Predicting the fidelity of JPEG2000 compressed CT images using DICOM header information

Kil Joong Kim
Department of Radiation Applied Life Science, Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul, 110-744, Korea

Bohyoung Kim and Hyunna Lee
School of Computer Science and Engineering, Seoul National University, 599 Kwanak-Ro, Kwanak-Gu, Seoul, 151-742, Korea

Hosik Choi
Department of Informational Statistics, Hoseo University, 165, Sechul-ri, Baebang-myeon, Asan-si, Chungcheongnam-do, 336-795, Korea

Jong-June Jeon
Department of Statistics, Seoul National University, 599 Kwanak-Ro, Kwanak-Gu, Seoul, 151-742, Korea

Jeong-Hwan Ahn
Korean Intellectual Property Office, Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon, 302-701, Korea

Kyoung Ho Lee[a)]
Department of Radiology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Institute of Radiation Medicine, and Seoul National University Medical Research Center, 300 Gumi-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-707, Korea

(Received 26 April 2011; revised 9 October 2011; accepted for publication 11 October 2011; published 10 November 2011)

Purpose: To propose multiple logistic regression (MLR) and artificial neural network (ANN) models constructed using digital imaging and communications in medicine (DICOM) header information in predicting the fidelity of Joint Photographic Experts Group (JPEG) 2000 compressed abdomen computed tomography (CT) images.
Methods: Our institutional review board approved this study and waived informed patient consent. Using a JPEG2000 algorithm, 360 abdomen CT images were compressed reversibly ($n = 48$, as negative control) or irreversibly ($n = 312$) to one of different compression ratios (CRs) ranging from 4:1 to 10:1. Five radiologists independently determined whether the original and compressed images were distinguishable or indistinguishable. The 312 irreversibly compressed images were divided randomly into training ($n = 156$) and testing ($n = 156$) sets. The MLR and ANN models were constructed regarding the DICOM header information as independent variables and the pooled radiologists' responses as dependent variable. As independent variables, we selected the *CR* (DICOM tag number: 0028, 2112), effective tube current-time product (0018, 9332), section thickness (0018, 0050), and field of view (0018, 0090) among the DICOM tags. Using the training set, an optimal subset of independent variables was determined by backward stepwise selection in a four-fold cross-validation scheme. The MLR and ANN models were constructed with the determined independent variables using the training set. The models were then evaluated on the testing set by using receiver-operating-characteristic (ROC) analysis regarding the radiologists' pooled responses as the reference standard and by measuring Spearman rank correlation between the model prediction and the number of radiologists who rated the two images as distinguishable.
Results: The *CR* and section thickness were determined as the optimal independent variables. The areas under the ROC curve for the MLR and ANN predictions were 0.91 (95% CI; 0.86, 0.95) and 0.92 (0.87, 0.96), respectively. The correlation coefficients of the MLR and ANN predictions with the number of radiologists who responded as distinguishable were 0.76 (0.69, 0.82, $p < 0.001$) and 0.78 (0.71, 0.83, $p < 0.001$), respectively.
Conclusions: The MLR and ANN models constructed using the DICOM header information offer promise in predicting the fidelity of JPEG2000 compressed abdomen CT images. © *2011 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3656963]

Key words: computed tomography, image compression, image fidelity, DICOM, machine learning algorithm

I. INTRODUCTION

Although the cost of storage and network resources have continued to drop, there is increasing demand for irreversible compression of computed tomography (CT) images for long-term preservation and efficient transmission of data.[1–4] As the compression level increases, the data size decreases; however, the amount of artifacts progressively increases, eventually affecting the diagnosis. Therefore, it is important to achieve an optimal compression level for a CT image, which provides maximum data reduction while preserving image fidelity.

Recently, several researchers[5–9] have demonstrated that it is feasible to compress a CT image adaptively and automatically to its own optimal compression level predetermined by a computerized image fidelity metric, which measures the fidelity of a distorted image by analyzing the characteristics of image contents in comparison to its original image.[10]

Compression artifacts are dependent on the characteristics of the image being compressed.[8,11–14] Having noticed that some of the image characteristics affecting the compression artifacts are already available in digital imaging and communications in medicine (DICOM) header describing patient information and image metadata,[15] we hypothesized that the fidelity of CT images compressed to a certain compression level can be predicted solely from the DICOM header information. An approach using such semantic information should be advantageous in terms of computational efficiency compared to the previously reported approaches that analyze the image characteristics themselves.

The purpose of this study was to examine the feasibility of multiple logistic regression (MLR) and artificial neural network (ANN) models constructed using the DICOM header information in predicting the fidelity of Joint Photographic Experts Group (JPEG) 2000 compressed abdomen CT images.

II. MATERIALS AND METHODS

II.A. Study overview

Our institutional review board approved this study and waived informed patient consent. A total of 360 abdomen CT images were obtained with six different section thicknesses (STs) ranging from 0.67 to 5 mm. Using a JPEG2000 algorithm, each image was compressed reversibly (n = 48, as negative control) or irreversibly (n = 312) to one of different compression ratios (CRs). Five radiologists independently determined whether the original and compressed images were distinguishable or indistinguishable.

The 312 irreversibly compressed images were divided randomly into two subsets: training (n = 156) and testing (n = 156) sets. Using the training set, MLR and ANN models were constructed regarding the DICOM header information as independent variables and pooled readers' responses as dependent variable, after determining an optimal subset of independent variables by backward stepwise selection in a four-fold cross-validation scheme. The constructed MLR and ANN models were evaluated on the testing set (a) by using the receiver-operating-characteristic (ROC) analysis regarding the radiologists' pooled responses as the reference standard and (b) by measuring the Spearman rank correlation between the model prediction and the number of radiologists who rated the compressed image as distinguishable from its original (Fig. 1).

II.B. Image acquisition and selection

By retrospectively reviewing contrast-enhanced abdomen CT scans of adults, obtained from 64-detector-row CT scanners (Brilliance; Philips Healthcare, Cleveland, OH) in Seoul National University Bundang Hospital in late 2009, an abdominal radiologist (K.H.L) with nine years of clinical experience, who did not participate in the human visual analysis, compiled 360 scans (187 males and 173 females; age range, 17–96 years) containing "common abnormalities". All scan parameters followed clinical scan protocols of our hospital. Automatic tube current modulation technique was used which maintains approximately the same preferred level of image noise across patients (patient-to-patient modulation; Automatic Current Selection [ACS], Philips Healthcare) and across images along the craniocaudal axis within an acquisition (z-axis modulation [Z-DOM]; Philips Healthcare). The reference tube current-time product was set to 260 mAs. Other scan parameters are provided in Table I.

The 360 scans were equally divided into six groups (60 scans per group) in random, and the 60 scans in each group were then reconstructed into one of six different *ST*s

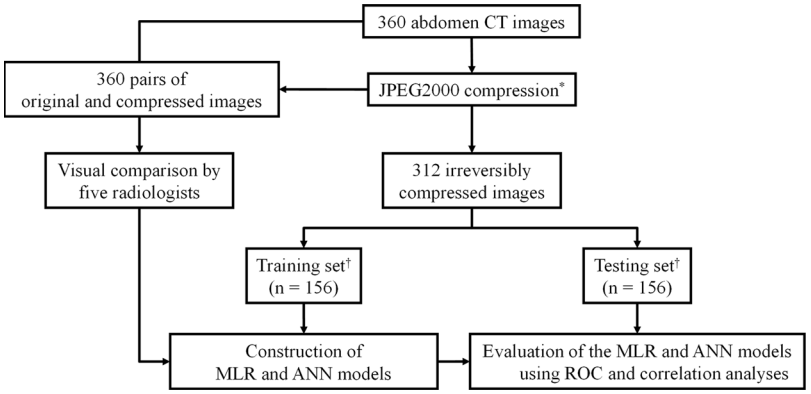

FIG. 1. Study design. *Randomly selected 48 images were reversibly compressed to serve as negative control in the human visual comparison. The remaining 312 images were irreversibly compressed to one of different compression ratios. †The 312 irreversibly compressed images were divided randomly into two subsets: training (n = 156) and testing (n = 156) sets. ANN: artificial neural network; CT: computed tomography; JPEG: Joint Photographic Experts Group; MLR: multiple logistic regression; ROC: receiver-operating-characteristic.

TABLE I. CT imaging parameters.

Parameter	Value
Contrast enhancement	
Contrast material	2 ml/kg (Ultravist 370; Schering, Berlin, Germany)
Flow rate	3 ml/s
Scanning	
Detector collimation	0.625 × 64 mm
Gantry rotation time	0.42 s
Tube potential	120 kVp
Pitch	0.70–1.17
Reference tube current-time product[a]	260 mAs
Effective tube current-time product[a]	138.0 ± 36.5 mAs
Effective radiation dose[a]	7.5 ± 1.9 mSv
Reconstruction	
Section thickness	0.67, 1.0, 2.0, 3.0, 4.0, or 5.0 mm
Reconstruction interval	50% of section thickness
Field of view[b)]	175–380 mm
Matrix size	512 × 512
Filter type	Standard soft tissue

[a]Automatic tube current modulation was used. Estimated by multiplying the dose-length product measured on the CT console by a conversion factor ($0.017\ mSv\ mGy^{-1}\ cm^{-1}$) (Ref. 43).
[b]The field of view was set for each patient to match the maximum transverse diameter of the abdomen with the image size.

(0.67, 1, 2, 3, 4, and 5 mm) with a reconstruction interval of 50% overlap. Detailed reconstruction parameters are also summarized in Table I. From each reconstructed image dataset, the same radiologist selected a single section that most clearly represented the pathology. Therefore, the final study sample was a set of 360 single-section images.

Using different *ST*s was to reproduce clinical practice where it is common to adjust the *ST* depending on the clinical purposes when reconstructing transverse image datasets from multi-helical projection data. Images of a smaller *ST* provide better through-plane spatial resolution, whereas those of a larger *ST* (e.g., 5 mm) are advantageous in low-contrast resolution as they contain less image noise.[16–20] Some radiologists even routinely reconstruct images with two or more different *ST*s from a single multihelical projection data.[4,21,22] The range of *ST*s tested in our study was chosen based on our clinical imaging protocol. In addition, using the different *ST*s would introduce heterogeneity in the noise level into the tested images. The heterogeneity is considered important in assessing the robustness of an image fidelity metric in predicting the fidelity of compressed CT images, as compression artifacts are significantly affected by the image noise level.[11,14]

II.C. Image compression

Each image had a bit depth of 12 bits/pixel and each pixel was packed on a two-byte boundary with four padding bits. We adjusted the compression level in terms of *CR* because the *CR* is the *de facto* standard index of the compression level of medical images.[8,23,24] The *CR* was defined as the ratio of the original image file size (16 bits/pixel) to the compressed size (bits/pixel).[25]

Of the 60 images with each *ST*, eight images were randomly selected and then reversibly compressed. The reversible compression data were used as negative control for the subsequent human visual analysis. The remaining 52 images were irreversibly compressed to one of different *CR*s. Images with 0.67- or 1-mm *ST*s were compressed to *CR*s which were randomly chosen between 4:1 and 8:1. Images with 2-, 3-, 4-, or 5-mm *ST*s were compressed to *CR*s between 6:1 and 10:1. These ranges of *CR*s were determined to be around the previously reported optimal *CR*s for abdomen CT images according to the *ST*s.[5,9,14,26,27] We compressed each image to a single *CR* rather than to multiple *CR*s to avoid a possible clustering effect in the subsequent statistical analyses. A JPEG2000 algorithm (Accusoft-Pegasus Imaging Co., Tampa, FL) with default encoder settings was used.[5,14] Minute differences in actual *CR*s from the nominal *CR*s, ranging from 0.00 to 0.07, were deemed unimportant.

II.D. Human observers and viewing conditions

Five abdominal radiologists with 10, 9, 8, 5, and 4 years of clinical experience participated in the visual analysis. Each of the 360 compressed (and then decompressed) images was paired with its original. The 360 image pairs were randomly assigned to six reading sessions. The order of the reading sessions changed for each reader. Sessions were separated by a minimum of 4 h to minimize readers' fatigue.

Images were displayed using viewing software in a monochrome monitor calibrated according to the DICOM part 14 (Grayscale Standard Display Function).[28] Since the reading distance would affect the readers' sensitivity to the compression artifacts,[29] the reading distance was limited to a range of the readers' habitual clinical reading distances (35–75 cm). Details of the display system and viewing conditions are summarized in Table II. The window level and width were set as 20 and 400 HU, respectively, which were the default window settings for abdomen CT images in our clinical practice.

II.E. Image discrimination task

Each image pair was alternately displayed on a single monitor while the order of the original and compressed

TABLE II. Display system and viewing conditions in human visual analysis.

Display system	
Display resolution	1536 × 2048 pixels
Display size	31.8 × 42.3 cm
Image resolution	1483 × 1483 pixels (stretched using bilinear interpolation)
Viewing software	PiViewSTAR (version 5.0.7, SmartPACS, Phillipsburg, NJ)
Luminance	1.3–398.3 cd/m^2
Viewing conditions	
Ambient room light	30 lux
Reading distance	35–75 cm
Window setting	width, 400 HU; level, 20 HU; not adjustable
Magnification	not allowed
Reading time	not constrained

images was randomized. The five readers, blind to the *CR*s, independently determined whether the two images were distinguishable or indistinguishable. This image comparison method is known to be very sensitive to image difference, and therefore, provides a conservative standpoint on image fidelity.[29,30]

Before the formal visual analysis, the readers were instructed on how to perform the visual analysis with ten example image pairs which were not included in the test dataset. The readers also had a chance to perform the analysis with another five example image pairs by themselves so that they could become familiar with the visual analysis.

Although we selected images that contained abnormalities, the readers were asked not to confine their analysis to the pathologic findings. Instead, they were asked to examine an entire image to find any image difference, paying attention to structural details, particularly small vessels and organ edges, and to the texture of solid organs and soft tissues.

To obtain an average response, the responses of the five readers were pooled. If three or more readers responded as distinguishable, the pooled response was considered as distinguishable (positive response); otherwise, the pooled response was considered as indistinguishable (negative response). These pooled data were used in the subsequent statistical analysis.

II.F. Multiple logistic regression and artificial neural network

To predict the image fidelity of compressed CT images using the semantic information from the DICOM header, we used the MLR and ANN regarding the header information as independent variables and the pooled readers' responses as dependent variable. These two algorithms have been widely used in constructing the prediction models in the medical fields.[31] A software (MATLAB R2009a, Mathworks, Nattick, Mass) was used to implement the models.

The logistic regression is a method which classifies a given data into one of two classes using the information from the independent variables.[32] It is based on an assumption that the logarithm of the odds of belonging to one class (i.e., logit) is a linear function of the independent variables used for classification, i.e.,

$$\begin{aligned}\text{logit}(p) &= \log\left(\frac{p}{1-p}\right) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_N X_N, \end{aligned} \quad (1)$$

where p is the probability of belonging to one class (distinguishable or indistinguishable), X_1, X_2, … X_N are independent variables from DICOM header, and β_1, β_2, … β_N are regression coefficients to be estimated based on the data. Among a variety of coefficient estimation methods, we applied the maximum likelihood which is the most widely used one.[32]

The ANN is a flexible mathematical structure which is capable of identifying complex nonlinear relationships between input and output data sets. There exist many forms of ANNs with a variety of layers numbers, node numbers, and training algorithms. The general description of the ANNs can be found elsewhere.[33] Our ANN was three-layer (one input layer, one output layer, and one hidden layer), feed-forward, and error back-propagation model, which is a method widely used in many previous studies.[34,35] The number of nodes in the hidden layer was three which was chosen empirically. A nonlinear sigmoid function was used as a transfer function for each of the neurons in the hidden and output layer. The output value of the ANN model (*ANN output*) was in the range of 0–1, indicating the likelihood that the compressed image is distinguishable from its original.

II.G. Selection of independent variables from the DICOM header

Among the DICOM header information, we selected the DICOM tags considered to affect the compression artifacts as the candidates for the independent variables. In addition to *CR* (DICOM tag number: 0028, 2112), we selected CT scanning and reconstruction parameters determining the level of image noise which is known to be an important factor affecting the compression artifacts.[8,11,14] First, we selected relevant DICOM tags describing the CT scanning parameters as follows. In general, noise level in CT images depends on the number of x-ray photons reaching the detector, which again depends on tube current, exposure time, pitch, and tube potential.[36] For the independent variable, we selected effective tube current-time product (tag number: 0018, 9332) as it accounts for tube current, exposure time, and pitch as follows: effective tube current-time product (mAs) = tube current [mA] × exposure time (s)/pitch.[36] Tube potential was not included in the independent variables because it was constant (i.e., 120 kVp) in our dataset. Second, we considered reconstruction parameters affecting image noise including *ST*, field of view (FOV), and reconstruction filter type.[36] Of those parameters, only *ST* (tag number: 0018, 0050) and FOV (tag number: 0018, 0090) were included in the independent variables because the reconstruction filter type was constant (i.e., standard soft tissue) in our dataset. Other factors known to affect the compression artifacts, such as body parts and compression algorithms,[1] were not included as those factors were also constant in our dataset.

II.H. Statistical analysis

The sample size of 360 images (156 images for training set, 156 images for testing set, and 48 images for negative control) was determined as follows. In the previous studies testing the image fidelity metrics in measuring the fidelity of compressed CT images, the areas under the ROC curve (AUCs) for the metrics ranged from 0.87 to 0.99.[5–9] We regarded 0.90 as an acceptable AUC value for an MLR model to be feasible in an adaptive compression. With this assumption, the sample size of a testing set was estimated to be 156 to construct a 95% confidence interval (CI) of the AUC with the width of no greater than 0.10. The sample size

calculation was based on Hanley and McNeil's papers.[37,38] In addition to the testing set, a separate sample of the same size[39] was required for the training set (n = 156). Finally, 48 reversibly compressed images (approximately 13% of the total sample size) were added into the final data sample as the negative control in the human visual analysis.

Interobserver agreement between the five readers was measured by using kappa statistic. To construct and validate the MLR and ANN models, we applied an analysis scheme widely used in the machine learning field.[40] The 312 irreversibly compressed images were divided randomly into two groups: 156 images for the training set and 156 images for the testing set. With the training set, an optimal subset of independent variables was determined using MLR by performing backward stepwise selection using the likelihood-ratio statistic ($p = 0.05$ for entry and $p = 0.10$ for removal) as a selection criterion.[40] At each step in the backward stepwise selection, four-fold cross-validation scheme was used. With the determined subset of independent variables, the final models were constructed by fitting the MLR and ANN to the entire training set.

The MLR and ANN models were then validated using the testing set. For each compressed image in the testing set, the prediction scores for the likelihood that the compressed image is distinguishable from its original were calculated from the MLR [i.e., logit(p)] and ANN (i.e., *ANN output*) models. Using these prediction scores, the performance of each model was evaluated in two ways. First, ROC analysis was performed regarding the readers' pooled responses as the reference standard. On the ROC curve, we identified cutoff values yielding the highest accuracy and 90% sensitivity. Using the cutoff values, the sensitivity, specificity, and accuracy were measured. Second, the Spearman rank correlation coefficient was measured between the prediction scores and the number of readers with positive response. A p-value of <0.05 was considered to indicate a statistical significance. Statistical software (MEDCALC, version 9.4.2, Mariakerke, Belgium and StatsDirect, version 2.5.6, Cheshire, United Kingdom) were used.

III. RESULT

III.A. Interobserver agreement

The overall kappa statistic between the five readers was 0.55 (95% CI, 0.52, 0.58; $p < 0.001$). All of the five readers determined that all of the 48 reversibly compressed images were indistinguishable from the originals. For 174 (55.8%) of the 312 irreversibly compressed images, all of the readers gave concordant responses as distinguishable (110 images) or as indistinguishable (64 images). For the remaining 138 irreversibly compressed images (44.2%), there was disagreement among the readers.

III.B. Overall response pattern

For each *ST*, each reader tended to respond that more compressed images were distinguishable as the *CR* increased. The pooled responses also tended to be positive more frequently as the *CR* increased (Fig. 2).

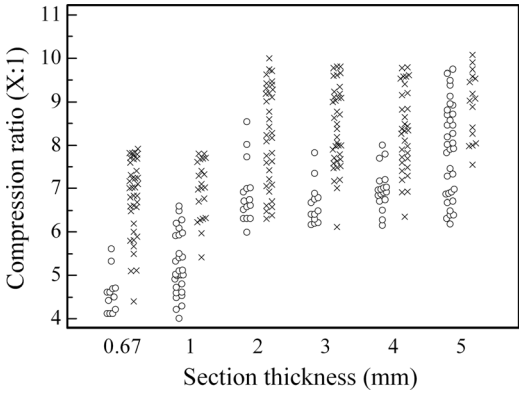

FIG. 2. Scatter plot of the radiologists' pooled responses at each section thickness and each compression ratio. The symbols ○ and × represent negative (indistinguishable image pairs) and positive responses (distinguishable image pairs), respectively.

III.C. Optimal variable selection and training of the models

According to the results of backward stepwise selection, the *ST* and *CR* were determined as the optimal subset of independent variables while the remaining variables—effective tube current-time product and FOV—were excluded. The regression coefficients (i.e., β) of the MLR model for the *ST* and *CR* were −1.15 (95% CI, −1.62, −0.67) and 1.78 (1.21, 2.34), respectively. The negative value of β for the *ST* implied that the likelihood that the compressed image is distinguishable from its original would decrease as the *ST* increases. The positive value of β for the *CR* implied that the likelihood would increase as the *CR* increases. The equation of the MLR model constructed on the training set was determined as follows:

$$\text{logit}(p) = -8.93 - 1.15 \times ST + 1.78 \times CR. \quad (2)$$

The connection weights between the nodes of the ANN model for the *ST* and *CR* were provided in Fig. 3.

III.D. Performance of the trained models (Table III)

Regarding the readers' pooled responses as the reference standard, the AUCs for the MLR and ANN predictions were 0.91 (95% CI, 0.86, 0.95) and 0.92 (0.87, 0.96), respectively, for the testing set. The results for the sensitivity, specificity, and accuracy at the cutoff values yielding the highest accuracy and 90% sensitivity are provided in Table III. The correlation coefficients of the MLR and ANN predictions with the number of readers with positive response were 0.76 (95% CI, 0.69, 0.82; $p < 0.001$) and 0.78 (0.71, 0.83, $p < 0.001$), respectively (Fig. 4). The prediction results of the models with all four variables are additionally provided in Table IV.

IV. DISCUSSION

In this study, we proposed MLR and ANN models which predicts the fidelity of JPEG2000 compressed CT images

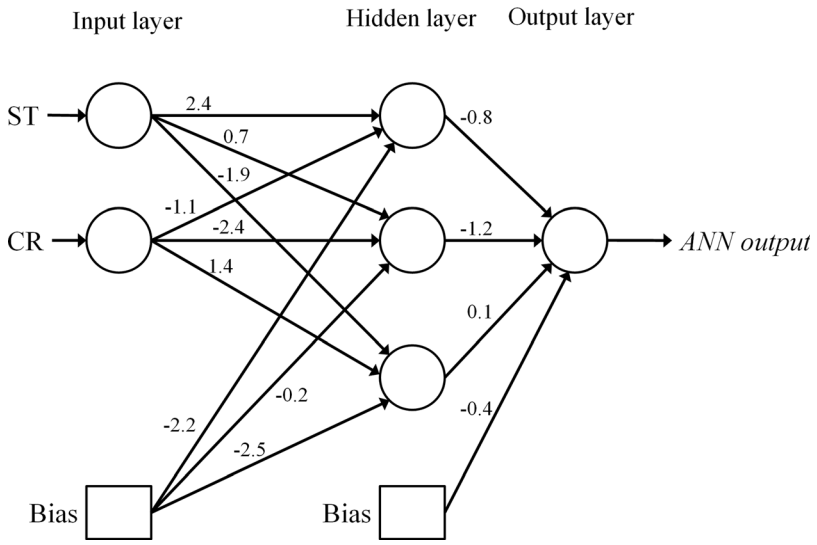

FIG. 3. Diagram of the artificial neural network constructed with optimal independent variables using training set. Each circle represents a node, while each line and the figure on the line represent a connection between the nodes and a connection weight, respectively. The nodes of network are arranged in three layers (input, hidden, output). *CR*: compression ratio; *ST*: section thickness.

using the DICOM header information. The AUCs of the MLR and ANN predictions with *CR* and *ST* as independent variables were 0.91 and 0.92, respectively, regarding the radiologists' pooled responses as the reference standards. The correlation coefficients of the MLR and ANN predictions with the number of radiologists with positive response (rating a compressed image as distinguishable from its original) were 0.76 and 0.78, respectively. These results imply that the proposed MLR and ANN models can be potentially used in an adaptive compression of abdomen CT images. To our knowledge, this study is the first attempt to propose a scheme of adaptive compression using solely semantic information available in the DICOM header.

Previous studies evaluated several image fidelity metrics, such as peak signal-to-noise ratio,[5–7,9] high-dynamic range visual difference predictor,[5–7,9] and multiscale structural similarity,[9] in measuring the fidelity of compressed CT images. The principal goal of these studies was to introduce the image fidelity metrics into an adaptive compression. However, those metrics have a couple of limitations from a practical viewpoint. First, they analyze the *image characteristics* of a compressed image in comparison to its original, requiring a substantial computational expense.[10] Second, to achieve an optimal compression level for an image, it is necessary to iteratively compress the image to multiple compression levels and to measure the image fidelity at each compression level until the measured fidelity reaches a predefined threshold.[5,6,8]

TABLE III. ROC analysis and correlation results for MLR and ANN models constructed with the optimal subset of independent variables.

	MLR	ANN
ROC analysis		
AUC value	0.91 (0.86, 0.95)	0.92 (0.87, 0.96)
At cutoff yielding the highest accuracy	0.92	0.58
Sensitivity (%)	78.4 (68.4, 86.5)	83.0 (73.4, 90.1)
Specificity (%)	92.7 (83.7, 97.5)	91.2 (81.8, 96.7)
Accuracy (%)	84.0 (77.1, 89.2)	86.5 (79.9, 91.3)
At cutoff yielding near 90% sensitivity	−0.27	0.41
Sensitivity (%)	89.8 (81.5, 95.2)	89.8 (81.5, 95.2)
Specificity (%)	69.1 (56.7, 79.8)	82.4 (71.2, 90.5)
Accuracy (%)	80.8 (73.5, 86.5)	86.5 (79.9, 91.3)
Correlation analysis		
Correlation coefficient	0.76 (0.69, 0.82)	0.78 (0.71, 0.83)
p–value	<0.001	<0.001

Note—Data in parentheses are the 95% CIs. ANN: artificial neural network; MLR: multiple logistic regression; ROC: receiver-operating-characteristic; AUC: area under ROC curve.

Although further comparative validation is needed, the proposed MLR and ANN models should be intrinsically more advantageous over the previous image fidelity metrics in terms of computing time. First, estimating the image fidelity involves purely *semantic* information from the DICOM header, and therefore, there is no need of analyzing the image characteristics. Second, without requiring the iterative compression and fidelity measurement at multiple compression levels, the optimal compression level (i.e., *CR*) can be calculated directly from the MLR model as follows:

$$CR = \frac{T + 1.15 \times ST + 8.93}{1.78}, \tag{3}$$

where T is the target threshold for logit(p) to achieve optimal image fidelity.

For an image with the *ST* of 5 mm, for example, the *CR* would be determined as 8.76 and 8.10 from the equation (3) if T is set as the cutoff yielding the highest accuracy (i.e., 0.92) and 90% sensitivity (i.e., −0.27) in our study, respectively. In case of an image with the *ST* of 0.67 mm, the *CR* would be determined as 5.97 and 5.30 if T is set as the cutoff

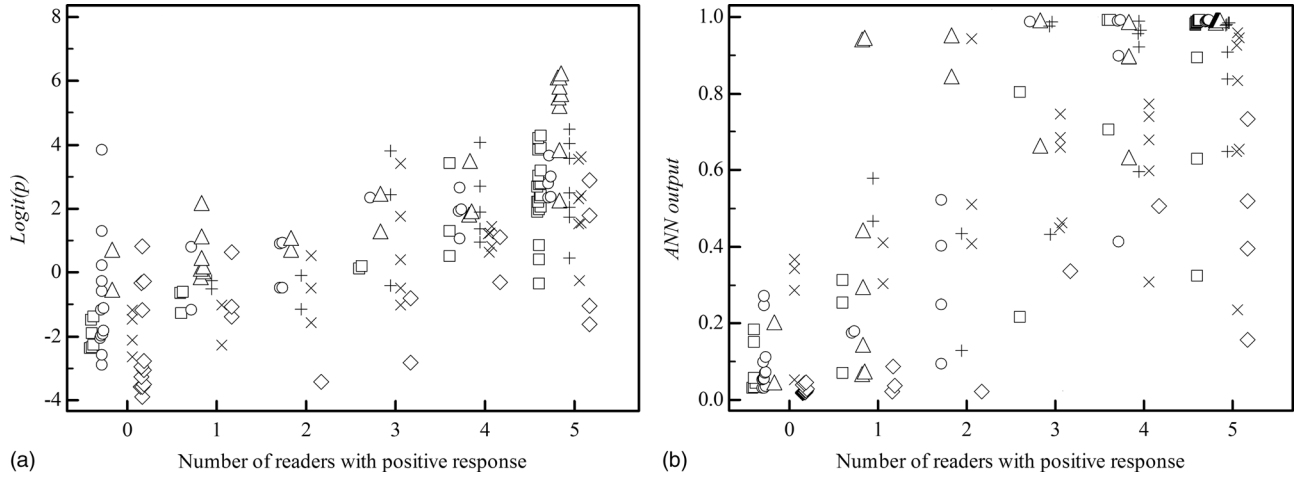

FIG. 4. Correlation between the number of readers with positive response (rating a compressed image as distinguishable from its original) and the prediction scores of the multiple logistic regression (a) and artificial neural network (b) models for the testing set. The symbols □, ○, Δ, +, ×, and ◇ represent the data with the section thickness of 0.67, 1, 2, 3, 4, and 5 mm, respectively. The Spearman correlation coefficients for multiple logistic regression and artificial neural network models were 0.76 (95% CI, 0.69, 0.82) and 0.78 (0.71, 0.83, $p < 0.001$), respectively.

yielding the highest accuracy and 90% sensitivity, respectively. The determined *CR*s are similar to the previous study results[14] based on the human experiments that the optimal *CR* for abdomen CT images with the *ST*s of 5 and 0.67 mm would be between 6:1 and 10:1 and below 6:1, respectively. However, it should be noted that such calculation would not be feasible for the ANN model, which is one of the disadvantages of ANN.[41]

Our results corroborate previous studies[11,14] in that the *ST* plays an important role in determining the fidelity of compressed images. Interestingly, the effective tube current-time product and FOV were not proven to be significant factors predicting the fidelity of compressed CT images, which may be considered a counterintuitive result as those two factors are theoretically related to image noise and thus to the compression artifacts. This result can be attributed to the fact that our image dataset was acquired using an automatic tube current modulation technique, which is currently adopted as a part of standard-of-practice.[42] As the tube current was modulated, the image noise level was maintained approximately constant across patients and across images regardless of the effective tube current-time product. Likewise, as the FOV was determined according to the patient body size, the image noise was also maintained approximately constant across patients, independent of the FOV. Consequently, the effects of the effective tube current-time product and FOV on the compression artifacts were not pronounced in our homogeneous study sample.

TABLE IV. ROC analysis and correlation results of the models constructed with all four independent variables (i.e., the compression ratio, effective tube current-time product, section thickness, and field of view).

	MLR	ANN
ROC analysis		
AUC value	0.91 (0.85, 0.95)	0.92 (0.87, 0.96)
At cutoff yielding the highest accuracy	1.18	0.80
Sensitivity (%)	75.0 (64.6, 83.6)	83.0 (73.4, 90.1)
Specificity (%)	97.1 (89.8, 99.6)	92.7 (83.7, 97.6)
Accuracy (%)	84.6 (77.8, 89.7)	87.2 (80.7, 91.8)
At cutoff yielding near 90% sensitivity	−0.07	0.58
Sensitivity (%)	89.8 (81.5, 95.2)	90.9 (82.9, 96.0)
Specificity (%)	72.1 (59.9, 82.3)	77.9 (66.2, 87.1)
Accuracy (%)	82.1 (74.9, 87.6)	86.5 (79.9, 91.3)
Correlation analysis		
Correlation coefficient	0.75 (0.67, 0.81)	0.77 (0.70, 0.83)
p–value	<0.001	<0.001

Note—Data in parentheses are the 95% CIs. ANN: artificial neural network; MLR: multiple logistic regression; ROC: receiver-operating-characteristic; AUC: area under ROC curve.

It should be noted that our MLR and ANN models cannot be directly generalized to other medical images. The tested images in our study included only single body part (i.e., abdomen), single compression algorithm (i.e., JPEG2000), single modality (i.e., CT), and single scanner type. Furthermore, CT imaging parameters were homogeneous. Had we used a dataset with different body parts, compression algorithms, modalities, scanner types, and/or imaging parameters such as reference tube current-time product, of which information are also available in DICOM header, different MLR and ANN models would have been obtained. It should be also noted that more sophisticated classification methods exist as alternatives to MLR or ANN, such as support vector machine and AdaBoost.[40] Therefore, we do not claim that our models can be used as a global compression guideline for all medical images. Instead, the implication of our study is rather proposing the scheme of adaptive compression by using solely semantic information available in the DICOM header.

In conclusion, the MLR and ANN models constructed using the DICOM header information offers promise in predicting the fidelity of JPEG2000 compressed abdomen CT images. Our results demonstrate the feasibility of adaptive compression by using solely semantic information available in the DICOM header, without the need of image analysis.

ACKNOWLEDGMENTS

This work was supported by Mid-career Researcher Program through National Research Foundation of Korea funded by the MEST (No. 2011-0012117).

APPENDIX

ROC analysis and correlation results of the models constructed with all four independent variables (i.e., the compression ratio, effective tube current-time product, section thickness, and field of view).

[a]Electronic mail: kholeemail@gmail.com

[1]D. Koff, P. Bak, P. Brownrigg, D. Hosseinzadeh, A. Khademi, A. Kiss, L. Lepanto, T. Michalak, H. Shulman, and A. Volkening, "Pan-Canadian evaluation of irreversible compression ratios ("Lossy" compression) for development of national guidelines," J. Digit Imaging **22**, 569–578 (2009).

[2]K. H. Lee, Y. H. Kim, B. H. Kim, K. J. Kim, T. J. Kim, H. J. Kim, and S. Hahn, "Irreversible JPEG 2000 compression of abdominal CT for primary interpretation: Assessment of visually lossless threshold," Eur. Radiol. **17**, 1529–1534 (2007).

[3]G. D. Rubin, "Data explosion: The challenge of multidetector-row CT," Eur. J. Radiol. **36**, 74–80 (2000).

[4]K. H. Lee, H. J. Lee, J. H. Kim, H. S. Kang, K. W. Lee, H. Hong, H. J. Chin, and K. S. Ha, "Managing the CT data explosion: Initial experiences of archiving volumetric datasets in a mini-PACS," J. Digit Imaging **18**, 188–195 (2005).

[5]B. Kim, K. H. Lee, K. J. Kim, R. Mantiuk, V. Bajpai, T. J. Kim, Y. H. Kim, C. J. Yoon, and S. Hahn, "Prediction of perceptible artifacts in JPEG2000 compressed abdomen CT images using a perceptual image quality metric," Acad. Radiol. **15**, 314–325 (2008).

[6]B. Kim, K. H. Lee, K. J. Kim, R. Mantiuk, S. Hahn, T. J. Kim, and Y. H. Kim, "Prediction of perceptible artifacts in JPEG2000 compressed chest CT images using mathematical and perceptual quality metrics," Am. J. Roentgenol. **190**, 328–334 (2008).

[7]B. Kim, K. H. Lee, K. J. Kim, R. Mantiuk, H. R. Kim, and Y. H. Kim, "Artifacts in slab average-intensity-projection images reformatted from JPEG 2000 compressed thin-section abdominal CT data sets," Am. J. Roentgenol. **190**, 342–350 (2008).

[8]K. J. Kim, B. Kim, K. H. Lee, R. Mantiuk, H. S. Kang, J. Seo, S. Y. Kim, and Y. H. Kim, "Objective index of image fidelity for JPEG2000 compressed body CT images," Med. Phys. **36**, 3218–3226 (2009).

[9]K. J. Kim, B. Kim, R. Mantiuk, T. Richter, H. Lee, H. S. Kang, J. Seo, and K. H. Lee, "A comparison of three image fidelity metrics of different computational principles for JPEG2000 compressed abdomen CT images," IEEE Trans. Med. Imaging **29**, 1496–1503 (2010).

[10]Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, 1st ed. (Morgan & Claypool, San Rafael, CA, 2006).

[11]V. Bajpai, K. H. Lee, B. Kim, K. J. Kim, T. J. Kim, Y. H. Kim, and H. S. Kang, "The difference of compression artifacts between thin- and thick-section lung CT images," Am. J. Roentgenol. **191**, 38–43 (2008).

[12]K. J. Kim, B. Kim, K. H. Lee, T. J. Kim, R. Mantiuk, H. S. Kang, and Y. H. Kim, "Regional difference in compression artifacts in low-dose chest CT Images: Effects of mathematical and perceptual factors," Am. J. Roentgenol. **191**, 30–37 (2008).

[13]T. J. Kim, K. W. Lee, B. Kim, K. J. Kim, E. J. Chun, V. Bajpai, Y. H. Kim, S. Hahn, and K. H. Lee, "Regional variance of visually lossless threshold in compressed chest CT Images: Lung versus mediastinum and chest wall," Eur. J. Radiol. **69**, 483–488 (2008).

[14]H. S. Woo, K. J. Kim, T. J. Kim, S. Hahn, B. H. Kim, Y. H. Kim, C. J. Yoon, and K. H. Lee, "JPEG 2000 compression of abdominal CT: Difference in compression tolerance between thin- and thick-section images," Am. J. Roentgenol. **189**, 535–541 (2007).

[15]R. R. William and R. P. David, "Extracting data from a DICOM file," Med. Phys. **32**, 1537–1541 (2005).

[16]D. D. Cody, "AAPM/RSNA physics tutorial for residents: Topics in CT. Image processing in CT," Radiographics **22**, 1255–1268 (2002).

[17]S. Kawata, T. Murakami, T. Kim, M. Hori, M. P. Federle, S. Kumano, E. Sugihara, S. Makino, H. Nakamura, and M. Kudo, "Multidetector CT: Diagnostic impact of slice thickness on detection of hypervascular hepatocellular carcinoma.," Am. J. Roentgenol. **179**, 61–66 (2002).

[18]K. H. Lee, Y. H. Kim, S. Hahn, K. W. Lee, T. J. Kim, S. B. Kang, and J. H. Shin, "Computed tomography diagnosis of acute appendicitis: Advantages of reviewing thin-section datasets using sliding slab average intensity projection technique," Invest. Radiol. **41**, 579–585 (2006).

[19]F. R. Verdun, A. Denys, J. Valley, P. Schnyder, and R. A. Meuli, "Detection of low-contrast objects: Experimental comparison of single- and multidetector row CT with a phantom," Radiology **223**, 426–431 (2002).

[20]M. Prokop, "Multislice CT: Technical principles and future trends," Eur. Radiol. **13**(Suppl 5), M3–M13 (2003).

[21]M. Prokop, "Principles of CT, spiral CT, and multislice CT," in *Spiral and Multislice Computed Tomography of the Body*, edited by M. Prokop and M. Galanski (Georg Tieme Verlag, Stuttgart, Germany, 2003), pp. 35–37.

[22]K. H. Lee, H. Hong, S. Hahn, B. H. Kim, K. W. Lee, H. J. Lee, B. J. Park, D. K. Jeong, and Y. H. Kim, "Summation or axial slab average intensity projection of abdominal thin-section computed tomography datasets: Can they substitute for the primary reconstruction from raw projection data?," J. Digit Imaging **4**, 422–432 (2008).

[23]ACR Technical Standard for Teleradiology, http://www.acr.org. Last accessed May 2010.

[24]Guidance for the submission of premarket notifications for medical image management devices. http://www.fda.gov/cdrh/ode/guidance/416.html. Last accessed May 2010.

[25]K. J. Kim, B. Kim, S. W. Choi, Y. H. Kim, S. Hahn, T. J. Kim, S. J. Cha, V. Bajpai, and K. H. Lee, "Definition of compression ratio: Difference between two commercial JPEG2000 program libraries," Telemed J E Health **14**, 350–354 (2008).

[26]H. Lee, K. H. Lee, K. J. Kim, S. Park, J. Seo, Y. Shin, and B. Kim, "Advantage in image fidelity and additional computing time of JPEG2000 3D in comparison to JPEG2000 in compressing abdomen CT image datasets of different section thicknesses," Med. Phys. **37**, 4238–4248 (2010).

[27]B. Kim, K. H. Lee, K. J. Kim, T. Richter, H. S. Kang, S. Y. Kim, Y. H. Kim, and J. Seo, "JPEG2000 3D compression vs. 2D compression: An assessment of artifact amount and computing time in compressing thin-section abdomen CT images," Med. Phys. **36**, 835–844 (2009).

[28]Digital Imaging and Communications in Medicine (DICOM). Part 14: Gray scale standard display function. http://medical.nema.org/dicom/2004/04_14pu.pdf. Last accessed May 2010.

[29]R. M. Slone, E. Muka, and T. K. Pilgram, "Irreversible JPEG compression of digital chest radiographs for primary interpretation: Assessment of visually lossless threshold," Radiology **228**, 425–429 (2003).

[30]B. Kim, H. Lee, K. J. Kim, J. Seo, S. Park, Y. G. Shin, S. H. Kim, and K. H. Lee, "Comparison of three image comparison methods for the visual assessment of the image fidelity of compressed computed tomography images," Med. Phys. **38**, 836–844 (2011).

[31]S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," J. Biomed. Inf. **35**, 352–359 (2002).

[32]T. Ayer, J. Chhatwal, O. Alagoz, C. E. Kahn, Jr., R. W. Woods, and E. S. Burnside, "Informatics in radiology: Comparison of logistic regression and artificial neural network models in breast cancer risk estimation," Radiographics **30**, 13–22 (2010).

[33]S. S. Cross, R. F. Harrison, and R. L. Kennedy, "Introduction to neural networks," Lancet **346**, 1075–1079 (1995).

[34]J. L. Jesneck, J. Y. Lo, and J. A. Baker, "Breast mass lesions: Computer-aided diagnosis models with mammographic and sonographic descriptors," Radiology **244**, 390–398 (2007).

[35]S. H. Kim, J. M. Lee, J. H. Kim, K. G. Kim, J. K. Han, K. H. Lee, S. H. Park, N. J. Yi, K. S. Suh, S. K. An, Y. J. Kim, K. R. Son, H. S. Lee, and B. I. Choi, "Appropriateness of a donor liver with respect to macrosteatosis: Application of artificial neural networks to US images—Initial experience," Radiology **234**, 793–803 (2005).

[36]M. F. McNitt-Gray, "AAPM/RSNA Physics tutorial for residents: Topics in CT. Radiation dose in CT," Radiographics **22**, 1541–1553 (2002).

[37]B. J. McNeil and J. A. Hanley, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology **143**, 29–36 (1982).

[38]J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," Radiology **148**, 839–843 (1983).

[39]A. Dupuy and R. M. Simon, "Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting," J. Natl. Cancer Inst. **99**, 147–157 (2007).
[40]T. Hastie, R. Tibshirani, and J. Friedman, "Model assessment and selection," in *The Elements of Statistical Learning*, edited by T. Hastie, R. Tibshirani, and J. Friedman (Springer, New York, NY, 2001), pp. 193–224.
[41]J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes* 1," J. Clin. Epidemiol. **49**, 1225–1231 (1996).
[42]C. H. Lee, J. M. Goo, H. J. Ye, S. J. Ye, C. M. Park, E. J. Chun, and J. G. Im, "Radiation dose modulation techniques in the multidetector CT era: From basics to practice," Radiographics **28**, 1451–1459 (2008).
[43]G. Bongartz, S. J. Golding, A. G. Jurik, M. Leonardi, E. van Persijn van Meerten, R. Rodríguez, K. Schneider, A. Calzado, J. Geleijns, K. A. Jessen, W. Panzer, P. C. Shrimpton, and G. Tosi, "European guidelines for multislice computed tomography," http://msct.eu/CT_Quality_Criteria.htm. Last accessed May 2010.