

Computerized assessment of motion-contaminated calcified plaques in cardiac multidetector CT

Martin King,^{a)} Maryellen L. Giger, Kenji Suzuki, Dianna M. E. Bardo, Brent Greenberg, Li Lan, and Xiaochuan Pan

Department of Radiology, Committee on Medical Physics, The University of Chicago, Chicago, Illinois 60637

(Received 10 May 2007; revised 17 September 2007; accepted for publication 9 October 2007; published 28 November 2007)

An automated method for evaluating the image quality of calcified plaques with respect to motion artifacts in noncontrast-enhanced cardiac computed tomography (CT) images is introduced. This method involves using linear regression (LR) and artificial neural network (ANN) regression models for predicting two *patient-specific, region-of-interest-specific, reconstruction-specific* and *temporal phase-specific* image quality indices. The first is a plaque motion index, which is derived from the actual trajectory of the calcified plaque and is represented on a continuous scale. The second is an assessability index, which reflects the degree to which a calcified plaque is affected by motion artifacts, and is represented on an ordinal five-point scale. Two sets of assessability indices were provided independently by two radiologists experienced in evaluating cardiac CT images. Inputs for the regression models were selected from 12 features characterizing the dynamic, morphological, and intensity-based properties of the calcified plaques. Whereas LR-velocity (LR-V) used only a single feature (three-dimensional velocity), the LR-multiple (LR-M) and ANN regression models used the same subset of these 12 features selected through stepwise regression. The regression models were parameterized and evaluated using a database of simulated calcified plaque images from the dynamic NCAT phantom involving nine heart rate/multi-sector gating combinations and 40 cardiac phases covering two cardiac cycles. Six calcified plaques were used for the plaque motion indices and three calcified plaques were used for both sets of assessability indices. In one configuration, images from the second cardiac cycle were used for feature selection and regression model parameterization, whereas images from the first cardiac cycle were used for testing. With this configuration, repeated measures concordance correlation coefficients (CCCs) and associated 95% confidence intervals for the LR-V, LR-M, and ANN were 0.817 [0.785, 0.848], 0.894 [0.869, 0.916], and 0.917 [0.892, 0.936] for the plaque motion indices. For the two sets of assessability indices, CCC values for the ANN model were 0.843 [0.791, 0.877] and 0.793 [0.747, 0.828]. These two CCC values were statistically greater than the CCC value of 0.689 [0.648, 0.727], which was obtained by comparing the two sets of assessability indices with each other. These preliminary results suggest that the variabilities of assessability indices provided by regression models can lie within the variabilities of the indices assigned by independent observers. Thus, the potential exists for using regression models and assessability indices for determining optimal phases for cardiac CT image interpretation. © 2007 American Association of Physicists in Medicine.

[DOI: [10.1118/1.2804718](https://doi.org/10.1118/1.2804718)]

Key words: cardiac CT, computed tomography, regression models, image analysis, computer-aided diagnosis, assessability, calcium score

I. INTRODUCTION

In recent years, cardiac multidetector computed tomography (CT) has been promoted as a next generation modality for diagnosing patients with coronary artery disease in an efficient and noninvasive manner. However, even with the recent advancements in technology such as the introduction of 64-slice and dual-source scanners,^{1,2} cardiac CT images are often plagued by motion artifacts. Artifacts due to cardiac and respiratory motion may degrade image quality such that the sharp delineation of the coronary arteries is lost. As a result, physicians are responsible for not only detecting a lesion, but also determining whether the quality of an image is sufficient for evaluating coronary structures in light of mo-

tion artifacts. The importance of accurately evaluating the quality of images with respect to motion artifacts is highlighted by studies, which have shown that motion artifacts can increase the variability of coronary artery calcium scores³⁻⁶ and reduce the diagnostic performance of cardiac CT for the detection of stenotic coronary lesions.⁷⁻⁹

The primary task of evaluating image quality with respect to motion artifacts is especially challenging. First, this task is subjective, and dependent on a physician's perception of whether the motion artifacts affecting the coronary structure would hinder his or her ability to arrive at an accurate clinical diagnosis. Since perceptions vary among different physicians and even within a single physician over time, inter- and

intra-observer variability may become issues. Second, this task is potentially time consuming. This is due in part to the task's subjective nature, and also to the secondary task of analyzing images of the same coronary structure at multiple cardiac phases if the initial image is deemed to have insufficient image quality. In other words, the primary task often leads to the secondary task of finding optimal phases for image interpretation.

Numerous studies have been performed that provide guidelines for rating the quality of images with respect to the types of motion artifacts affecting a coronary structure.^{10–13} These studies commonly propose image quality rating scales, which in some cases span from 1 (no motion artifacts) to 5 (severe motion artifacts). In one study on rating images of calcified plaques in noncontrast-enhanced CT images, plaques with tail-shaped artifacts are given a rating of 3, whereas plaques with star-shaped artifacts are given a rating of 4.⁶ These studies are useful in that standards for rating the quality of images are explicitly stated in terms of the types of motion artifacts that are present.

However, since the characterizations of motion artifacts are largely descriptive, physicians may have different perceptions on how these motion artifacts affect the overall quality of a given image. Thus, if the motion artifacts affecting a plaque contain characteristics from two ratings, such as a plaque with one long tail (rating=3) but smaller “tentacles” that may qualify it as star-shaped (rating=4), it is entirely possible that two physicians may arrive at different ratings. As a result, image quality rating scales still retain an element of subjectivity. Although “good” inter-observer agreeability may be obtained for physicians who are experienced in reading cardiac CT images,^{13,14} it may suffer for those who are not as experienced in characterizing motion artifacts.

With respect to the secondary task of finding optimal phases for image interpretation, many studies publish optimal phases for analyzing coronary structures based on image quality rating schemes applied to a large database of patient images.^{11–13} Although the optimal phases suggested by these studies are useful for the majority of the patient population, they may find limited applicability for critically ill patients with high and/or arrhythmic heart rates or those unable to maintain adequate breath holds. For these cases, physicians must often undertake the time-consuming task of analyzing images at multiple phases in order to find images of acceptable quality.

Methods for selecting optimal phases based on computer-generated indices reflecting states of high or low cardiac motion also have been developed. These methods are useful in that they can provide suggested phases that are specific for a given patient and scan, and may be used instead of electrocardiogram data for phase-correlated image reconstruction.^{15–17} However, since these methods are based on obtaining global indices reflective of the amount of motion contamination of an entire image volume or across an entire transverse slice, they may not reflect the true motion characteristics of specific coronary segments.

Methods which select cardiac phases based on the velocities of well-known coronary land-marks also have been developed.^{18,19} These methods are advantageous in that suggested phases are selected based on the dynamic characteristics (lowest mean velocities) of specific coronary land-marks within a given patient. However, the manner in which motion artifacts manifest themselves is not solely dependent on mean velocity. Instead, motion artifacts are dependent on a myriad of factors including the relations of the dynamic trajectories of a moving object to the gantry angles used for reconstruction. In fact, our studies have shown that, in some cases, images of calcified plaques with moderate mean velocities exhibit less motion artifacts than images of plaques with lower mean velocities.

Therefore, efficient methods for characterizing cardiac CT images with respect to motion artifacts based on multiple quantitative indices are needed. If such methods can provide image quality ratings that can closely mimic those provided by physicians who are experts in evaluating cardiac CT images, these methods may be implemented as potentially more consistent and efficient ways of accomplishing the primary task of evaluating image quality using the implicit rating patterns established by these experts. Moreover, these methods may be applied to the secondary task of determining optimal phases for interpreting cardiac CT images in a much less time-consuming manner.

In a previous publication, we have presented the use of 12 features (quantitative indices) for characterizing motion-contaminated calcified plaques in noncontrast-enhanced cardiac CT images.²⁰ These features include two dynamic features [e.g., three-dimensional (3D) velocity], six morphological features (e.g., irregularity), and four intensity-based features (e.g., mean intensity). Although individual features, such as 3D velocity, may be imperfect descriptors of motion artifacts in particular and the diagnostic utility of the image in general, linear regression (LR) and artificial neural network (ANN) regression models may be used to merge these features into comprehensive indices. These comprehensive indices, when based on physicians' perceptions of image quality with respect to motion artifacts, potentially can be used to characterize the amount of motion contamination affecting a calcified plaque in a given image as well as select optimal phases for image interpretation in a more consistent and less time-consuming manner. ANN classifiers have been used in many computer-aided diagnosis schemes that involve estimating the probability of malignancy of radiological lesions.^{21–24}

The purpose of this research is to develop an automated computerized scheme for evaluating calcified plaques in noncontrast-enhanced cardiac CT images with respect to motion artifacts. This scheme will be based on using LR and ANN regression models for merging the *temporal phase-specific* dynamic, morphological, and intensity-based features that characterize motion artifacts affecting a calcified plaque within a given region-of-interest (ROI) image. An automated method for obtaining these features directly from projection data through the rapid phase-correlated ROI (RP-ROI) tracking algorithm as well as detailed descriptions of

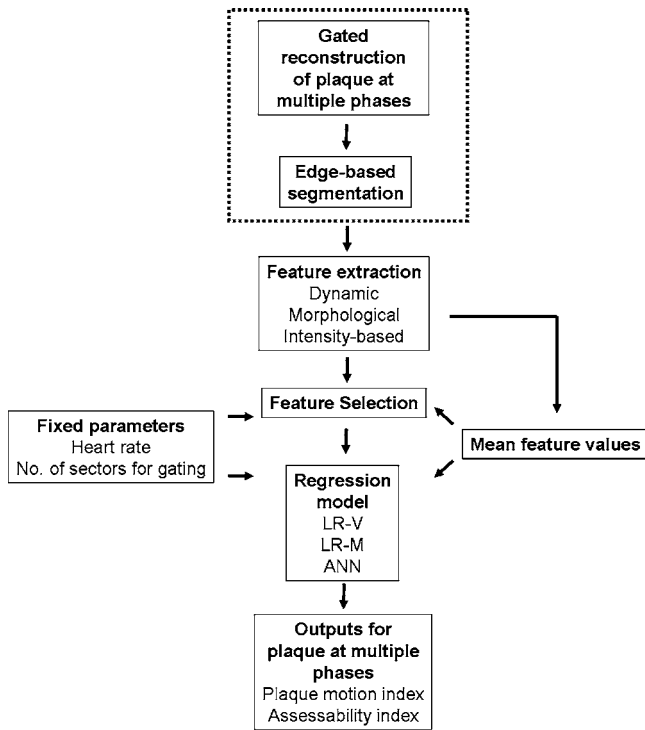


FIG. 1. Flow chart for the computerized assessment of calcified plaques. The dotted square encompasses components of the rapid phase-correlated ROI (RP-ROI) tracking algorithm (Ref. 20). Mean feature values were calculated across different plaques. LR-V, LR-M, and ANN stand for linear regression-velocity, linear regression-multiple, and artificial neural network, respectively.

these features are included in Ref. 20. In addition to these features, additional fixed parameters including *patient-specific* (e.g., heart rate) and *reconstruction-specific* (e.g., number of sectors for gating) values also will be included in the models. The regression models will provide two *patient-specific*, *ROI-specific*, *reconstruction-specific* and *temporal phase-specific* indices: a continuous plaque motion index derived from the actual trajectory of the plaque, as well as an assessability index that is analogous to the five-point image quality rating scales (e.g., 1=excellent: minimal motion artifacts; 5=poor: severe motion artifacts) used in previous studies. The assessability index will reflect the degree to which a calcified plaque within a ROI image is affected by motion artifacts. In addition to developing and evaluating the performance of this scheme, we will examine the potential of using this scheme to determine optimal phases for interpreting cardiac CT images.

This paper is organized as follows. In Sec. II, the relevant materials and methods for developing and evaluating the proposed scheme are discussed. In Sec. III, the results of the proposed experiments are presented. A discussion of the results as well as implications of this research are included in Sec. IV.

II. MATERIALS AND METHODS

The flow chart for the computerized scheme is given in Fig. 1. In short, images of calcified plaques were recon-

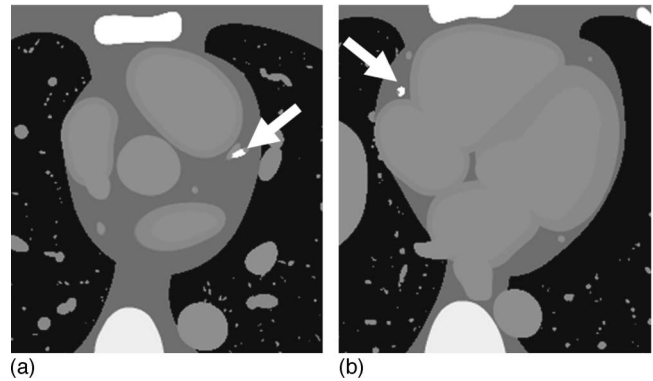


FIG. 2. Transverse slices of the NCAT phantom through the following calcified plaques: (a) LAD2 and (b) RCA1. In each image, a solid white arrow points to the calcified plaque. L: 50 HU/W: 400 HU.

structed and segmented using the RP-ROI algorithm.²⁰ Features were extracted from these images, and then selected through a stepwise regression method. Then, LR and ANN regression models were used to predict the plaque motion indices and assessability indices. The performances of the regression models in producing accurate indices were then evaluated.

II.A. Image database

A database of 2160 simulated images of six calcified coronary plaques from the NURBS-based cardiac-torso (NCAT) phantom^{25–28} was used in this study. Two of these plaques (LAD1 and LAD2) were extracted from the left anterior descending artery; one (LCX) was obtained from the left circumflex artery; and three (RCA1, RCA2, and RCA3) were obtained from the right coronary arteries. The LAD2 and RCA1 plaques are shown in Fig. 2. For each plaque, 360 images were reconstructed at nine heart rate/multi-sector gating combinations. These nine combinations spanned four heart rates of 50, 66, 80, and 90 bpm. Single-sector (G1) gating was used for all heart rates, and two-sector (G2) and three-sector (G3) gating were implemented at and above heart rates of 66 and 80 bpm, respectively. Forty phase-correlated images covering two cardiac cycles at 5% R-R intervals were obtained for each calcified plaque and heart rate/gating combination. All images were reconstructed on a cubic ROI. Each edge of the ROI contained 62 voxels and spanned 24.22 mm. The rapid phase-correlated ROI (RP-ROI) tracking algorithm was used to obtain all reconstructed images in an automated fashion.²⁰

II.B. Image features

Twelve features were extracted from each of these images using the methodology provided in Ref. 20. These features, which are listed in Table I, were classified into the following three categories of dynamic, morphological, and intensity-based features. Plaque 3D velocity (VEL) and 3D acceleration (ACC) were the dynamic features. Morphological features included edge-based volume (VOL-E), threshold-based volume (VOL-T), sphericity (SPHER), irregularity (IRREG),

TABLE I. Twelve features characterizing the motion-contaminated calcified plaques.

Feature	Feature type	Feature description
VEL	Dynamic	3D Velocity
ACC	Dynamic	3D Acceleration
VOL-E	Morphological	Edge-based volume
VOL-T	Morphological	Threshold-based volume
SPHER	Morphological	Sphericity
IRREG	Morphological	Irregularity
MG	Morphological	Margin gradient
VMG	Morphological	Variance of margin Gradient
MAX INT	Intensity-based	Maximum intensity
MEAN INT	Intensity-based	Mean intensity
MIN INT	Intensity-based	Minimum intensity
STD INT	Intensity-based	Standard deviation of intensity

average margin gradient (MG), and variance of margin gradient (VMG). The intensity-based features consisted of maximum intensity (MAX INT), mean intensity (MEAN INT), minimum intensity (MIN INT), and the standard deviation of intensity (STD INT). Note that square roots of the 3D velocity and 3D acceleration features, as well as the square of the sphericity feature, were used, since a previous study had shown that these transformations improved the linearity of the relationships between these features and the plaque motion index described below.²⁰

II.C. Ground truths

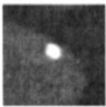
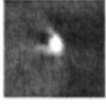
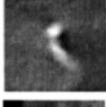


II.C.1. Plaque motion index

For a plaque image reconstructed at a given cardiac phase $\bar{\phi}$, the plaque motion index was defined as the maximum distance between any two points of the actual plaque trajectory mapped out over the phase interval required for reconstruction (single-sector, two-sector, or three-sector) divided by the time window for single-sector short-scan reconstruction.²⁰ Since a phantom was used for this study, the trajectory of the plaque was known for all phases. The motion index was calculated for images from all six plaques in the database.

II.C.2. Assessability index

Assessability indices were assigned by two independent radiologists experienced in examining cardiac CT images, and ranged on an ordinal scale from one to five as shown in Table II. A score of one referred to images of excellent quality, in which the margins of the calcified plaques were well defined and no motion artifacts were apparent. A score of two signified good quality images with well-defined margins and minimal motion artifacts, such as minor blurring. A score of three was assigned to moderate quality images, in which calcified plaques were clearly visible but were affected by moderately sized tail-shaped artifacts. A score of four was given to images of calcified plaques that had poorly defined margins and more severe tail-shaped and smearing artifacts.

TABLE II. Assessability indices and representative examples.

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 indeterminate. Excessive motion artifacts consisting of fragmented plaque regions.

A score of five was assigned to images of calcified plaques that were affected by such excessive motion artifacts that the plaques became fragmented into rather disjointed regions.

Each of the two radiologists provided a separate set of assessability indices for 540 images corresponding to the first cardiac cycle of the LAD2, LCX, and RCA1 plaques. For each plaque, 180 images covering all nine heart rate/gating combinations were rated. The radiologists were presented with images at separate sessions on a computerized interface developed in our lab. Within this interface, images were presented in a random fashion across the three plaques, nine heart rate/gating combinations, and 20 cardiac phases. The radiologists were able to window the images and extract attenuation values in Hounsfield Units (HU) from all of the voxels within the images. However, they were not provided

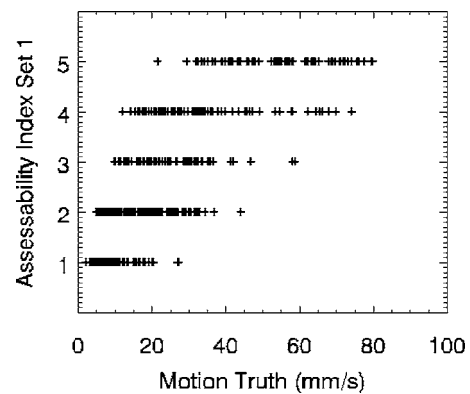


FIG. 3. Ordinal assessability indices and continuous plaque motion indices. Relationship between the set of assessability indices from observer one and the continuous plaque motion index. The point-polyserial coefficient between the two indices was 0.798.

TABLE III. Table showing how the database of plaque images was divided into feature selection, training, and testing groups based on cardiac cycle. In configurations C1 and C2, images in the testing group were taken from the first and second cardiac cycles, respectively.

Group	C1	C2
Feature selection	2nd	1st
Training	2nd	1st
Testing	1st	2nd

with any other information. In this article, the two sets of assessability indices were designated as sets from observer one and observer two.

In order to double the number of images with assessability indices for this study, assessability indices assigned to images within the first cardiac cycle were directly mapped to those corresponding to the second cardiac cycle. As a result, all 1080 images corresponding to the LAD2, LCX, and RCA1 plaques had assessability indices. Although slight variations in indices may have existed for cases in the second cardiac cycle, it is important to note that plaque motion indices of corresponding cases between these two cycles should be identical.

II.C.3. Truth comparisons

Point-polyserial correlation coefficients²⁹ between the continuous plaque motion index and each set of assessability indices were 0.798 and 0.767. Both correlations had p values of less than 0.01 with respect to null hypotheses of zero correlations. However, a given assessability index value often corresponded to a wide range of motion indices, as shown for the case of observer one in Fig. 3. In terms of the agreeability between the two sets of assessability indices provided by the two observers, a repeated measures concordance correlation coefficient (CCC) and associated 95% bootstarp confidence interval of 0.689 [0.648, 0.727] was obtained.³⁰⁻³² As discussed in Sec. II G 1, this statistic was used because it was able to take into account the clustered nature of the feature and truth values, in which all values corresponding to a given plaque were considered repeated measurements and therefore inherently correlated. The mean and standard deviation of the differences in assessability indices between observers one and two were also calculated and determined to be -0.541 ± 0.667 . Although the assess-

ability indices assigned by the two observers were somewhat precise, the indices assigned by observer one were lower on average than those from observer two.

II.D. Experimental design

II.D.1. Experiment A: Prediction of plaque motion indices

The complete database involving all six calcified plaques (2160 images) was divided into three groups for purposes of feature selection, training (parameterization) and testing of the regression models. The following two configurations shown in Table III were used. In the first configuration (C1), the feature selection and training groups consisted of the 1080 images corresponding to the second cardiac cycle, whereas the testing group consisted of all 1080 images corresponding to the first cardiac cycle. For the cases within the second cardiac cycle, one-quarter of the cases were selected for feature selection and three-quarters of the cases were used for training. In the second configuration (C2), associations between the three groups and two cardiac cycles were reversed.

II.D.2. Experiment B: Prediction of observer-assigned assessability indices

All 1080 images involving the LAD2, LCX, and RCA1 plaques were divided into training and testing groups based on the C1 and C2 configurations discussed above. However, since the feature selection results from the previous experiment (experiment A) were used for this experiment, the feature selection groups were merged into the training groups. The two sets of assessability indices provided by the two radiologists were used as separate sets of truths so that the robustness of the regression models against different truths could be evaluated.

II.D.3. Agreeability of feature values over both cardiac cycles

Since the training and testing groups discussed above differed solely in terms of cardiac cycle, differences of motion artifacts and feature values between the two cardiac cycles were evaluated. As seen in Fig. 4, motion artifacts and feature values of the RCA1 plaque (H66/G1) at corresponding phases of the %R-R interval often differed quite noticeably between the two cardiac cycles. Repeated measures CCCs

		$\bar{\phi}$	0.30	0.50	0.70	0.89	1.09	1.29	1.49	1.68	1.88	2.08
Rec												
Feature Values	VEL		36.9	8.4	82.3	15.5	74.0	48.7	1.2	77.2	25.2	44.0
	VOL-E		107.6	79.7	104.0	85.2	125.2	114.4	58.5	165.0	84.8	141.1
	MEAN INT		124	143	118	141	109	106	183	96	123	111
	MG		3.13	3.72	2.80	3.62	2.66	3.49	4.10	2.74	3.97	2.68
"Truths"	Motion Index		45.7	21.3	70.9	22.3	53.9	47.4	18.1	71.1	24.4	52.2
	Assess 1		4	4	5	4	5	4	4	5	4	5
	Assess 2		5	4	5	4	5	5	4	5	4	5

FIG. 4. Reconstructions, selected feature values, and truth values for images of the RCA1 plaque (H66/G1) at selected cardiac phases $\bar{\phi}$ spanning two cardiac cycles. Values for the 3D velocity VEL (mm/s), edge-based volume VOL-E (mm³), mean intensity MEAN INT (HU), and average margin gradient MG features are given. Truths include the plaque motion index (mm/s) and the two sets of assessability indices. Note that assessability indices assigned to the first five images on the left were mapped to the corresponding five images on the right.

TABLE IV. Repeated measures concordance correlation coefficients (CCC) for individual features comparing cases from all six plaques obtained during the first cardiac cycle with those obtained from the second. The CCC for the plaque motion index was 0.992.

Feature	CCC	Feature	CCC
VEL	0.517	ACC	0.620
VOL-E	0.887	IRREG	0.866
VOL-T	0.953	MG	0.588
SPHER	0.906	VMG	0.442
MAX INT	0.928	MIN INT	0.497
MEAN INT	0.954	STD INT	0.929

comparing feature values from all cases in the first cardiac cycle with those from the second were also calculated. As shown in Table IV, the seven features of edge-based volume, threshold-based volume, sphericity, irregularity, maximum intensity, mean intensity, and standard deviation of intensity had repeated measures CCCs greater than or equal to 0.866. The five features of 3D velocity, 3D acceleration, average margin gradient, variance of margin gradient, and minimum intensity had CCCs less than or equal to 0.620. The CCC for the plaque motion index was 0.992.

II.E. Feature selection

Input features for the LR-multiple (LR-M) and ANN regression models discussed below were selected from the 12 available features by applying a stepwise regression method^{33,34} to the feature selection databases. In the stepwise regression method, the continuous plaque motion index was modeled as the dependent variable. The independent variables consisted of the subset of 12 features selected in and out of the regression model as well as seven additional parameters. Two of these were fixed parameters of patient heart rate and the number of sectors used for gating (1, 2, 3). An additional five binary-valued dummy coded variables representing the six plaques were used to take into account differences of mean feature values across the six plaques.³⁵ These seven parameters were retained in the stepwise regression model over all iterations.

Since the stepwise regression model can produce variable results depending on the data, this method was repeated 5000 times through a bootstrap resampling scheme, in which cases from the feature selection databases discussed above were sampled with replacement.³⁶ For each configuration, the feature combination selected with the greatest frequency was used in the LR-M and ANN regression models. In order to determine how strongly the selected features correlated with each other, partial correlation coefficients adjusted for heart rate, number of sectors for gating, and mean feature values for all possible pairs of selected features were analyzed.²⁰

II.F. Regression models

Two implementations of a linear regression (LR) model and one implementation of an ANN regression model were used for predicting plaque motion indices and assessability

indices. With respect to the linear regression model,³⁵ LR-velocity (LR-V) used the 3D velocity feature as its key independent variable, whereas LR-multiple (LR-M) used the feature combinations that were selected with the stepwise regression method discussed above. The ANN regression model³⁷ also used these same feature combinations.

The three regressions were parametrized using the training cases, and evaluated with the testing cases. Fixed parameters of heart rate, number of gated sectors, as well as mean feature values for a given plaque, heart rate, and gating combination were included as independent variables in order to adjust for these covariates. Mean feature values were included only for those features implemented in a given regression model. Furthermore, separate mean feature values were calculated for images in the training and testing groups. Dummy codes were not used in order to ensure that regression models parametrized with one subset of plaque images could be applied to another subset of images from different plaques.

II.F.1. Linear regression models

For the continuous plaque motion index, both LR-V and LR-M were based on ordinary least squares regression. For the assessability indices, the LR-V and LR-M used ordinal logistic regression formulated in terms of the proportional odds model. Since assessability indices represented an ordinal dependent variable, the ordinal logistic regression model was better suited for predicting assessability indices.^{35,38,39}

II.F.2. Artificial neural network regression

The ANN regression model³⁷ was used because this model was capable of mapping local nonlinearities between the selected features and a given truth. The ANN regression 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 linear output activation function was used, because an ANN using this type of activation function was better capable of reproducing the entire range of output values in regression tasks. For an ANN with a sigmoidal output activation function, which is used in classification tasks, output values near both extremes of this range would have been much more difficult to reproduce.⁴⁰ The ANN model used the same independent variables as the LR-M. Either the plaque motion index or assessability index served as the dependent variable. The manner in which the ANN was implemented differed based on the truth that was used.

For the continuous plaque motion index, an N_{in} -20-1 (number of input-hidden-output units) structure was utilized. The N_{in} input units corresponded to the number of independent variables, and the one output unit represented the value of the continuous truth. The 20 hidden units were used to generate mappings between the input and output units. The number of training epochs was set to 2000. The number of hidden units and the number of training epochs were tuned using the repeated measures CCC as a metric of regression model performance.

The ANN regression models for obtaining assessability indices used an N_{in} -10-5 structure. Since assessability indices were defined on a five-point ordinal scale, the five output units were used to represent assessability indices using “thermometer” code.⁴¹ In thermometer code, each of the five units represented a binary variable. The value of the assessability index was equivalent to the number of consecutive nodes beginning from the first node that were set to one. For example, an assessability index of two was represented by the first two nodes being set to one and the rest of the nodes set to zero. The ANN regression model was trained with thermometer code using 500 training epochs, and then implemented on the testing database. Since the values for the output units of the testing cases were continuous, these values first were rounded and then summed up to re-obtain assessability indices on the five-point scale. The values were not summed and then rounded in order to prevent the accumulation of small deviations from zero and one across the five nodes from affecting the final result. The ten hidden units and 500 training epochs were tuned by using the repeated measures CCC as a metric of regression model performance. Although the numbers of hidden units and training epochs were less than those used for the continuous plaque motion indices, it is important to remember that the assessability indices were defined on an ordinal scale.

II.G. Evaluation

II.G.1. Performance evaluations of regression models

The three regressions were evaluated using the repeated measures concordance correlation coefficient (CCC)³⁰⁻³² and metrics from Bland–Altman analysis.⁴²

II.G.1.a. Repeated measures concordance correlation coefficient The repeated measures CCC was used to evaluate how well the predicted values for the plaque motion index and assessability index, as obtained from the three regressions, matched with their respective actual and assigned values for cases in the testing group. The nonoverlapping block bootstrap resampling technique^{43,44} was used to obtain a mean and 95% confidence interval for each CCC value. In this re-sampling scheme, 5000 bootstrap samples of nonoverlapping blocks consisting of images at five consecutive phases were obtained from both the training and testing groups. Confidence intervals were computed using the percentile method.^{45,46}

In experiment A, which involved the prediction of plaque motion indices, one CCC value was calculated for each of the three regressions and two configurations C1 and C2. Each CCC value was calculated from the 1080 test images involving all six plaques. For experiment B, which involved the prediction of assessability indices, two CCC values were calculated for each of the three regressions and two configurations. The two values corresponded to the sets of assessability indices from observers one and two. All CCC values in Experiment B were calculated from the 540 test images corresponding to the LAD2, LCX, and RCA1 plaques. Re-

peated measures CCCs were also used to compare the performances of the three regressions for each of the nine heart rate/gating combinations.

II.G.1.b. Bland–Altman analysis The accuracy and precision of the predicted plaque motion and assessability indices were calculated using metrics from Bland–Altman analysis in the following manner. First, differences Δ between the predicted and actual/assigned values for each truth were calculated. The accuracy was represented as the mean of these differences, whereas the precision was represented by the standard deviation of these differences. For the continuous plaque motion indices, Bland–Altman plots showing the relationships of differences Δ to the means of the predicted and actual values were created.⁴² These plots were not created for the assessability indices due to their ordinal nature.

II.G.2. Comparative evaluation of predicted indices and their corresponding images

Individual graphs showing temporal relationships of plaque motion indices and assessability indices versus cardiac phase were created for different plaque, heart rate, and gating combinations. For each combination, predicted motion and assessability indices were compared with actual motion and assigned assessability indices, respectively. Images with low values of plaque motion and/or assessability indices were of special interest, since these images possibly corresponded to optimal phases for image interpretation.

III. RESULTS

III.A. Feature selection

Table V shows the top three feature combinations selected by the bootstrapped stepwise regression method discussed in Sec. II E. For the C1 configuration, the following six features of 3D velocity, 3D acceleration, edge-based volume, threshold-based volume, sphericity, and standard deviation of intensity were selected the greatest number of times (1446 out of 5000 trials). For the C2 configuration, the following eight features of 3D velocity, 3D acceleration, edge-based volume, threshold-based volume, sphericity, mean intensity, standard deviation of intensity and variance of margin gradient were selected the greatest number of times (464 out of 5000 trials).

In terms of correlations between the selected features, the feature pair of 3D velocity and standard deviation of intensity exhibited the highest partial correlation coefficient in terms of absolute value of 0.640 for the features selected from the C1 configuration. For the features selected from the C2 configuration, the feature pair of mean intensity and standard deviation of intensity had the highest partial correlation coefficient in terms of absolute value of 0.744.

III.B. Experiment A: Prediction of plaque motion indices

In Table VI, repeated measures CCCs are presented for the three regressions from the C1 and C2 configurations. For both C1 and C2 configurations, the ANN provided higher

TABLE V. Bootstrap resampling ($N=5000$) of the stepwise regression method for feature selection, as discussed in Sec. II E. The tables in (a) and (b) show the top three feature combinations selected by this method for the C1 and C2 configurations, respectively.

	Selected features	No. of selected features	Frequency
(a)	VEL, ACC, VOL-E, VOL-T, SPHER, STD INT	6	1446
	VEL, ACC, VOL-E, VOL-T, SPHER, IRREG, STD INT	7	454
	VEL, ACC, VOL-E, VOL-T, SPHER, MEAN INT, STD INT	7	384
	Selected features	No. of selected features	Frequency
(b)	VEL, ACC, VOL-E, VOL-T, SPHER, MEAN INT, STD INT, VMG	8	464
	VEL, ACC, VOL-E, VOL-T, SPHER, IRREG, MEAN INT, STD INT, VMG	9	347
	VEL, ACC, VOL-T, STD INT	4	264

CCC values than those from the LR-M and LR-V. However, CCC values between the ANN and LR-M were quite comparable. The LR-V, on the other hand, produced much lower CCC values. For a given regression, the CCC values and confidence intervals also differed slightly between the C1 and C2 configurations. The different sets of features chosen, as shown in Table V, and variations in the training and testing cases may have accounted for some of these disparities.

Table VI also provides means and standard deviations of differences Δ between predicted and actual plaque motion indices from the three regressions. Although no distinct trends were observed in the mean differences, the standard deviations were the smallest for the ANN and the largest for the LR-V over both C1 and C2 datasets. The results for the ANN were the most precise. In the Bland–Altman plots shown in Fig. 5 for the C1 configuration, the differences Δ were more tightly clustered around zero for the ANN and were more scattered for the LR-V. Also, for the LR-V, the differences seemed to exhibit a pronounced nonlinear trend with respect to the mean values of the predicted and actual motion indices.

Figure 6(a) contains a graph of CCC values for the three regressions over the nine heart rate/gating combinations for the C1 configuration. As seen in this figure, the ANN was associated with the highest CCC values over most combina-

tions. However, the CCC values from the LR-M were quite comparable, especially for the H80/G2, H80/G3, and H90/G2 combinations. The LR-V, on the other hand, performed worse than the LR-M over all combinations.

III.C. Experiment B: Prediction of assembly indices

CCC values for both sets of assessability indices are provided in Table VII. As seen in this table, both LR-M and ANN were capable of providing indices that matched with the assessability indices from both sets of observers quite well. Although mean CCC values for the ANN were greater than those for the LR-M, the 95% confidence intervals demonstrated considerable overlap for both configurations and sets of indices. The CCC values for the LR-V, on the other hand, were lower than those obtained from the other regressions.

In terms of means and standard deviations of differences Δ between predicted and assigned assessability indices, Table VII also shows that all three regressions were associated with low mean differences for most conditions considered. In terms of precision, the LR-V tended to exhibit the largest standard deviations. The standard deviations between the LR-M and ANN, on the other hand, were much more comparable.

TABLE VI. Experiment A: Regression model performance for the prediction of plaque motion indices. Repeated measures concordance correlation coefficients (CCC) with 95% confidence intervals, as well as means and standard deviations of differences Δ between predicted and actual indices, are included. Means and standard deviations of differences Δ are provided as metrics of accuracy and precision, respectively. Values are included for the linear regression-velocity (LR-V), linear regression-multiple (LR-M), and the artificial neural network (ANN) regressions for the C1 and C2 configurations.

Model	CCC		Difference Δ	
	C1 Configuration	C2 Configuration	C1 Configuration	C2 Configuration
LR-V	0.817 [0.785, 0.848]	0.816 [0.784, 0.845]	0.074 \pm 9.170	0.211 \pm 9.232
LR-M	0.894 [0.869, 0.916]	0.904 [0.885, 0.920]	-0.017 \pm 6.902	0.070 \pm 6.549
ANN	0.917 [0.892, 0.936]	0.911 [0.888, 0.931]	-0.237 \pm 5.598	-0.101 \pm 6.116

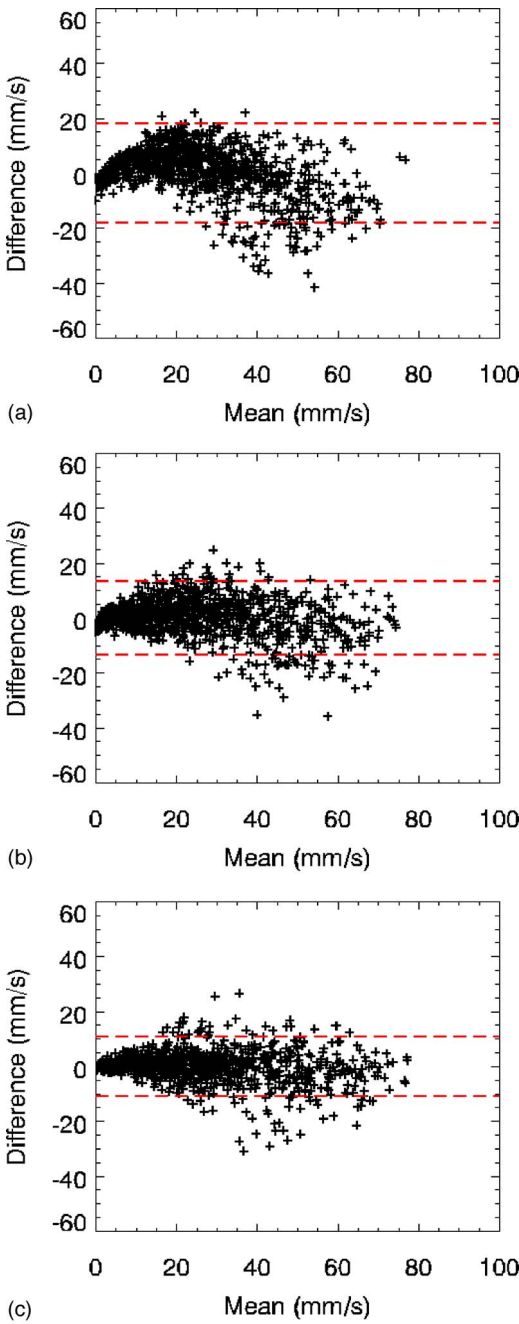


FIG. 5. Bland–Altman plots of predicted and actual plaque motion indices for the (a) linear regression-velocity (LR-V), (b) linear regression-multiple (LR-M), and (c) ANN regressions. In each plot, the two dashed lines represent the 95% confidence interval for the differences in predicted and actual motion indices.

In Figs. 6(b) and 6(c), graphs of CCC values for both sets of assessability indices are shown for the nine heart rate/gating combinations. In both graphs, the performances of the LR-M and ANN seemed quite comparable. The LR-V, on the other hand, had the lowest CCC values over all combinations.

In Fig. 7, the percentages of images with specified differences between ANN-predicted and assigned assessability indices for the two sets of assessability indices under the C1

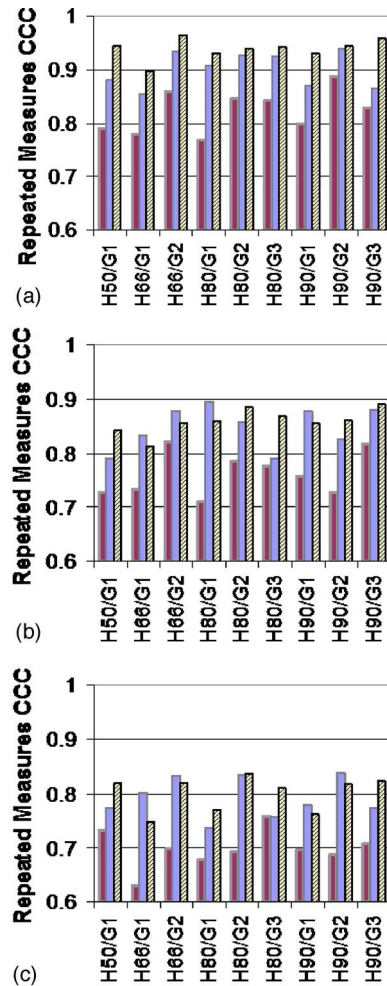


FIG. 6. Bar plots of repeated measures concordance correlation coefficients (CCC) for (a) predicted and actual plaque motion indices, and predicted and assigned assessability indices from observers (b) one and (c) two over the nine heart rate/gating combinations. The darkest bars correspond to LR-V regression. The semidark bars are associated with LR-M regression. The lightest bars with the dark hashes correspond to ANN regression.

and C2 configurations are shown. As seen in this table, the majority of the images were assigned the correct assessability index, whereas most of the images with discrepancies had differences of -1 or 1 . No images had differences of more than 2 in magnitude.

III.D. Comparative evaluation of predicted indices and their corresponding images

In Fig. 8, graphs showing the temporal relationships of ANN-predicted plaque motion indices and ANN-predicted assessability indices versus cardiac phase are provided for the LAD2 and RCA1 plaques at 66 bpm.

III.D.1. LAD2 plaque

For the LAD2 plaque at 66 bpm, the predicted plaque motion indices compared very well with the actual motion indices for both G1 and G2 cases, as shown in Figs. 8(a) and 8(b). As seen in these figures, the ANN model was able to predict assessability indices that differed from their assigned

TABLE VII. Experiment B: Regression model performance for the prediction of the two sets of assessability indices provided by observers (a) one and (b) two. Repeated measures concordance correlation coefficients (CCC) with 95% confidence intervals, as well as means and standard deviations of differences Δ between predicted and assigned indices, are included.

		CCC		Difference Δ	
Model		C1 Configuration	C2 Configuration	C1 Configuration	C2 Configuration
(a)	LR-V	0.761 [0.715, 0.802]	0.727 [0.675, 0.776]	-0.145±0.759	-0.037±0.843
	LR-M	0.834 [0.794, 0.868]	0.786 [0.733, 0.827]	-0.063±0.606	0.018±0.710
	ANN	0.843 [0.791, 0.877]	0.808 [0.764, 0.846]	-0.052±0.582	-0.061±0.682
		CCC		Difference Δ	
Model		C1 Configuration	C2 Configuration	C1 Configuration	C2 Configuration
(b)	LR-V	0.692 [0.643, 0.735]	0.676 [0.628, 0.717]	-0.013±0.712	0.043±0.756
	LR-M	0.782 [0.738, 0.820]	0.703 [0.621, 0.762]	-0.006±0.586	-0.030±0.686
	ANN	0.793 [0.747, 0.828]	0.745 [0.696, 0.788]	0.007±0.568	0.032±0.660

value by a value of one for most cases. The predicted assessability index for image C in panel (b) was two values lower than the assigned index, although the motion artifacts affecting the plaque did not seem to be much greater in severity. Images A and E in panel (a) had the lowest predicted assessability indices, and low levels of motion artifacts. For the G2 images in panel (b), images A, B, D, and E had the lowest predicted assessability indices, and seemed to exhibit lower levels of motion artifacts.

III.D.2. RCA1 Plaque

Figures 8(c) and 8(d) contained graphs and reconstructions of the RCA1 plaque at 66 bpm. Similar to the previous case, the predicted and assigned assessability indices matched quite well, especially for the G2 case. Predicted and assigned assessability indices differed by a value of one at most. For the G1 case, images B and D in panel (c) had the lowest predicted assessability indices. These images also seemed to contain the least amounts of motion artifacts. For the G2 case, images B and D had the lowest predicted assessability indices and the clearest plaque margins.

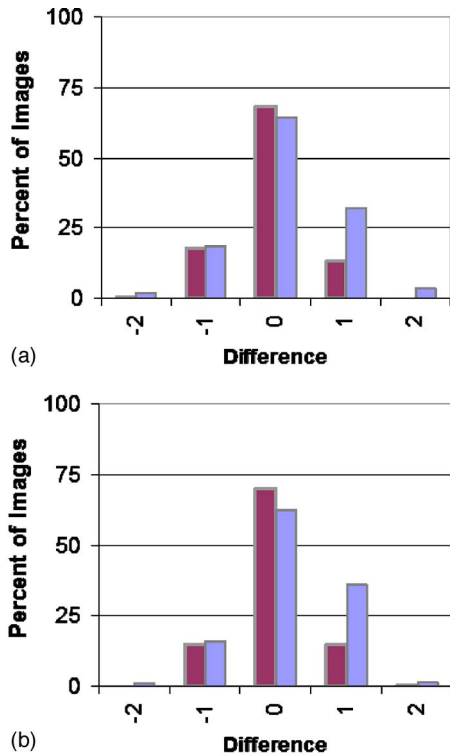


FIG. 7. Percentages of images (out of 540) with shown differences between ANN-predicted and assigned assessability indices from observers (a) one and (b) two. The darkest bars correspond to images from the C1 configuration, and the lighter bars correspond to images from the C2 configuration. No images had absolute differences that were greater than two.

IV. DISCUSSION

In this paper, a computerized scheme for evaluating calcified plaques with respect to motion artifacts in noncontrast-enhanced cardiac CT images has been presented. The linear regression (LR) and artificial neural network (ANN) regression models were used to predict the continuous plaque motion indices and ordinal assessability indices. LR-velocity (LR-V) used 3D velocity as its main independent variable, whereas LR-multiple (LR-M) and ANN regressions used multiple features selected through stepwise regression. The results of this study show that the ANN provided the most agreeable plaque motion and assessability indices in terms of the repeated measures concordance correlation coefficient (CCC). Also, this model was the most precise in terms of having the smallest standard deviations of differences between the predicted and actual values of these two indices. However, the LR-M performed quite comparably in predicting both plaque motion and assessability indices. Furthermore, it did not require the additional training time associated with the ANN model.

The lower performance of the LR-V was expected, since it only used the 3D velocity feature for predicting motion and assessability indices. The LR-M and ANN models used multiple features, which can allow for better predictive performance. The slightly better performance of the ANN with

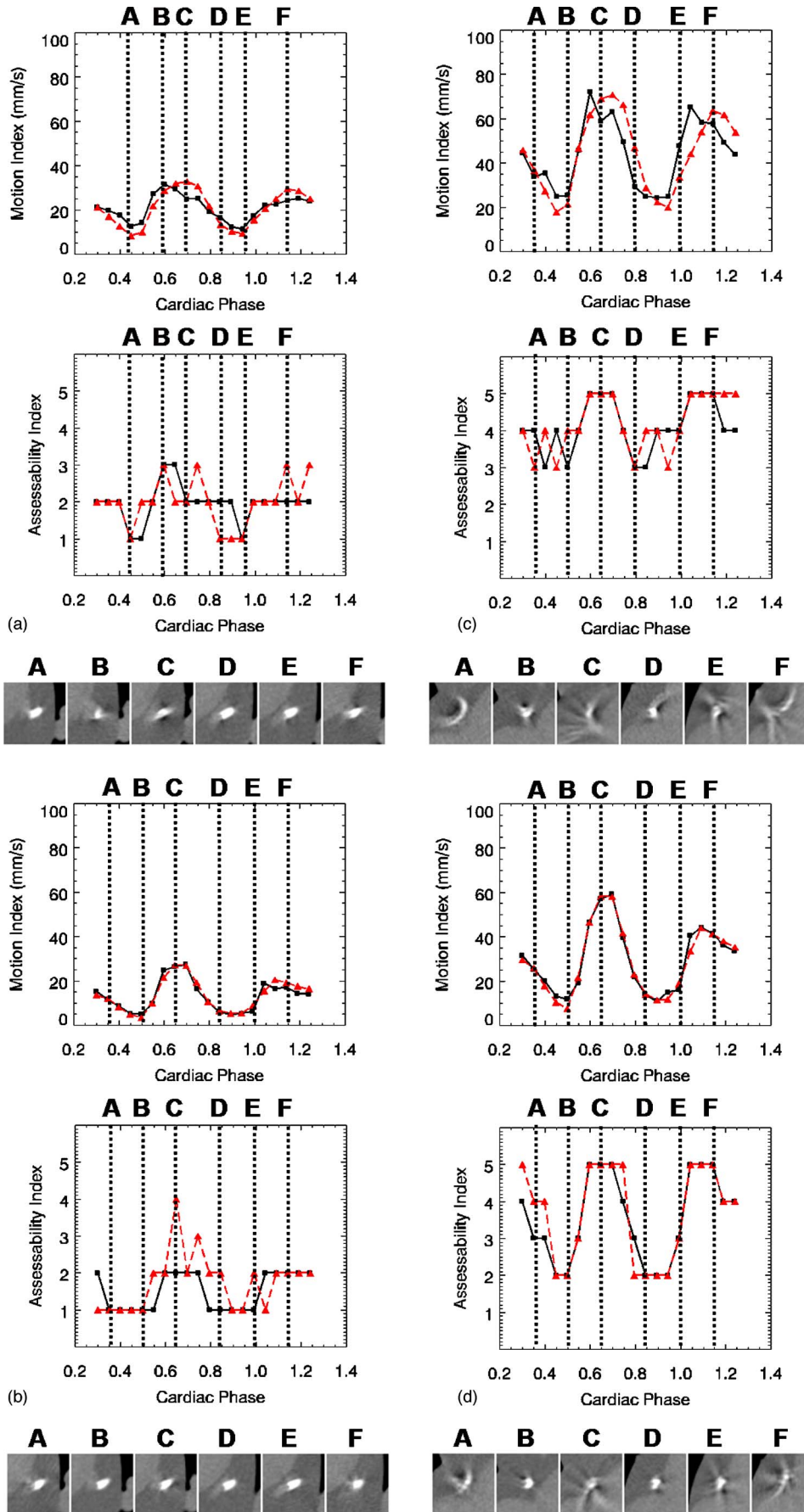


FIG. 8. Graphs showing relationships between assessability indices from observer one and cardiac phase, as well as plaque motion indices and cardiac phase, for the (a)-(b) LAD2 plaque at 66 bpm and (c)-(d) RCA1 plaque at 66 bpm. Panels (a) and (c) correspond to G1 reconstructions, whereas panels (b) and (d) correspond to G2 reconstructions. In all plots, the dark solid lines represent predicted values from the ANN regression model, and the light dashed lines represent the actual motion indices or assigned assessability indices.

respect to the LR-M may be due in part to the ANN's natural ability for mapping local nonlinearities between independent (feature) and dependent (truth) variables.

In terms of assessability indices, separate regression models were trained for the two sets of indices provided by the two radiologists. For the C1 configuration, CCC values and 95% confidence intervals for the ANN model were 0.843 [0.791, 0.877] and 0.793 [0.747, 0.828] for the two sets of indices. These two CCC values were statistically greater than the CCC value of 0.689 (95% confidence interval [0.648, 0.727]), which was obtained by comparing the two sets of assessability indices with each other. These preliminary results suggest that the variabilities of assessability indices assigned by the ANN model can lie within the variabilities of the indices assigned by independent observers.

The graphs depicting temporal relationships between the continuous plaque motion indices and cardiac phase for different plaques and heart rate/gating combinations (see Fig. 8) show that the ANN regression model was capable of predicting motion indices that matched with the actual indices quite well over multiple cardiac phases. For the two-sector (G2) and three-sector (G3) (not shown) reconstructions, the predicted motion indices followed the actual indices very closely. For the single-sector (G1) reconstructions, small deviations between the predicted and actual indices slightly distorted the overall trends of the indices for some plaques. In terms of the five-point assessability indices, the predicted indices provided by the ANN model matched the assigned indices for the majority of the cases (see Fig. 7). Differences from the assigned indices usually spanned a factor of one, although differences of two were found in a smaller percentage of cases. Despite these differences, the predicted assessability indices were useful in that cases with lower levels of motion artifacts could be identified. Our results show that assessability indices potentially can be used to select optimal phases for image interpretation.

The regression models achieved better performance in terms of the repeated measures CCC for the motion indices than for the assessability indices. One key reason was that the motion and assessability indices were defined on different scales. Since plaque motion indices were defined on a continuous scale, the computer may have been able to better differentiate images with more subtle differences in extracted feature characteristics during parameterization of the regression models. Assessability indices were provided on a five-point ordinal scale, and the regression models may not have been able to differentiate these subtle differences. Second, the assessability indices assigned by the radiologists contained an extra element of uncertainty due to the subjective nature of rating images with respect to motion artifacts. For a few cases, this uncertainty may have presented itself as inconsistent ratings between images with similar types and extents of motion artifacts.

The following limitations should be noted when interpreting the results of this study. First, although this study showed that the LR-M and ANN regression models could assign accurate motion and assessability indices to different plaques at different heart rate/gating combinations, the feature selec-

tion, training, and testing groups used in this study were not completely independent. These groups drew from repeated measurements of the same six plaques obtained using nine heart rate/gating combinations over two cardiac cycles. The main difference between the cases in the training and testing groups were that they were drawn from different cardiac cycles. It is shown in Table IV and Fig. 4 that the features at the same %R-R intervals between the two cardiac cycles were related but not absolutely agreeable. However, additional studies involving truly independent training and testing groups of plaques with widely varying properties such as size, shape, and intensity are needed in order to validate the proposed techniques. Given that the regression models in this study used between six and eight independent features, repeated measurements of over 100 independent plaque samples may be needed to build regression models that can handle completely independent training and testing groups.

A second limitation is that both the LR-M and ANN regression models used the same set of features selected through stepwise regression. ANN-specific methods for variable selection, such as node pruning,⁴⁷ may be well suited for larger plaque databases. A third limitation is that this scheme was applied to simulated calcified plaque images instead of clinical images.

In terms of potential clinical applications, one key benefit of this proposed scheme, which integrates image reconstruction, image segmentation, feature selection, and regression, is that it is completely automated. The only required user interaction is the specification of the plaque location at an initial cardiac phase. As described in Ref. 20, the rapid phase-correlated ROI (RP-ROI) tracking algorithm uses this location to produce phase-correlated reconstructions of the plaque through advancing cardiac phases. From these reconstructions, the computer can automatically segment the images, extract the relevant features, and provide values of motion indices or assessability indices based on already trained regression models. For the assessability indices, if the physicians who assigned these indices are expert observers, the assessability indices provided by the ANN model potentially may be used as an aid by other physicians for evaluating image quality with respect to motion artifacts or finding optimal phases for image interpretation.

The proposed scheme potentially can find immediate applicability for coronary calcium scans obtained with retrospective gating, such as those using low-dose protocols.⁴⁸ Once a physician specifies the location of the calcified plaque at an initial cardiac phase, the scheme can output graphs showing the temporal relationships of motion indices or assessability indices to cardiac phase, as shown in Fig. 8. In addition, a select group of images with low motion indices or assessability indices can be displayed on the screen. The physician would only need to evaluate a few images with low assessability indices rather than search through images at multiple cardiac phases in order to find diagnostic images. Substantial time savings potentially can be achieved.

The greatest clinical benefit of the proposed scheme may be for patients with highly varying or arrhythmic heart rates. Since current suggestions of optimal phase for image inter-

pretation, such as phases in mid-diastole for lower heart rate patients and late systole for higher heart rate patients,¹² may not apply to these cases, physicians must often spend a great deal of time searching for these optimal phases. By using the proposed computerized scheme, which assigns assessability indices on an image-by-image basis, the physicians could obtain optimal phases for evaluating calcified coronary plaques much more quickly. Assessability indices have the potential to become a *patient-specific*, *ROI-specific*, *reconstruction-specific* and *temporal phase-specific* method for evaluating image quality with respect to motion artifacts.

Our future work involves applying our methodology to clinical images of calcified plaques at multiple cardiac phases, and validating this methodology using independent training and testing databases. Furthermore, our group will attempt to extend this proposed framework for evaluating contrast-enhanced coronary arteries in CT coronary angiography.

V. CONCLUSION

We have investigated computerized image analysis methods, which output assessability indices characterizing the levels of motion artifacts affecting calcified plaques in noncontrast-enhanced cardiac CT scans. The results of this study suggest that the assessability indices obtained from these methods fall within the range of inter-observer variability. Furthermore, output from such eventual online analyses of computer-extracted features is expected to provide optimal phases for the interpretation of calcified plaques in cardiac CT.

ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health Medical Scientist Training Program Grant, National Institutes of Health Grants 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, Dr. Lorenzo Pesce, Amy Feng, and Zach Rodgers for the helpful discussions. The authors also would like to thank Dr. William Segars and Dr. Benjamin Tsui for granting permission to use the NCAT phantom.

^{a)}Electronic mail: mtking@uchicago.edu

¹A. W. Leber, A. Knez, F. von Ziegler, A. Becker, K. Nikolaou, S. Paul, B. Wintersperger, 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. Budoff, H. Bakhsheshi, and S. 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. As-*

sist. Tomogr. **26**, 880–885 (2002).

⁵R. Detrano, M. Anderson, J. Nelson, N. Wong, J. Carr, M. McNitt-Gray, and D. 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).

⁷K. Nieman, M. Oudkerk, B. Rensing, P. van Ooijen, A. Munne, R. van Geuns, and P. de Feyter, "Coronary angiography with multi-slice computed tomography," *Lancet* **357**, 599–603 (2001).

⁸D. 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).

⁹G. 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).

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. 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).

¹⁴A. Kopp, S. Schroeder, A. Kuettner, M. Heuschmid, C. Georg, B. Ohnesorge, R. Kuzo, and C. Claussen, "Coronary arteries: Retrospectively ECG-gated multi-detector row CT angiography with selective optimization of the image reconstruction window," *Radiology* **221**, 683–688 (2001).

¹⁵M. Kachelriess, D. Sennst, W. Maxlmoser, and W. A. Kalender, "Kymogram detection and kymogram-correlated image reconstruction from sub-second spiral computed tomography scans of the heart," *Med. Phys.* **29**, 1489–1503 (2002).

¹⁶G. Wang, S. Y. Zhao, and D. Heuscher, "A knowledge-based cone-beam x-ray CT algorithm for dynamic volumetric cardiac imaging," *Med. Phys.* **29**, 1807–1822 (2002).

¹⁷R. Manzke, T. Köhler, T. Nielsen, D. Hawkes, and M. Grass, "Automatic phase determination for retrospectively gated cardiac CT," *Med. Phys.* **31**, 3345–3362 (2004).

¹⁸S. Achenbach, D. Ropers, J. Holle, G. Muschiol, W. G. Daniel, and W. Moshage, "In-plane coronary arterial motion velocity: Measurement with electron-beam CT," *Radiology* **216**, 457–463 (2000).

¹⁹M. Vembar, M. Garcia, D. Heuscher, R. Haberl, D. Matthews, G. Bohme, and N. Greenberg, "A dynamic approach to identifying desired physiological phases for cardiac imaging using multislice spiral CT," *Med. Phys.* **30**, 1683–1693 (2003).

²⁰M. 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).

²¹S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. (Prentice-Hall, Upper Saddle River, NJ, 1999).

²²Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," *Radiology* **187**, 81–87 (1993).

²³Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign mass lesions on digitized mammograms," *Acad. Radiol.* **5**, 155–168 (1998).

²⁴Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz,

- "Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness," *Acad. Radiol.* **7**, 1077–1084 (2000).
- ²⁵W. 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).
- ²⁶W. P. Segars, Ph.D. thesis, The University of North Carolina (2001).
- ²⁷J. 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).
- ²⁸W. 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**, 18 (2006).
- ²⁹H. Kraemer, "Correlation coefficients in medical research: From product moment correlation to the odds ratio," *Stat. Methods Med. Res.* **15**, 525–545 (2006).
- ³⁰T. S. King, V. M. Chinchilli, and J. L. Carrasco, "A repeated measures concordance correlation coefficient," *Stat. Med.* **16**, 3095–3113 (2007).
- ³¹L. I. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics* **45**, 255–268 (1989).
- ³²T. S. King and V. M. Chinchilli, "A generalized concordance correlation coefficient for continuous and categorical data," *Stat. Med.* **20**, 2131–2147 (2001).
- ³³M. L. Thompson, "Selection of variables in multiple regression: Part I: a review and evaluation," *Int. Statist. Rev.* **46**, 1–19 (1978).
- ³⁴M. L. Thompson, "Selection of variables in multiple regression: Part II: chosen procedures, computations, and examples," *Int. Statist. Rev.* **46**, 129–146 (1978).
- ³⁵J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. (Lawrence Erlbaum, Mahwah, NJ, 2003).
- ³⁶B. Efron and G. Gong, "A leisurely look at the bootstrap, the jackknife and cross-validation," *Am. Stat.* **37**, 36–48 (1983).
- ³⁷K. 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).
- ³⁸P. McCullagh, "Regression models for ordinal data," *J. R. Stat. Soc. Ser. B (Methodol.)* **42**, 109–142 (1980).
- ³⁹J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression* (Chapman and Hall, Boca Raton, FL, 2006).
- ⁴⁰K. Suzuki, S. G. Armato III, F. Li, S. Sone, and K. Doi, "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography," *Med. Phys.* **30**, 1602–1617 (2003).
- ⁴¹M. Smith, *Neural Networks for Statistical Modeling* (Van Nostrand Reinhold, New York, 1993).
- ⁴²J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **1**, 307–310 (1986).
- ⁴³E. Carlstein, "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *Ann. Stat.* **14**, 1171–1179 (1986).
- ⁴⁴P. Hall, J. L. Horowitz, and B.-Y. Jing, "On blocking rules for the bootstrap with dependent data," *Biometrika* **82**, 561–574 (1995).
- ⁴⁵B. Efron, "Nonparametric standard errors and confidence intervals," *Can. J. Stat.* **9**, 139–158 (1981).
- ⁴⁶J. Carpenter and J. Bithell, "Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians," *Stat. Med.* **19**, 1141–1164 (2000).
- ⁴⁷J. Sietsma and R. J. F. Dow, "Creating artificial neural networks that generalize," *Neural Networks* **4**, 67–79 (1991).
- ⁴⁸J. Horiguchi, H. Yamamoto, N. Hirai, Y. Akiyama, C. Fujioka, K. Marukawa, H. Fukuda, and K. Ito, "Variability of repeated coronary artery calcium measurements on low-dose ECG-gated 16-MDCT," *AJR, Am. J. Roentgenol.* **187**, W1–W6 (2006).