

Automatic Quantification for Phantom-based Image Quality Assessment in Bone Spect: Computerized Self-classification of Detectability Using a Novel Index

Hajime ICHIKAWA

Toyohashi municipal hospital <https://orcid.org/0000-0001-5331-3391>

Kazunori Kawakami

Fujifilm Toyama Chemical

Masahisa Onoguchi (✉ onoguchi@staff.kanazawa-u.ac.jp)

Takayuki Shibutani

Kanazawa University

Kazuki Nagatake

Fujifilm Toyama Chemical

Tetsuo Hosoya

Fujifilm Toyama Chemical

Toshimune Ito

Saiseikai Yokohamashi Tobu Hospital

Toyohiro Kato

Toyohashi municipal hospital

Original research

Keywords: bone scintigraphy, single-photon emission computed tomography, automatic quantification, detectability equivalence volume, contrast-to-noise ratio

Posted Date: June 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-33205/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: We have previously developed a custom-design thoracic bone scintigraphy-specific phantom (hereafter referred to as “SIM² bone phantom”) to assess image quality in bone single-photon emission computed tomography (SPECT). This study aimed to develop an automatically assessment system for imaging technology in bone SPECT and to demonstrate the validity of this system.

Methods: We were able to generate more realistic phantom configurations of the thorax, spine, mediastinum, and lung. Four fillable spheres as bone metastases with diameters of 13, 17, 22, and 28 mm were inserted in the vertebral body. The newly developed software with statistical parametric mapping version 2 was automatically calculated for quantitative indexes (e.g., contrast-to-noise ratio, % coefficient of variance, % detectability equivalence volume, sharpness index). The repeatability and reproducibility of the contrast-to-noise ratio was assessed and compared between the automatic and manual methods. A detectability score (DS) was used to define the four observation types to score the hot spherical lesions which were classified with the quantitative indexes using decision tree analysis. The gold standard for DSs was independently classified by three expert board-certified nuclear medicine technologists using a 4-point classification (1, poor; 2, average; 3, adequate; 4, excellent); thereafter, a consensus was reached. To assess the validity of the DS, the DS classified with Hone Graph was compared to the visual evaluation by 11 technologists (using the same score).

Results: The repeatability and reproducibility of the software were shown to be excellent. Decision tree analysis produced seven terminal groups, and four quantitative indexes were used to classifying the DS. The automatically classified DSs exhibited an almost perfect agreement with the gold standard. The agreement between the Hone Graph and the experienced (n = 4), moderately experienced (n = 3), and inexperienced (n = 4) observers was almost perfect, substantial, and moderate, respectively.

Conclusions: The developed software could automatically classify the detectability of hot lesions in the SIM² bone phantom using the self-calculated quantitative indexes. These findings suggest that this novel software could provide a means to automatically perform analysis after data input that is both excellent in convenience and repeatability.

Background

Bone scintigraphy has been widely used to detect bone metastases in patients with cancer for several decades. The sensitivity for single-photon emission computed tomography (SPECT) imaging has been reported to be higher than that for planar imaging in bone metastases detection [1–3]. Recently, with the rapid development of imaging technology, such as applied iterative reconstruction or SPECT coupled with computed tomography (SPECT/CT), the necessity of bone SPECT has become a highlight in clinical practice [4]. Consequently, bone scintigraphy has gradually shifted from reliance on planar whole-body imaging with additional spot planar imaging to additional SPECT or SPECT/CT imaging [5]. In particular, spine SPECT can be useful since the spine is the most common site of skeletal metastases [2].

Assessment of image quality for bone SPECT, especially detectability, is important to better understand the variations in image quality with system performance and imaging technology. Although visual analysis is needed to assess the detectability of lesions, it has poor inter-observer agreement [6]. Rose et al. established that a contrast-to-noise ratio (CNR) greater than 5 was required to detect a small lesion; hereafter, referred to as Rose criteria [7]. Moreover, several studies in nuclear medicine have recently addressed the association between the CNR and the detectability of hot lesions [8–11]. However, it is known that Rose criteria [7] will change with either the lesion size or pixel size [12]. A few studies have reported that multiplying CNR by any measure of volume for lesion (e.g., square root of numbers of pixels), gave an alternative index that was relevant to detectability [8, 13].

Low accumulations may be undetectable due to the display, the whitish nuclear medicine images are generally displayed in inverted grayscale using 256 steps from white to black. Consequently, the detectability of hot lesions is affected the maximum intensity values in the entire slices, and for a lesion to be detectable, it will need to be over a certain count compared to the maximum count. In other words, a lesion will be no longer detectable when the counts of the lesion fall below a certain threshold of the maximum count, similar to areas that are excluded from the delineating region of interest (ROI) using a threshold value.

Imaging technology for bone SPECT has been poorly documented, partly due to the lack of an appropriate experimental instrument, such as anthropomorphic phantom. Even if it was possible to use a phantom, the image quality analysis would be extremely laborious, especially with regards to visual analysis. Despite this, bone SPECT has been widely usable for more than a decade, and previous studies using SPECT have mainly focused on diagnostic accuracy. In bone SPECT practice, immediate optimizing of imaging technology to detect bone metastases is of critical importance. We have previously developed a custom-design thoracic phantom to assess the imaging technology for bone SPECT [14]. The aim of the current study was to contribute to this growing field of research by developing a phantom-based automatic assessment system for the imaging technology of bone SPECT. In the current study, we focused on the relationship between detectability and the hot lesion volume at a certain threshold value of the maximum value in the SPECT image, and proposed a novel quantitative index for detectability. Furthermore, in order to elucidate the association between the quantitative index and visual detection, we performed decision tree analysis and subsequent verification.

Methods

Thoracic bone phantom

We have previously developed a custom-design thoracic bone scintigraphy-specific phantom (hereafter referred to as “SIM² bone phantom,” Kyoto kagaku co., Ltd., Japan) for bone SPECT to assess image quality (Fig. 1) [14]. A phantom should be anatomically realistic, functional, and simulate a patient thoracic vertebra with bone metastases. Hence, a fillable section within a spine required highly radioactivity concentrations within a vertebral body. Furthermore, all sections in the phantom must have

similar radioactivity and attenuation coefficients to the tissue. We were able to generate more realistic phantom configurations of the thorax, spine, mediastinum, lung, and bone metastases by constructing a SIM² bone phantom with an elliptical shape, a major axis of 310 mm, a minor axis of 210 mm, and a height of 320 mm. The vertebral body and reference section were constructed in a cylinder shape with a diameter of 36 mm, and height of 207 and 35 mm, respectively. Spherical lesion with diameters of 13, 17, 22, and 28 mm were inserted in the vertebral body. The spinous and transverse processes were T-shaped containers with a major axis of 90 mm, minor axis of 40 mm, and a height of 250 mm. The lungs were a semicircular column 170 mm in diameter, with a height of 300 mm. The phantom contained a ^{99m}Tc solution to simulate the mediastinum, bone section (vertebral body, reference section, spherical lesion, spinous and transverse processes) contained a bone equivalent solution of K₂HPO₄ (Molecular Imaging Labo Inc. Japan) and ^{99m}Tc, while the lungs contained lung-equivalent solid (tough lung phantom LD; Kyoto Kagaku Co., Ltd., Japan).

Study design

Our project developed a phantom-based automatic quantification software package (hereafter referred to as “Hone Graph”) to assess the imaging technology for bone SPECT images. The objectives were as follows: 1) To automatically measure quantitative indexes for the SIM² bone phantom; 2) to assess the repeatability and reproducibility of the indexes; 3) to automatically classify the detectability score (DS) with the indexes using decision tree analysis; and 4) to assess the validity of the DS compared to visual evaluation.

Algorithm for image analysis

The algorithm for image analysis is as follows:

STEP 1

Using statistical parametric mapping (SPM) version 2, an input image (SIM² bone phantom SPECT image) is rigidly registered with a reference image. This process is performed to automatically set the volume of interest (VOI) template for the quantitative index calculation in STEP 2. The reference image was created by rigidly registering the teaching SIM² bone phantom SPECT image with the SIM² bone phantom CT image. The input image after rigid registration is resampled to a matrix size of 360 × 360, 360 slices, with resulting voxel dimensions of 1 × 1 × 1 mm. The maximum count is then normalized with a pixel value of 30,000.

STEP 2

The VOI template for quantitative index calculation is automatically set in the input image after rigid registration (Fig. 2), and the quantitative indexes are calculated. This VOI template was created from the CT image of the SIM² bone phantom. The calculated indexes are as follows

- The percentage of detectability equivalence volume (%DEV): As an alternative measure of detectability, we defined %DEV as the percentage of volume within a respective spherical lesion when applying a 40% threshold of maximum counts in the SPECT image.
- The CNR was defined here as follows:

$$CNR = (mean\ counts_{SL} - mean\ counts_{NB}) / SD_{NB}$$

where $mean\ counts_{SL}$ and $counts_{NB}$ indicate the mean counts in the respective spherical lesion (13–28 mm) and normal bone, respectively. SD_{NB} indicates the standard deviation of the normal bone.

- The correcting CNR for %DEV ($CNR_{\%DEV}$) as a measure of detectability; multiplying the CNR by the %DEV gives the $CNR_{\%DEV}$.
- The percentage of coefficient of variation on normal bone and reference section ($\%CV_{NB}$ and $\%CV_{RS}$): The $\%CV_{NB}$ and $\%CV_{RS}$ were calculated by dividing the SD by the average count of VOIs within the normal bone and reference section, respectively.
- The target-to-normal bone ratio (TNR): The TNR was calculated as the count ratio of the respective spherical lesion (13–28 mm) to the normal bone.
- The sharpness index (SI): The SI is an index of spatial resolution used to eliminate the influence of background [15], which was taken as the maximal slope at 6 pixels by calculating the mean line-profile curve on 160 central transverse images of the spinous process (Fig. 2b).

STEP 3

Using the calculated quantitative indexes, automatically classify DS (see below).

STEP 4

Output the image quality assessment result (performance report) of the input image.

All SPECT data in this study were automatically measured for quantitative indexes using Hone Graph (version 1.0; Bone SPECT conference, Japan).

Radioactivity concentration

The amount of radioactivity usually administrated for bone scintigraphy (including SPECT) is 740 MBq; 50 kBq/mL of ^{99m}Tc phosphate compounds was taken as an appropriate amount of radioactivity to represent the radioactivity uptake in the normal vertebra approximately 3 hours after intravenous injection [16, 17]. The radioactivity concentration of bone metastases should be 4-fold greater than that of normal bone [18]. Therefore, the mediastinum, normal vertebra, and hot spherical lesions of the SIM² bone phantom were filled with 8 kBq/mL ^{99m}Tc solution and 50 and 200 kBq/mL ^{99m}Tc K₂HPO₄ solution

(Molecular Imaging Labo Inc.) of bone equivalent density, respectively [19]. It is not always necessary to use bone equivalent density solution to prepare the phantom.

Image acquisition and reconstruction

SPECT images were acquired with a dual-head gamma camera equipped with a low-energy high-resolution collimator (Symbia Intevo hybrid SPECT/CT scanner, Siemens Healthineers, USA). The SIM² bone phantom was positioned at the center of rotation and acquired for a total of 15 minutes. SPECT acquisitions were performed using the following parameters: Energy window, 15% at 140 keV; lower sub window, 15% for scatter correction; 128 × 128 matrix and 1.0 zoom (with a 4.8 × 4.8 mm² pixel); 90 views over 360 degrees; 60 seconds per rotation; and continuous mode through the elliptical orbit.

SPECT images were reconstructed for the 3-, 5-, 7-, 10-, and 15-minute timepoints after acquisition of the full 15 minutes using ordered subset expectation maximization with depth-dependent three-dimension resolution recovery (Flash3D) reconstruction; 2 subsets; 24 iterations; and 0 (i.e., without Gaussian post-filter), 4.8, and 9.6 mm full width at the half maximum (FWHM) Gaussian post-filter. Scatter correction, using lower scatter window subtraction (dual energy window method), and CT-based attenuation correction were performed, and uncorrected images were also reconstructed. Thereby, 30 SPECT image sets were reconstructed for the development and validation of Hone Graph. In addition, to evaluate of the repeatability and repeatability of Hone Graph, the 15-minute acquisition data were reconstructed with an approximately ± 50-mm X, Y, and Z-axis shift to the rotation center.

CT scanning, involving a 4.0-mm slice thickness, was performed at 130 kV, with dose modulation and a quality reference of 60 mAs. The CT data were reconstructed using the kernels of H08 SPECT AC for CT-based attenuation correction.

Image analysis

For the assessment of repeatability and reproducibility, the CNR was calculated at seven SPECT images with (6 images) or without (1 image) a pixel shift to the rotation center by four expert board-certified nuclear medicine technologists (BCNMTs). Automatic VOI-analysis using Hone Graph and manual ROI setting using ImageJ software (1.51 National Institutes of Health, Bethesda, MD, USA) were performed. Repeatability was evaluated as the mean %CV of seven SPECT images by four BCNMTs. Furthermore, reproducibility was evaluated as the mean %CV of four BCNMTs on seven SPECT images. Four BCNMTs were blinded to the information about the SPECT images that the seven SPECT images were reconstructed from same projection data.

Visual analysis

To define the gold standard for the DS, three expert BCNMTs with more than 15 years of clinical experience in nuclear medicine practice independently classified the detectability of hot spherical lesions based on the following criteria:

The DS of lesions is as follows:

1. Poor (not observed as a hot lesion)
2. Average (just acceptable)
3. Adequate (definitely observed but with diagnostically relevant noise and/or blurring)
4. Excellent (definitely observed without diagnostically relevant noise and blurring)

To assess the validity of the classified DS with Hone Graph, four experienced BCNMTs (≥ 10 years), three moderately experienced (3–10 years), and four inexperienced (< 3 years) nuclear medicine technologists (who were offered study specific instructions and opinions on image interpretation) classified the DS based on the aforementioned criteria. Hot lesions of 13, 17, and 22 mm in the vertebra of 30 SPECT images (i.e., 90 lesions) were observed in axial, sagittal, and coronal views, respectively, in inverted gray scale.

Statistical analysis

The CNR measurement reliability was assessed by intraclass correlation coefficient (ICC). Decision tree analysis was performed to analyze quantitative indexes equating the DS to the gold standard. A conditional inference tree was built using R software (version 3.5.3) [20]. Differences between two values of each DS group were tested for significance using Mann-Whitney U test. The categories of DS were compared with an unweighted kappa coefficient for assessment of agreements between the result of the Hone Graph and all observers. Agreement was graded as slight (kappa value < 0.20), fair (≥ 0.20 and < 0.40), moderate (≥ 0.40 and < 0.60), substantial (≥ 0.60 and < 0.80), and almost perfect (≥ 0.8 up to 1). Statistical analyses were performed using SPSS 23.

Results

The SIM² bone phantom was simulated a patient thoracic vertebra with various sizes of bone metastases, as demonstrated in Fig. 3. The transverse slices (Fig. 3a, c, d) correspond to a plane intersecting the 17-mm sphere, while the sagittal slice (Fig. 3b) corresponds to a plane intersecting the 13–28-mm sphere and the reference section insert.

Quantitative index

Figure 4 shows a brief version of performance report. The detectability of hot lesions and the noise characteristics of normal bone and background are indicated by color variations (red to green). Additional data about the quantitative indexes can be found in a detailed version of the performance report.

Repeatability and reproducibility of the quantitative index

Figure 5 shows the distribution of the CNRs by four experienced BCNMTs (A to D) using both the Hone Graph and the manual method. In the manual method, the CNRs were shown to be up to 3 or 4-fold different at most in terms of individual lesion sizes, whereas, Hone Graph yielded almost the same result in each lesion. The repeatability of the existing manual method was 39.6% (range, 18.3–55.8), whereas

that of Hone Graph was consistent across each of the four BCNMTs in the seven SPECT images (i.e., 0%). The reproducibility of the manual method and Hone Graph were 13.2% (range, 5.4–24.6) and 1.7% (1.2–2.2), respectively. The intra-observer reliability (ICC) of the manual method and Hone Graph for every hot lesion was 0.60 (95% confidence interval, 0.43–0.77) and 1.00 (95% confidence interval, 1.00–1.00), respectively.

Decision tree analysis

The results of decision tree analysis based on quantitative indexes and DS are depicted in Fig. 6. The analysis produced seven terminal groups, and four quantitative indexes were used to classify the DS. The first decision node at the top shows that %DEV was the index that was most associated with DS. The branches show that a cut-off value of 14.215% was best to stratify for DS 1 to 2 and DS 3 to 4. Subsequently, the decision nodes in both branches were further divided by CNR. No further split was performed in DS 1 and 2, or in lesions with a CNR less than 10.45. Those with a CNR more than 10.45 were further stratified by TNR, anew CNR and SI. In the comparison of the classifications by Hone Graph and the gold standard, the resulting percentage agreement (PA) and kappa coefficient were 99% and 0.98, respectively (Table 1). The software neither overestimated nor underestimated the classifications, and the correlation between Hone Graph and the gold standard was almost perfect.

Table 1
Frequency table showing classifications made by Hone Graph compared to the gold standard

| Hone Graph | Gold standard | | | | Total |
|------------|---------------|---|----|----|-------|
| | 1 | 2 | 3 | 4 | |
| 1 | 24 | 1 | 0 | 0 | 25 |
| 2 | 0 | 5 | 0 | 0 | 5 |
| 3 | 0 | 0 | 29 | 0 | 29 |
| 4 | 0 | 0 | 0 | 31 | 31 |
| Total | 24 | 6 | 29 | 31 | 90 |

Validity of the DS with Hone Graph

A comparison of the classifications by Hone Graph and eleven observers are shown in Table 2. The PA and kappa coefficient between Hone Graph and all observers were calculated. On average, the PA between Hone Graph and experienced BCNMTs, moderately experienced, and inexperienced nuclear medicine technologists were 85% (range, 79–92), 67% (51–80), and 38% (11–62), and the mean kappa was 0.82 (range, 0.70–0.94), 0.54, (0.34–0.72), and 0.32 (0.15–0.49), respectively. As a result, the correlation between Hone Graph and experienced BCNMTs, moderately experienced, and inexperienced

nuclear medicine technologists was almost perfect, moderate, and fair, respectively. As such, Hone Graph corresponded closely to the observation of experienced BCNMTs.

Table 2 Frequency table showing classifications made by Hone Graph compared to eleven observers. Observer A to D, E to G and H to K are experienced BCNMTs, moderately experienced, and inexperienced nuclear medicine technologists, respectively.

| Hone Graph | Experienced BCNMTs | | | | | | | | | | | | | | | | |
|------------|--------------------|---|----|----|------------|---|----|----|------------|----|----|----|------------|---|----|----|-------|
| | Observer A | | | | Observer B | | | | Observer C | | | | Observer D | | | | Total |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | |
| 1 | 22 | 3 | 0 | 0 | 23 | 2 | 0 | 0 | 23 | 2 | 0 | 0 | 23 | 2 | 0 | 0 | 25 |
| 2 | 0 | 5 | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 3 | 0 | 0 | 0 | 5 | 0 | 0 | 5 |
| 3 | 0 | 0 | 28 | 1 | 0 | 0 | 25 | 4 | 0 | 5 | 20 | 4 | 0 | 1 | 15 | 13 | 29 |
| 4 | 0 | 0 | 0 | 31 | 0 | 0 | 1 | 30 | 0 | 0 | 6 | 25 | 0 | 0 | 0 | 31 | 31 |
| Total | 22 | 8 | 28 | 32 | 23 | 7 | 26 | 34 | 25 | 10 | 26 | 29 | 23 | 8 | 15 | 44 | 90 |

| Hone Graph | Moderately experienced nuclear medicine technologists | | | | | | | | | | | | Total |
|------------|---|---|----|----|------------|----|----|---|------------|---|----|----|-------|
| | Observer E | | | | Observer F | | | | Observer G | | | | |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | |
| 1 | 19 | 6 | 0 | 0 | 25 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 25 |
| 2 | 0 | 3 | 2 | 0 | 4 | 1 | 0 | 0 | 1 | 4 | 0 | 0 | 5 |
| 3 | 0 | 0 | 10 | 19 | 1 | 15 | 13 | 0 | 0 | 5 | 18 | 6 | 29 |
| 4 | 0 | 0 | 0 | 31 | 0 | 0 | 24 | 7 | 0 | 0 | 6 | 25 | 31 |
| Total | 19 | 9 | 12 | 50 | 30 | 16 | 37 | 7 | 26 | 9 | 24 | 31 | 90 |

| Hone Graph | Inexperienced nuclear medicine technologists | | | | | | | | | | | | | | | | |
|------------|--|----|----|---|------------|----|----|---|------------|----|----|---|------------|----|----|----|-------|
| | Observer H | | | | Observer I | | | | Observer J | | | | Observer K | | | | Total |
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | |
| 1 | 22 | 3 | 0 | 0 | 25 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 18 | 7 | 0 | 0 | 25 |
| 2 | 1 | 8 | 1 | 0 | 5 | 0 | 0 | 0 | 2 | 3 | 0 | 0 | 0 | 5 | 0 | 0 | 5 |
| 3 | 0 | 11 | 14 | 1 | 9 | 18 | 2 | 0 | 5 | 18 | 6 | 0 | 1 | 13 | 15 | 0 | 29 |
| 4 | 0 | 1 | 23 | 5 | 0 | 11 | 17 | 3 | 0 | 4 | 21 | 6 | 0 | 1 | 13 | 17 | 31 |
| Total | 23 | 23 | 38 | 6 | 39 | 29 | 19 | 3 | 32 | 25 | 27 | 6 | 19 | 26 | 28 | 17 | 90 |

Discussion

The aim of this study was to develop an automatic quantification package for phantom-based image quality assessment of bone SPECT and to assess its validity. The software could automatically classify the DS of hot lesions in the SIM² bone phantom using the self-calculated quantitative indexes. The excellent agreement of DS between Hone Graph and the observation of expert BCNMTs was indicated. The most interesting finding was that the %DEV was the most strongly associated index with DS.

Hone Graph, allows easy assessment and reliable results of image quality (e.g., CNR, %CV, and detectability) with excellent repeatability and reproducibility since, the software can automatically calculate the indexes simply by selecting SPECT image files. Moreover, the report will be visually easier to understand for a wide range of experiences. Our findings underscore that the detectability calculated by Hone Graph almost perfectly agreed with that of expert BCNMTs. Thus, Hone Graph is not only easier, but also may improve repeatability and reproducibility for quality assessment in bone SPECT. The repeatability and reproducibility for Hone Graph was significantly superior than those for the results of experienced BCNMTs performed ROI analysis. Hone Graph provided the same results when the same SPECT data were repeatedly measured. Notably, misalignment of the settings of the phantom will not substantially affect the results. Owing to the rigid alignment to the reference image and resampling to 1-mm pixel size, the influence of shift alignment will be significantly decreased, especially in small spheres

such as those that are 13-mm, given that the pixel size of bone SPECT images is generally approximately 5 mm and the bone SPECT images will be affected by shift alignment.

In the several indexes calculated using the software, the most strongly associated index with DS was %DEV; thus, this was selected as the first node in the decision tree analysis. We defined %DEV as an alternative index for detectability, and in our preliminary experiment, a threshold value of 40% was optimal. Subsequently, classifications between DS 1 and 2 and between DS 3 and 4 showed detection performance with a certainty of 100% and required %DEV for 14.215%. Then, the classification between DS 1 and 2 indicated whether hot lesion was barely detected, was performed with CNR for 5.25. This value is in agreement with our earlier observations, which showed that observable lesions have a CNR greater than 5 with the Rose criteria [7]. Indeed, in the present study, only one out of 65 detectable lesions did not have a CNR greater than 5. Although decision tree analysis indicates a strong correlation between DS and not $CNR > 5$, $\%DEV > 14\%$ was found on detection of hot lesion. Hitherto, several previous reports have demonstrated that decision tree analysis provides a new incremental value in nuclear medicine [21, 22]. In the current study, the reason for the weak correlation between DS and CNR can be attributed to the fact that only hot lesions with a CNR from 4 to 6 were included. Indeed, many studies have only focused on only CNR for the detection of hot lesions; thus, its potential effects on the lesion volume and maximum density in the image may be neglected [11, 12, 23, 24]. However, prior studies have also highlighted the importance of a quantitative index with CNR and lesion volume for detectability [8, 13]. These previous reports focus on the number of pixels in the lesion ROI, although comparison of images of different pixel sizes is impossible. In the current study, $CNR_{\%DEV}$ was calculated according to the CNR, and a factor for lesion volume was not also associated with the detectability of hot lesions. One unexpected finding was that the SI, free from influence of background and noise as a resolution characteristic, was poorly associated with the DS, despite the finding that hot lesions were blurred in virtue of the 9.6 mm Gaussian filter.

Given that visual analysis may lead to greater inter-observer variation and more laborious work, previous studies have substituted CNR for detectability [8–11]. Although quantitative indexes, such as CNR, strongly correlate with detectability, measurement of CNR using manual ROI settings showed significant inter-observer variation in this study. In the manual measurement of CNR, variation between observers is one of the primary reasons for variation in the CNR, irrespective of the length of experience [25]. In the manual method, the CNRs exhibited great difference of more than three times that of inter-observer, whereas Hone Graph provided stable results. The intra-observer variation for the manual method was approximately 10%, although the method significantly underperformed Hone Graph. Very few studies have evaluated the inter-observer reproducibility of ROI analysis, and it has been largely neglected in phantom-based studies. The excellence in repeatability and reproducibility of Hone Graph is supported by previous studies that demonstrated the validity of SPM [26, 27]. Regarding detectability, the results showed that classification of the DS using Hone Graph was in almost perfect agreement with the gold standard, and in excellent agreement with experienced BCNMTs. However, with each decrease in year of experience, the agreement with the DS of Hone Graph also decreased. This variation between years of

experience is in line with previous studies that demonstrated that the performance in the interpretations improved with experience, although between clinical study and phantom study is different as the observers know where to look for the lesions [28, 29].

Even when using a different device or acquisition parameter (e.g., pixel size or acquisition time), Hone Graph will likely be useful to serve as comparison for image-to-image variation due to pixel size resampling and normalization of count distribution. Hone Graph, the phantom-based method, is considered to be a simple and convenient method to understand differences in gamma camera system or imaging technology. As an extension of this idea, standardization of bone SPECT imaging technology may be possible utilizing Hone Graph. Quantitative indexes are usually calculated on the central slice though the target (i.e., two-dimensions only), whereas Hone Graph using VOI (i.e., three-dimensions) though showed excellent reproducibility. The CNR measurement using ROI retains the potential to pose a serious problem for reproducibility due to variation in pixel size or a slice though the target that is off-center. The use of ROI analysis will lead to difficulty in comparison of CNR, even in variation of pixel size alone [12].

Many studies published on lesion detection in bone SPECT highlight the importance for [1–5] nuclear medicine technologists to maintain adequate imaging technology for bone SPECT. However, to the best of our knowledge, almost no reports have described imaging technology for bone SPECT. The rapid SPECT/CT protocol has a potential impact on the acquisition of SPECT/CT, although the scan protocol has not been shown to be able to detect smaller or weaker lesions. Conversely, phantom-based image quality assessment assists in developing a greater understanding of imaging technology, and the ability to detect small lesion. In terms of quality assessment, the accuracy of quantification, repeatability, and reproducibility are of the utmost importance, all of which are possible with the software.

As a limitation in our study, we assessed the proposed software using only one gamma camera system and limited parameter ranges of imaging technology. The maximum value in the reference section is a criterion value that is used to assess image quality using Hone Graph, although the maximum value in the reference section may be decreased by Gibb's artifact, which can lead to overestimation of the DS in lesions [30]. Even after optimization based on the proposed method, additional assessment with a physician for clinical image is likely to be needed due to the structural difference between the phantom and the human body. It is conceivable that the Hone Graph program can assess bone SPECT images among multiple centers; however, the quantitative index, other than the DS calculated by the program, has no criterion value. Further multicenter studies are required to define criteria to optimize imaging technology for bone SPECT. With increased volumes of SPECT images, Hone Graph may evolve to machine learning in the same way as random forest or support vector machine.

Conclusions

We developed quantification software for bone SPECT that performs after data input to output automatically, and is excellent in terms of reproducibility and accuracy. Detectability was classified with

both %DEV, CNR, TNR, and SI. Hone Graph may have the potential to be used as a tool for the standardization of bone SPECT imaging technology; however, there is a need for further studies to test our findings with other devices in the future.

Abbreviations

CNR

Contrast-to-noise ratio, ROI:Region of interest, VOI:Volume of interest, SPM:Statistical parametric mapping, %DEV:Percentage of detectability equivalence volume, %CV:Percentage of coefficient of variation, TNR:Target-to-normal bone ratio, SI:Sharpness index, FWHM:Full width at the half maximum, BCNMT:Board-certified nuclear medicine technologists, DS:Detectability score, ICC:Intra-observer reliability, PA:Percentage agreement

Declarations

Ethics approval and consent to participate

Based on the nature of this study, the requirement for ethical review was waived by the medical ethics review committee of our hospital.

Consent for publication

Not applicable.

Availability of data and materials

The data generated and analyzed in this study are available from the corresponding author on reasonable request.

Competing interests

Hajime Ichikawa has a collaborative research work for developing software with FUJIFILM Toyama Chemical, Co. Ltd, Tokyo, Japan. Kazunori Kawakami, Kazuki Nagatake and Tetsuo Hosoya are employees of FUJIFILM Toyama Chemical Co., Ltd. All other authors declare that they have no competing interests.

Funding

This study was supported by FUJIFILM Toyama Chemical, Co. Ltd, Tokyo, Japan to publish the study.

Authors' contributions

HI and KK conceived the study. KN, TH, and KK developed the software and modified the program. HI, TS, TI, and TK acquired the SPECT scans, reconstructed the images, and analyzed the image data. HI, KK, and MO provided key input to the conceptual design and the structure of the paper. All authors critically reviewed and revised the manuscript and consented to its publication. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Hiroto Yoneyama, Hirotatsu Tsuchikame, Asuka Mizutani, Takahiro Konishi, Tomoya Banno, Taiki Kato, Yu Kawakita, and Jumpei Osanai for taking part in the human observer study, and Yoshinao Misu for supporting the data analysis.

References

1. Schirrmeister H, Glatting G, Hetzel J, Nüssle K, Arslanemir C, Buck AK, Dziuk K, Gabelmann A, Reske SN, Hetzel M. Prospective evaluation of the clinical value of planar bone scans, SPECT, and ¹⁸F-labeled NaF PET in newly diagnosed lung cancer. *J Nucl Med* 2001;42(12):1800-4.
2. Even-Sapir E, Metser U, Mishani E, Lievshitz G, Lerman H, Leibovitch I. The detection of bone metastases in patients with high-risk prostate cancer: ^{99m}Tc-MDP planar bone scintigraphy, single- and multi-field-of-view SPECT, ¹⁸F-fluoride PET, and ¹⁸F-fluoride PET/CT. *J Nucl Med* 2006;47(2):287-97.
3. Abikhzer G, Gourevich K, Kagna O, Israel O, Frenkel A, Keidar Z. Whole-body bone SPECT in breast cancer patients: the future bone scan protocol? *Nucl Med* 2016;37(3):247-53.
4. Römer W, Nömayr A, Uder M, Bautz W, Kuwert T. SPECT-guided CT for evaluating foci of increased bone metabolism classified as indeterminate on SPECT in cancer patients. *J Nucl Med* 2006;47(7):1102-6.
5. Ichikawa H, Miwa K, Okuda K, Shibutani T, Kato T, Nagaki A, et al. Current state of bone scintigraphy protocols and practice in Japan. *Asia Ocean J Nucl Med Biol*. 2020.
6. Sadik M, Suurkula M, Höglund P, Järund A, Edenbrandt L. Quality of planar whole-body bone scan interpretations—a nationwide survey. *Eur J Nucl Med Mol Imaging*. 2008;35(8):1464-72.
7. Vision RA. 1st ed. US: Springer; 1973.
8. Carlier T, Eugène T, Bodet-Milin C, Garin E, Ansquer C, Rousseau C, Ferrer L, Barbet J, Schoenahl F, Kraeber-Bodéré F. Assessment of acquisition protocols for routine imaging of Y-90 using PET/CT. *EJNMMI Res*. 2013;3(1):11.
9. Beijst C, de Keizer B, Lam MGEH, Janssens GO, Tytgat GAM, de Jong HWAM. A phantom study: should ¹²⁴I-mIBG PET/CT replace ¹²³I-mIBG SPECT/CT? *Med Phys*. 2017(2473-4209 (Electronic));44(5):1624-31.

10. 10. AlqahtaniMS, LeesJE, BugbySL, Samara-RatnaP, NgAH, PerkinsAC. Design and implementation of a prototype head and neck phantom for the performance evaluation of gamma imaging systems. *EJNMMI* 2017;4(1):19.
11. 11. AdlerS, SeidelJ, ChoykeP, KnoppMV, BinzelK, ZhangJ, BarkerC, ConantS, Maass-MorenoR. Minimum lesion detectability as a measure of PET system performance. *EJNMMI* 2017;4(1):13.
12. 12. ØenSK, AasheimLB, EikenesL, KarlbergAM. Image quality and detectability in Siemens Biograph PET/MRI and PET/CT systems-a phantom study. *EJNMMI* 2019;6(1):16.
13. 13. BeijstC, KistJW, ElschotM, ViergeverMA, HoekstraOS, de KeizerB, de JongHW. Quantitative comparison of 124I PET/CT and 131I SPECT/CT detectability. *J Nucl* 2016;57(1):103-8.
14. 14. IchikawaH, KatoT, ShimadaH, WatanabeY, MiwaK, MatsutomoN, et al. Detectability of thoracicbonescintigraphyevaluatedusing a novelcustom-designedphantom. *The JpnJ NuclMedTechnol.* 2017;37(3):229-38.
15. 15. ImbertL, PoussierS, FrankenPR, SongyB, VergerA, MorelO, WolfD, NoelA, KarcherG, MariePY. Compared performance of high-sensitivity cameras dedicated to myocardial perfusion SPECT: a comprehensive analysis of phantom and human images. *J Nucl* 2012;53(12):1897-903.
16. 16. CachovanM, VijaAH, HorneggerJ, KuwertT. Quantification of (99m)Tc-DPD concentration in the lumbar spine with SPECT/CT. *EJNMMI Res.* 2013;3(1):45.
17. 17. KanetaT, OgawaM, DaisakiH, NawataS, YoshidaK, InoueT. SUV measurement of normal vertebrae using SPECT/CT with Tc-99m methylene diphosphonate. *Am J Nucl Med Mol Imaging.* 2016;6(5):262-8.
18. 18. Kujil, YamaneT, SetoA, YasumizuY, ShirotakeS, OyamaM. Skeletal standardized uptake values obtained by quantitative SPECT/CT as an osteoblastic biomarker for the discrimination of active bone metastasis in prostate cancer. *Eur J Hybrid Imaging.* 2017;1(1):2.
19. 19. de DreuilleOd, StrijckmansV, AmeidaP, Loc'hC, BendriemB. Bone equivalent liquid solution to assess accuracy of transmission measurements in SPECT and PET. *IEEE Trans NuclSci.* 1997;44(3):1186-90.
20. 20. ZhangZ. Decision tree modeling using R. *Ann Transl* 2016;4(15):275.
21. 21. AssanteR, ZampellaE, NicolaiE, AcampaW, VergaraE, NappiC, GaudieriV, FiumaraG, KlainM, PetrettaM, CuocoloA. Incremental value of sestamibi SPECT/CT over dual-phase planar scintigraphy in patients with primary hyperparathyroidism and inconclusive ultrasound. *Front Med (Lausanne).* 2019;6:164.
22. 22. GuerratyMA, RaoHS, AnjanVY, SzaparyH, MankoffDA, PrymaDA, RaderDJ, DubroffJG. The role of resting myocardial blood flow and myocardial blood flow reserve as a predictor of major adverse cardiovascular outcomes. *PLOS ONE.* 2020;15(2):e0228931.
23. 23. DickerscheidD, LavalayeJ, RomijnL, HabrakenJ. Contrast-noise-ratio (CNR) analysis and optimisation of breast-specific gamma imaging (BSGI) acquisition protocols. *EJNMMI Res.* 2013;3(1):21.

24. 24. DietzeMMA, KunnenB, van der VeldenS, SteenbergenJHL, KoppertWJC, ViergeverMA, de JongHWAM. Performance of a dual-layer scanner for hybrid SPECT/CBCT. *Phys Med* 2019;64(10):105020.
25. 25. VerhoeffNP, KapucuO, Sokole-BusemannE, van RoyenEA, JanssenAG. Estimation of dopamine D2 receptor binding potential in the striatum with iodine-123-IBZMSPECT: technical and interobserver variability. *J Nucl* 1993;34(12):2076-84.
26. 26. ClausJJ, van HarskampF, BretelerMMB, KrenningEP, van der CammenTJM, HofmanA, HasanD. Assessment of cerebral perfusion with single-photon emission tomography in normal subjects and in patients with Alzheimer's disease: effects of region of interest selection. *EurJ NuclMed*. 1994;21(10):1044-51.
27. 27. LiuHG, MountzJM, InampudiC, San PedroEC, DeutschG. A semiquantitative cortical circumferential normalization method for clinical evaluation of rCBF brain SPECT. *ClinNuclMed*. 1997;22(9):596-604.
28. 28. SadikM, SuurkulaM, HöglundP, JärundA, EdenbrandtL. Quality of planar whole-body bone scan interpretations—a nationwide survey. *Eur J Nucl Med Mol Imaging*. 2008;35(8):1464-72.
29. 29. ArsanjaniR, XuY, HayesSW, FishM, LemleyM, , GerlachJ, DorbalaS, BermanDS, GermanoG, SlomkaP. Comparison of fully automated computer analysis and visual scoring for detection of coronary artery disease from myocardial perfusion SPECT in a large population. *J Nucl Med*. 2013;54(2):221-8.
30. 30. ZengGL. Gibbs artifact reduction by nonnegativity constraint. *J Nucl Med* 2011;39(3):213-9.

Figures

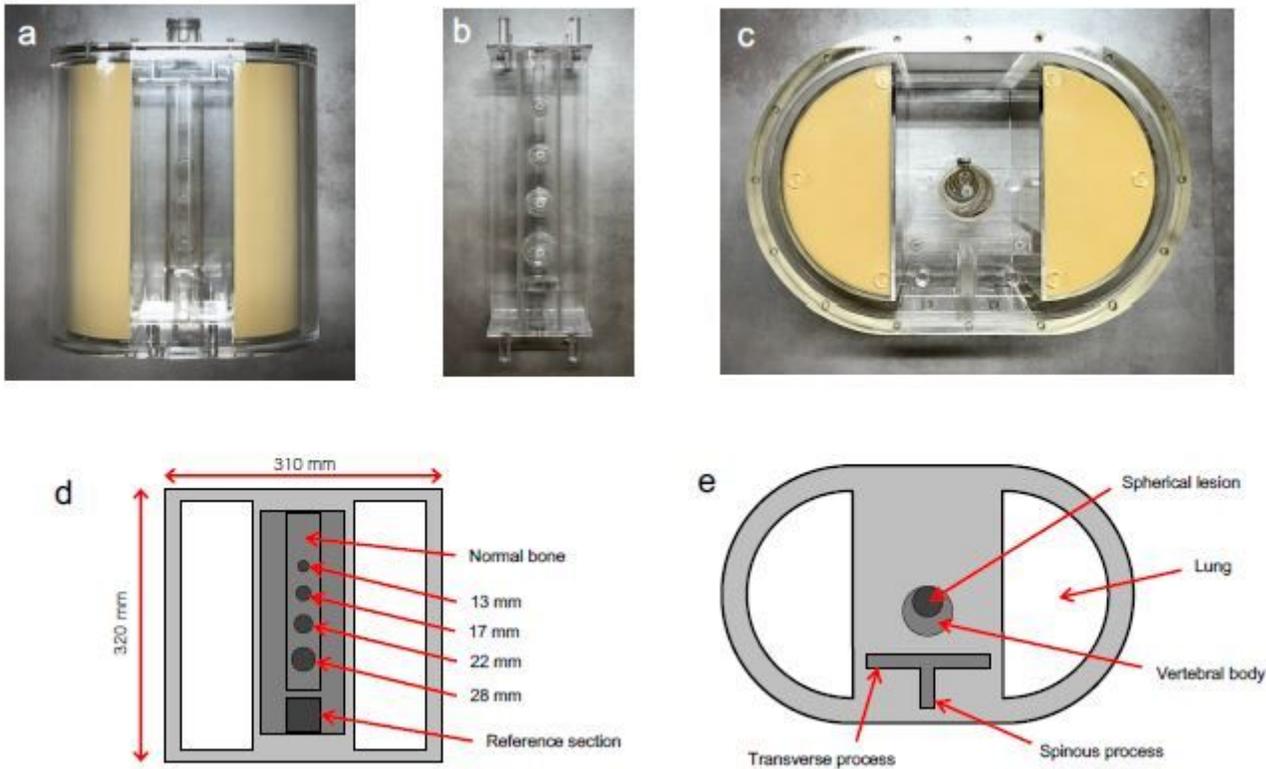


Figure 1

SIM2 bone phantom with spine components. a) Frontal view, b) spine components, and c) superior view of the SIM2 bone phantom. Diagrams of the SIM2 bone phantom: d) Frontal view and e) superior view.

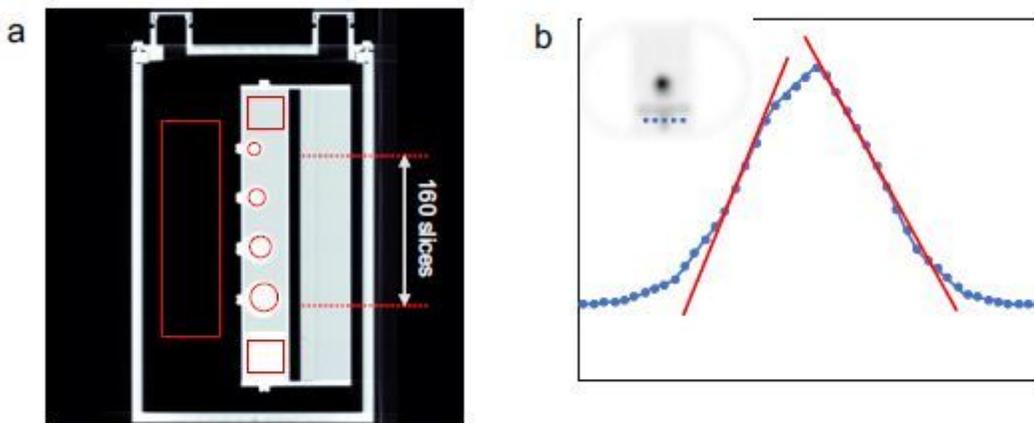


Figure 2

Volumes of interest (VOI) and line-profile setting. a) VOI template. b) The mean count profile curve (blue line) of the central 160 slices is indicated by the dashed lines. Two red lines indicate the left-right maximal slope at 6 pixels. Distance and pixel intensity are expressed in mm and normalized pixel values, respectively.

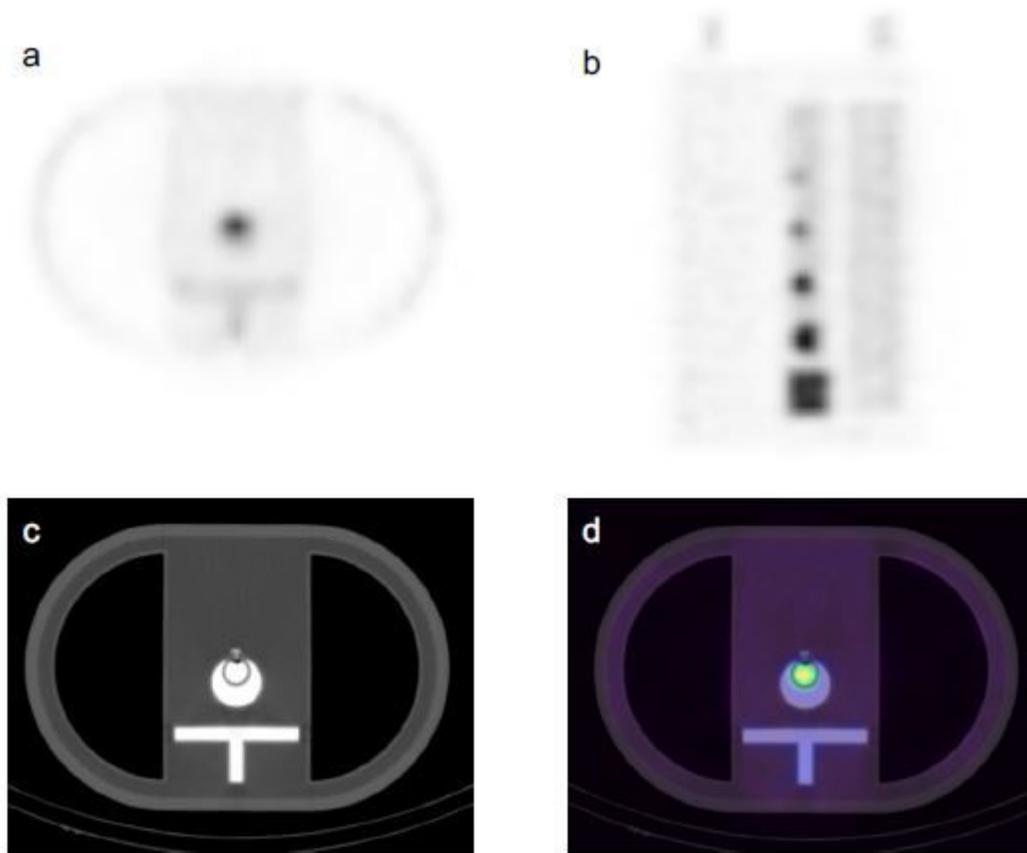


Figure 3

Typical SPECT and CT image of the SIM2 bone phantom. Top row: a) transverse, and b) sagittal SPECT images of the SIM2 bone phantom. Bottom row: c) CT, and d) fused single-photon emission computed tomography coupled with computed tomography images. Vertebral body, tumor, and processus contained K_2HPO_4 solution with a density equivalent to bone and ^{99m}Tc , and the body phantom contained ^{99m}Tc solution.

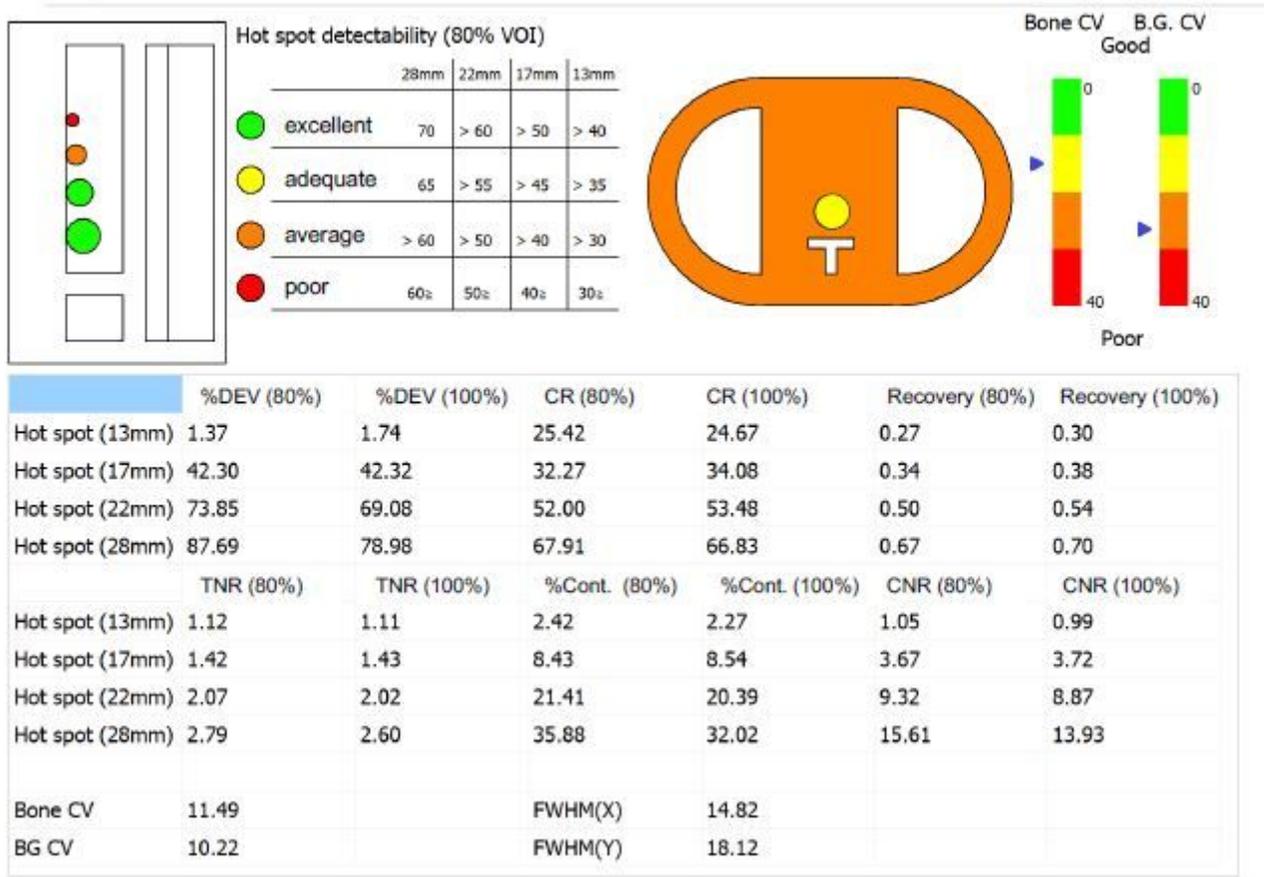


Figure 4

Performance report for the Hone Graph software. For example, with a detectability score of 13-mm, the hot lesion showed as red is not observed as a metastatic.

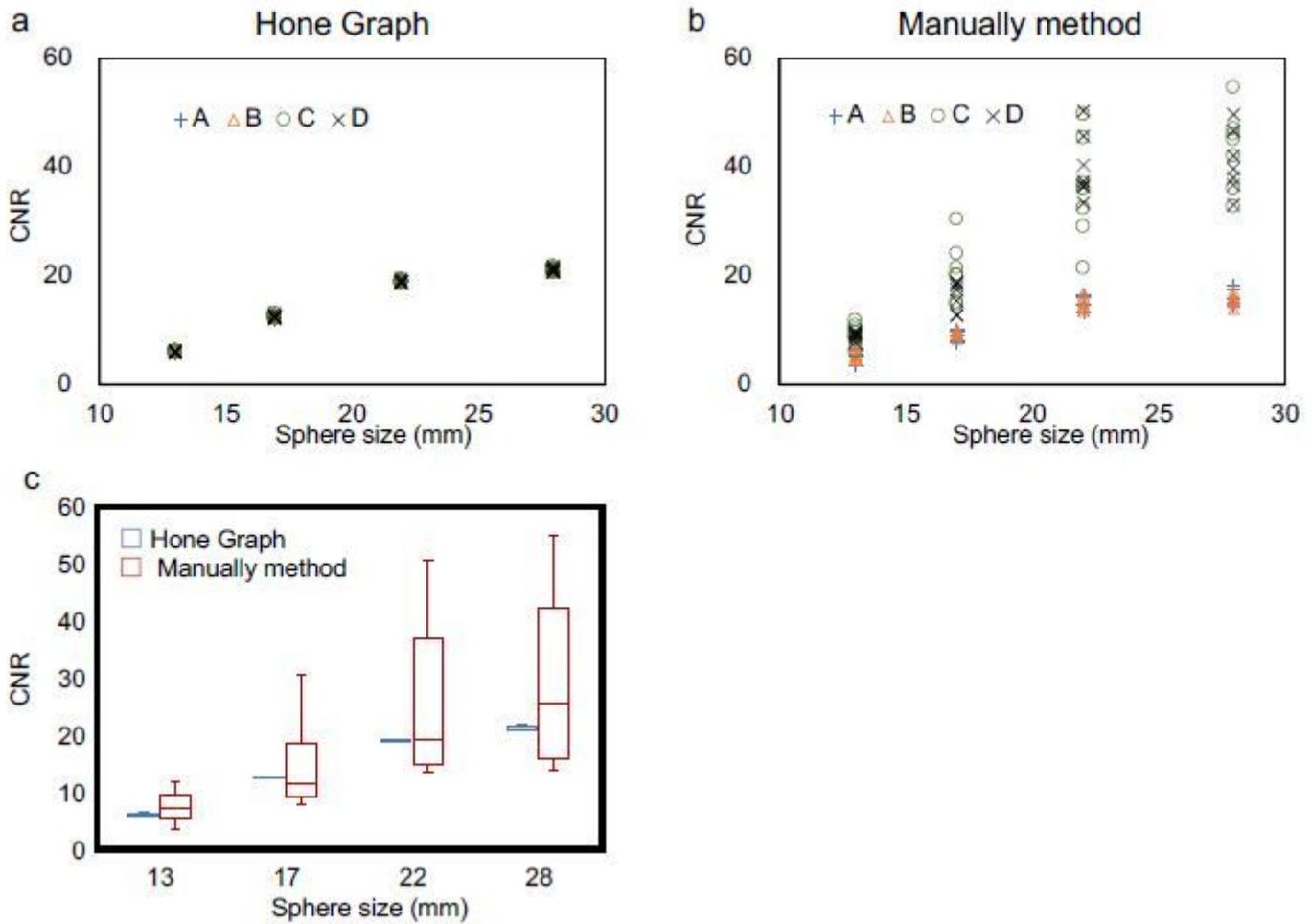


Figure 5

Contrast-to-noise ratios (CNRs) using automatic and manual methods. CNRs for a) Hone Graph, and b) manual method by four expert board-certified nuclear medicine technologists (A–D). c) Boxplot of CNRs for the Hone Graph and manual method (same data as a and b).

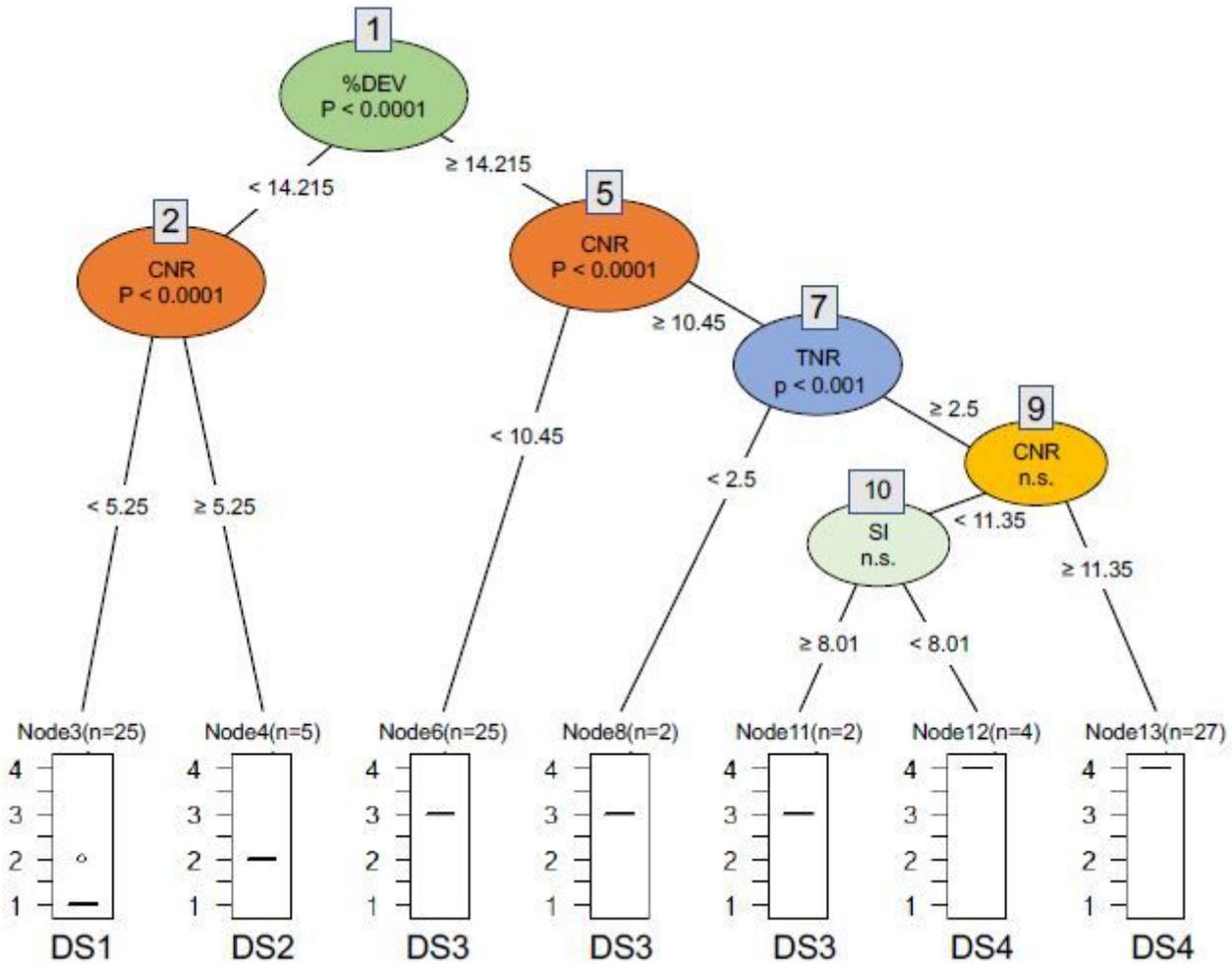


Figure 6

Decision tree analysis for the classification of the detectability score based on quantitative indexes. Input variable and P-values are given for each inner node (1, 2, 5, 7, 9, 10), and a boxplot of the detectability score is displayed for each terminal node.

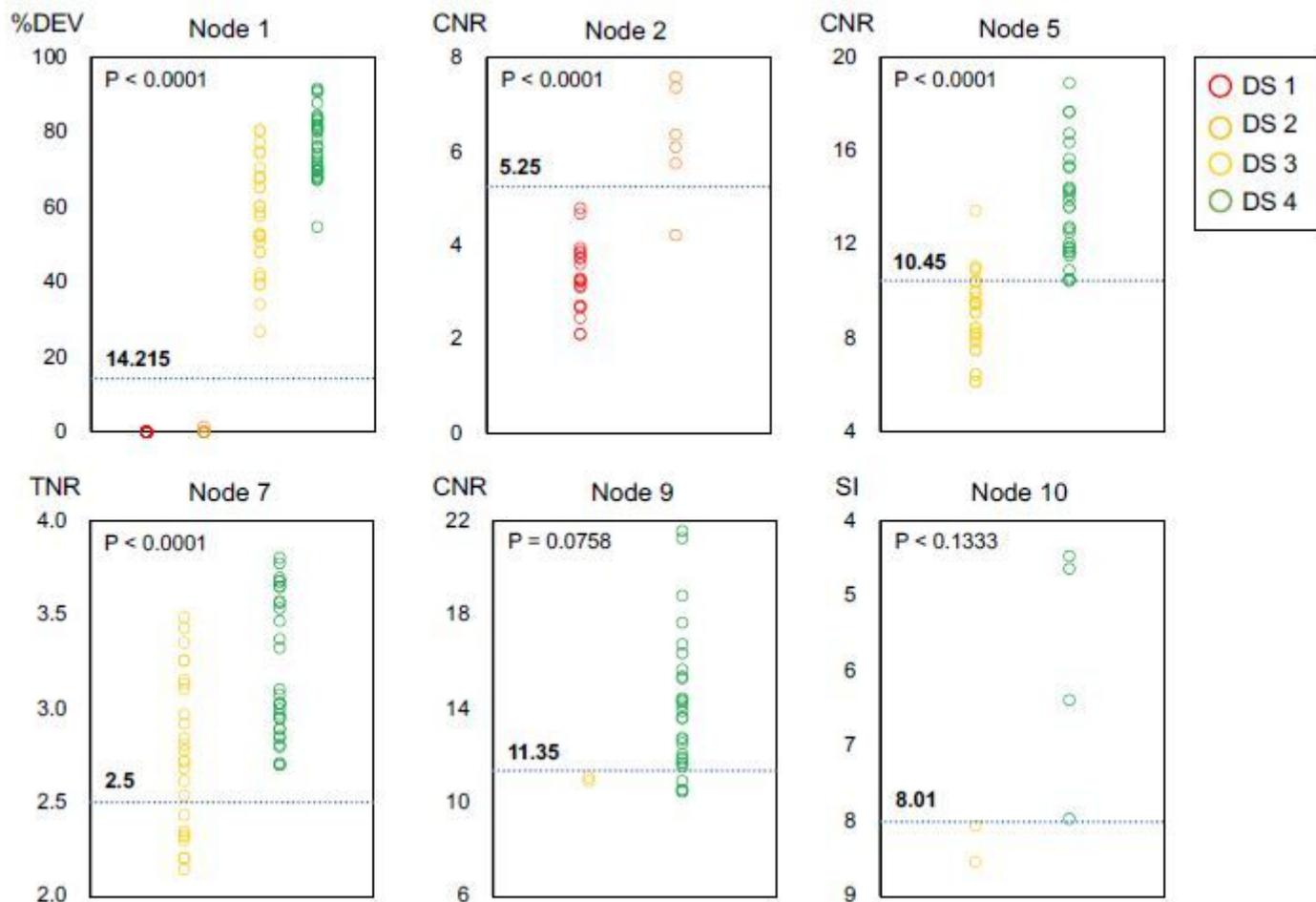


Figure 7

Scatter plot showing the observed quantitative indexes against the detectability score of the gold standard. In node 2, only one detectability score, using automatically classification, was diverged from that of the gold standard. The detectability scores classified (1–4) in the human observer study (gold standard) are presented in red, orange, yellow, green, respectively.