectively reduced after the application of the ANN model. The mean AUC_{ANN1} and mean AUC_{ANN2} were 0.88 and 0.90, respectively.

The noticeable improvement in mean AUC_{ANN1} as compared to mean AUC_{OBS1} may reflect the greater intraobserver variability in the indices provided by the first observer. In

FIG. 8. Performance characteristics of the observer-assigned and ANN-predicted assessability indices for the same iteration of the bootstrap resampling scheme corresponding to the plaques in Fig. 7. (a) Bubble plot of OBS1-ANN1 assessability indices. Circle size correlates with the number of images associated with a given pairing of assessability indices. The CCC was 0.52. (b) Bubble plot of OBS2-ANN2 assessability indices. The CCC was 0.61. (c) ROC curves of assessability indices provided by OBS1 (light dashed), ANN1 (light solid), OBS2 (dark dashed), and ANN2 (dark solid) for the task of identifying diagnostic quality images. Corresponding AUC values were 0.81, 0.92, 0.87, and 0.88 for this iteration.

other words, some of the inherent inconsistencies in this observer's assigned ratings may have been averaged out by the artificial neural network. For example, in Fig. 7(a), OBS1 assigned assessability ratings of 4, 1, and 4, for the plaque at

70%, 80%, and 90% phases, respectively. However, ANN1 assigned assessability indices of 2 to all three images, and all three images had diagnostic calcium scores. Furthermore, despite the overlapping confidence intervals, the mean agreeability for the OBS1-ANN1 indices (CCC of 0.48) was lower than the mean agreeability for the OBS2-ANN2 indices (CCC of 0.59). This lower mean agreeability of the OBS1-ANN1 indices may have been necessary to generate the improvement in the mean AUC_{ANN1} . Thus, the results of this study provide additional validation to the previously described automated method for evaluating the image quality of calcified plaques based on computer-extracted features and artificial neural networks.^{18,20,23}

The physical dynamic cardiac phantom provided an excellent model system for conducting this study. First, it allowed for the calculation of a gold standard metric, the absolute error in coronary calcium score, which was used to define images with diagnostic coronary calcium scores. Since this metric was associated with a single image, it allowed for a straightforward evaluation of the diagnostic performance of assessability indices. Second, the phantom allowed for the diagnostic evaluation of assessability indices to occur across precisely controlled conditions of heart rate, cardiac phase, and plaque location while maintaining the use of a clinical cardiac CT scanner. Of note, this dynamic cardiac phantom has been used previously for evaluating temporal and spatial resolution, as well as assessing coronary calcium score accuracy.^{22,39,40}

However, the use of a physical phantom does introduce inherent limitations to the applicability of this study. The images of the cardiac phantom do not represent the true anatomy of the beating heart in human subjects. Besides being a gross simplification of the actual four-chambered heart, the cylindrical phantom had Hounsfield units that were much greater than that of the usual human heart. Furthermore, the calcified plaques, which were not placed along the typical coronary distribution of the heart, were exposed to air instead of the hilar and mediastinal structures surrounding the human heart.

Since both expert observers were not accustomed to analyzing calcified plaques in such an artificial and simplified context, the expert observers had to visualize calcified plaque images from the initial pilot study in order to learn the spectrum of motion-contaminated and motion-free images. Whether or not the assessability indices assigned to plaques in this study would have mirrored indices assigned to plaques from human subjects under similar motion-contaminated conditions remains in question. Although additional expert observers may have allowed us to perform a more robust comparison of ANN-assigned and observer-assigned indices in the identification of images with diagnostic calcium scores, two observers were sufficient in showing feasibility for this type of phantom study.

An additional limitation to this study is that the gold standard of absolute calcium score error cannot be calculated for actual human subjects. As a result, a direct extrapolation of the numerical results of this paper to related studies from human data cannot be made. Although the reproducibility

metric comparing coronary calcium scores from successive scans⁶ would have been better suited for this specific purpose, it would not have been the best gold standard for designating an image with a diagnostic calcium score due to its dependence on two motion-contaminated images.

Despite these limitations, however, this study demonstrates that images with diagnostic calcium scores can be identified by using a computerized method for evaluating image quality based on ratings provided by expert observers. By incorporating the inherent thought processes of expert observers in the designation of images with diagnostic calcium scores, the method proposed in this study is fundamentally unique. Previous phantom studies have attempted to characterize the accuracy of calcium scoring based on pertinent image acquisition parameters (i.e., heart rate, phase interval, scanning trajectory, and imaging modality).^{6,39,41} Another phantom study demonstrated a method for correcting coronary calcium scores based on calcium density, temporal resolution, and estimated linear coronary velocity of plaques.⁴²

Although the proposed method in this study does not attempt to correct for inaccurate calcium scores, it does attempt to designate images with accurate calcium scores based on emulating the impressions provided by expert observers. It does not make this determination solely based on image acquisition parameters and plaque intensity, given that the ability to predict image quality of calcified plaques from these parameters in the context of complex plaque morphologies and circuitous motion trajectories remains unknown. Assessability indices, on the other hand, have been shown to be able to predict image quality of calcified plaques adequately on the anatomically based dynamic NCAT phantom.²⁰ Thus, if the ANN regression model is trained on a large database of calcified plaque images from patient scans using the reproducibility metric as a gold standard, it potentially can be applied to patients with diverse coronary anatomies, plaque morphologies, and plaque motion trajectories.

In addition, the scope of this study can be extended to contrast-enhanced coronary angiography. The proposed computerized method, for example, could provide assessability indices that could be translated into a metric of confidence for the percent stenosis of a stenotic lesion. This type of analysis could particularly be useful for patients with fast heart rates or high coronary calcium burden, as the impressive sensitivities and specificities for detecting significant stenotic lesions in coronary vessels have been shown to decline for patients with these conditions on single-source 64-slice CT.⁸

Moreover, this type of analysis potentially could be used to aid physicians in the identification of unstable coronary plaques. One study demonstrated that patients with plaques exhibiting both low attenuation values and positive vessel remodeling have a markedly increased risk of developing acute coronary syndrome in subsequent years.⁴³ Our experience indicates that even with state-of-the-art CT scanners, these plaques remain difficult to detect and evaluate due to image quality factors such as object motion, inadequate contrast administration, and noise. The proposed computerized

method potentially could provide assessability indices that could also act as a metric of confidence for the likelihood that a potential lesion is indeed anatomically unstable.

ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health Medical Scientist Training Program Grant, National Institutes of Health Grant Nos. EB00225 and EB02765, as well as the Lawrence H. Lanzl Graduate Student Fellowship in Medical Physics (Committee on Medical Physics, The University of Chicago). The authors would like to thank Dr. Michael Vannier for his tremendous help in acquiring the cardiac phantom, Philips Medical Systems for loaning the cardiac phantom, Arkadiusz Wdowiak for the aid that he provided in scanning the phantom, Lorenzo Pesce for helpful discussions regarding the statistical analysis needed for this project, Kenji Suzuki for providing the artificial neural network model, and Li Lan for designing the workstation used by the expert physician observers.

^a)Electronic mail: mtking@uchicago.edu

^b)Author to whom correspondence should be addressed. Electronic mail: m-giger@uchicago.edu; Telephone: (773) 702-6778; Fax: (773) 702-0371.

¹A. W. Leber, A. Knez, F. von Ziegler, A. Becker, K. Nikolaou, S. Paul, B. Winter-sperger, M. Reiser, C. R. Becker, G. Steinbeck, and P. Boekstegers, "Quantification of obstructive and nonobstructive coronary lesions by 64-slice computed tomography: A comparative study with quantitative coronary angiography and intravascular ultrasound," *J. Am. Coll. Cardiol.* **46**, 147–154 (2005).

²T. R. Johnson, K. Nikolaou, B. J. Wintersperger, A. W. Leber, F. von Ziegler, C. Rist, S. Buhmann, A. Knez, M. F. Reiser, and C. R. Becker, "Dual-source CT cardiac imaging: Initial experience," *Eur. Radiol.* **16**, 1409–1415 (2006).

³S. Mao, M. J. Budoff, H. Bakhsheshi, and S. C. Liu, "Improved reproducibility of coronary artery calcium scoring by electron beam tomography with a new electrocardiographic trigger method," *Invest. Radiol.* **36**, 363–367 (2001).

⁴J. Horiguchi, T. Nakanishi, A. Tamura, and K. Ito, "Coronary artery calcium scoring using multicardiac computed tomography," *J. Comput. Assist. Tomogr.* **26**, 880–885 (2002).

⁵R. C. Detrano, M. Anderson, J. Nelson, N. Wong, J. J. Carr, M. McNitt-Gray, and D. E. Bild, "Coronary calcium measurements: Effect of CT scanner type and calcium measure on rescan reproducibility—MESA study," *Radiology* **236**, 477–484 (2005).

⁶J. Horiguchi, H. Fukuda, H. Yamamoto, N. Hirai, F. Alam, H. Kakizawa, M. Hieda, T. Tachikake, K. Marukawa, and K. Ito, "The impact of motion artifacts on the reproducibility of repeated coronary artery calcium measurements," *Eur. Radiol.* **17**, 81–86 (2007).

⁷D. Ropers, U. Baum, K. Pohle, K. Anders, S. Ulzheimer, B. Ohnesorge, C. Schlundt, W. Bautz, W. Daniel, and S. Achenbach, "Detection of coronary artery stenoses with thin-slice multi-detector row spiral computed tomography and multiplanar reconstruction," *Circulation* **107**, 664–666 (2003).

⁸G. L. Raff, M. J. Gallagher, W. W. O'Neill, and J. A. Goldstein, "Diagnostic accuracy of noninvasive coronary angiography using 64-slice spiral computed tomography," *J. Am. Coll. Cardiol.* **46**, 552–557 (2005).

⁹A. W. Leber, A. Knez, A. Becker, C. Becker, F. von Ziegler, K. Nikolaou, C. Rist, M. Reiser, C. White, G. Steinbeck, and P. Boekstegers, "Accuracy of multidetector spiral computed tomography in identifying and differentiating the composition of coronary atherosclerotic plaques: A comparative study with intracoronary ultrasound," *J. Am. Coll. Cardiol.* **43**, 1241–1247 (2004).

¹⁰X. Hamoir, T. Flohr, V. Hamoir, L. Labaki, J. Tricquet, A. Duhamel, and J. Kirsch, "Coronary arteries: Assessment of image quality and optimal reconstruction window in retrospective ECG-gated multislice CT at 375-ms gantry rotation time," *Eur. Radiol.* **15**, 296–304 (2005).

- ¹¹B. J. Wintersperger, K. Nikolaou, F. von Ziegler, T. Johnson, C. Rist, A. Leber, T. Flohr, A. Knez, M. F. Reiser, and C. R. Becker, "Image quality, motion artifacts, and reconstruction timing of 64-slice coronary computed tomography angiography with 0.33-second rotation speed," *Invest. Radiol.* **41**, 436–442 (2006).
- ¹²C. Herzog, M. Arning-Erb, S. Zangos, K. Eichler, R. Hammerstingl, S. Dogan, H. Ackermann, and T. J. Vogl, "Multi-detector row CT coronary angiography: Influence of reconstruction technique and heart rate on image quality," *Radiology* **238**, 75–86 (2006).
- ¹³S. Leschka, L. Husmann, L. M. Desbiolles, O. Gaemperli, T. Schepis, P. Koepfli, T. Boehm, B. Marincek, P. A. Kaufmann, and H. Alkadhi, "Optimal image reconstruction intervals for non-invasive coronary angiography with 64-slice CT," *Eur. Radiol.* **16**, 1964–1972 (2006).
- ¹⁴W. P. Segars, D. S. Lalush, and B. M. W. Tsui, "A realistic spline-based dynamic heart phantom," *IEEE Trans. Nucl. Sci.* **46**, 503–506 (1999).
- ¹⁵W. P. Segars, "Development of a new dynamic NURBS-based cardiac-torso (NCAT) phantom," Ph.D. thesis, The University of North Carolina, 2001.
- ¹⁶J. M. Garrity, W. P. Segars, S. B. Knisley, and B. M. W. Tsui, "Development of a dynamic model for the lung lobes and airway tree in the NCAT phantom," *IEEE Trans. Nucl. Sci.* **50**, 378–383 (2003).
- ¹⁷W. P. Segars, K. Taguchi, G. S. K. Fung, E. K. Fishman, and B. M. W. Tsui, "Effect of heart rate on CT angiography using the enhanced cardiac model of the 4D NCAT," *Proc. SPIE* **6142**, 61420I (2006).
- ¹⁸M. King, M. Giger, K. Suzuki, and X. Pan, "Feature-based characterization of motion-contaminated calcified plaques in cardiac multidetector CT," *Med. Phys.* **34**, 4860–4875 (2007).
- ¹⁹K. Suzuki, I. Horiba, and N. Sugie, "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1582–1596 (2003).
- ²⁰M. King, M. Giger, K. Suzuki, D. Bardo, B. Greenberg, L. Lan, and X. Pan, "Computerized assessment of motion-contaminated calcified plaques in cardiac multidetector CT," *Med. Phys.* **34**, 4876–4889 (2007).
- ²¹H. Timinger, S. Krueger, J. Borgert, and R. Grewer, "Motion compensation for interventional navigation on 3D static roadmaps based on an affine model and gating," *Phys. Med. Biol.* **49**, 719–732 (2004).
- ²²P. G. C. Begemann, U. van Stevendaal, R. Manzke, A. Stork, F. Weiss, C. Nolte-Ernsting, M. Grass, and G. Adam, "Evaluation of spatial and temporal resolution for ECG-gated 16-row multidetector CT using a dynamic cardiac phantom," *Eur. Radiol.* **15**, 1015–1026 (2005).
- ²³Z. B. Rodgers, M. King, M. L. Giger, M. Vannier, D. M. E. Bardo, K. Suzuki, and L. Lan, "Computerized assessment of coronary calcified plaques in CT images of a dynamic cardiac phantom," *Proc. SPIE* **6915**, 69150M (2008).
- ²⁴T. Q. Callister, B. Cooil, S. Raya, N. J. Lippolis, D. J. Russo, and P. Raggi, "Coronary artery disease: Improved reproducibility of calcium scoring with an electron-beam CT volumetric method," *Radiology* **208**, 807–814 (1998).
- ²⁵C. H. McCollough, S. Ulzheimer, S. S. Halliburton, K. Shanneik, R. D. White, and W. A. Kalender, "Coronary artery calcium: A multi-institutional, multimanufacturer international standard for quantification at cardiac CT," *Radiology* **243**, 527–538 (2007).
- ²⁶M. King, D. Xia, X. Pan, M. Vannier, T. Koehler, P. L. Riviere, E. Sidky, and M. Giger, "Chord-based image reconstruction from clinical projection data," *Proc. SPIE* **6193**, 61932G (2008).
- ²⁷M. Smith, *Neural Networks for Statistical Modeling* (Van Nostrand Reinhold, New York, 1993).
- ²⁸J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **1**, 307–310 (1986).
- ²⁹T. S. King, V. M. Chinchilli, and J. L. Carrasco, "A repeated measures concordance correlation coefficient," *Stat. Med.* **26**, 3095–3113 (2007).
- ³⁰L. I. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics* **45**, 255–268 (1989).
- ³¹T. S. King and V. M. Chinchilli, "A generalized concordance correlation coefficient for continuous and categorical data," *Stat. Med.* **20**, 2131–2147 (2001).
- ³²C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* **21**, 720–733 (1986).
- ³³C. E. Metz and X. Pan, "proper" binormal ROC curves: Theory and maximum-likelihood estimation," *J. Math. Psychol.* **43**, 1–33 (1999).
- ³⁴E. Carlstein, "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *Ann. Stat.* **14**, 1171–1179 (1986).
- ³⁵P. Hall, J. L. Horowitz, and B.-Y. Jing, "On blocking rules for the bootstrap with dependent data," *Biometrika* **82**, 561–574 (1995).
- ³⁶T. DiCiccio and B. Efron, "Bootstrap confidence intervals," *Stat. Sci.* **11**, 189–228 (1996).
- ³⁷B. Efron, "Nonparametric standard errors and confidence intervals," *Can. J. Stat.* **9**, 139–158 (1981).
- ³⁸J. Carpenter and J. Bithell, "Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians," *Stat. Med.* **19**, 1141–1164 (2000).
- ³⁹P. G. C. Begemann, U. van Stevendaal, R. Koester, A. H. Mahnken, A. Kooops, G. Adam, M. Grass, and C. Nolte-Ernsting, "Evaluation of the influence of acquisition and reconstruction parameters for 16-row multidetector CT on coronary calcium scoring using a stationary and dynamic cardiac phantom," *Eur. Radiol.* **17**, 1985–1994 (2007).
- ⁴⁰D. T. Boll, E. M. Merkle, E. K. Paulson, R. A. Mirza, and T. R. Fleiter, "Calcified vascular plaque specimens: Assessment with cardiac dual-energy multidetector CT in anthropomorphically moving heart phantom," *Radiology* **249**, 119–126 (2008).
- ⁴¹J. M. Groen, M. J. W. Greuter, R. Vliegthart, C. Suess, B. Schmidt, F. Zijlstra, and M. Oudkerk, "Calcium scoring using 64-slice MDCT, dual source CT and EBT: A comparative phantom study," *Int. J. Cardiovasc. Imaging* **24**, 547–556 (2008).
- ⁴²M. J. W. Greuter, J. M. Groen, L. J. Nicolai, H. Dijkstra, and M. Oudkerk, "A model for quantitative correction of coronary calcium scores on multidetector, dual source, and electron beam computed tomography for influences of linear motion, calcification density, and temporal resolution: A cardiac phantom study," *Med. Phys.* **36**, 5079–5088 (2009).
- ⁴³S. Motoyama, M. Sarai, H. Harigaya, H. Anno, K. Inoue, T. Hara, H. Naruse, J. Ishii, H. Hishida, N. D. Wong, R. Virmani, T. Kondo, Y. Ozaki, and J. Narula, "Computed tomographic angiography characteristics of atherosclerotic plaques subsequently resulting in acute coronary syndrome," *J. Am. Coll. Cardiol.* **54**, 49–57 (2009).