Computer-Assisted Interpretation of Planar Whole-Body Bone Scans

May Sadik[1], Iman Hamadeh[2], Pierre Nordblom[2], Madis Suurkula[1], Peter Höglund[3], Mattias Ohlsson[4], and Lars Edenbrandt[1,2,5]

[1]Department of Molecular and Clinical Medicine, Clinical Physiology, Sahlgrenska University Hospital, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden; [2]EXINI Diagnostics AB, Lund, Sweden; [3]Competence Centre for Clinical Research, Lund University Hospital, Lund, Sweden; [4]Department of Theoretical Physics, Lund University, Lund, Sweden; and [5]Department of Clinical Sciences, Malmö, Lund University, Sweden

The purpose of this study was to develop a computer-assisted diagnosis (CAD) system based on image-processing techniques and artificial neural networks for the interpretation of bone scans performed to determine the presence or absence of metastases. **Methods:** A training group of 810 consecutive patients who had undergone bone scintigraphy due to suspected metastatic disease were included in the study. Whole-body images, anterior and posterior views, were obtained after an injection of ^{99m}Tc-methylene diphosphonate. The image-processing techniques included algorithms for automatic segmentation of the skeleton and automatic detection and feature extraction of hot spots. Two sets of artificial neural networks were used to classify the images, 1 classifying each hot spot separately and the other classifying the whole bone scan. A test group of 59 patients with breast or prostate cancer was used to evaluate the CAD system. The patients in the test group were selected to reflect the spectrum of pathology found in everyday clinical work. As the gold standard for the test group, we used the final clinical assessment of each case. This assessment was based on follow-up scans and other clinical data, including the results of laboratory tests, and available diagnostic images, such as from MRI, CT, and radiography, from a mean follow-up period of 4.8 y. **Results:** The CAD system correctly identified 19 of the 21 patients with metastases in the test group, showing a sensitivity of 90%. False-positive classification of metastases was made in 4 of the 38 patients not classified as having metastases by the gold standard, resulting in a specificity of 89%. **Conclusion:** A completely automated CAD system can be used to detect metastases in bone scans. Application of the method as a clinical decision support tool appears to have significant potential.

Key Words: computer-assisted diagnosis; radionuclide imaging; bone metastases; image processing; neural networks

J Nucl Med 2008; 49:1958–1965
DOI: 10.2967/jnumed.108.055061

Bone scanning is widely accepted as a method of choice for initial diagnosis of bone and joint changes in patients with oncologic diseases (*1–4*). Because the choice of treatment strategy is influenced by the presence or absence of bone metastases, the correct interpretation of the bone scans is important. Classification of the bone scans is, however, a subjective task, and some of the previous studies in this field have shown that false-negative interpretations are unacceptably high (*5,6*). Our group recently performed a nationwide survey to investigate observer variation and performance regarding the interpretation of bone metastases (*7*). Thirty-seven observers with various levels of experience working at 18 different hospitals in Sweden participated. The sensitivities for the observers ranged from 52% to 100%, with an average of 77% at a mean specificity of 96%. The study also showed moderate interobserver agreement among the 37 observers when they were pairwise compared with each other (mean κ, 0.48). Our findings were in agreement with those of Rossing et al. (*5*), in whose study 3 experienced observers reread 842 bone scans from 12 different hospitals. With the interpretation of the 3 panel observers considered as the gold standard, the sensitivities and specificities of the original reports were 78% and 84%, respectively (*5*). Peters et al. performed a clinical audit in nuclear medicine and found that 19 of 220 reports (8.6%) were classified as having nontrivial interpretation errors, in which the physicians failed to mention increased uptake in their reports on patients with cancer (*6*).

Computer-assisted diagnosis (CAD) systems have recently become a part of the clinical routine work for detection of breast cancers on mammograms at many screening sites and hospitals in the United States (*8–12*). These systems have been shown to significantly improve the performance of the physician in finding cancers (*10–12*). Similar CAD systems have also been shown to increase the sensitivity of less experienced physicians in detecting polyps in CT colonography (*13*) and to improve performance and decrease interobserver variability regarding interpretations in myocardial perfusion imaging (*14*). CAD systems could, therefore, be of

Received Jun. 10, 2008; revision accepted Aug. 18, 2008.
For correspondence or reprints contact: May Sadik, Department of Molecular and Clinical Medicine, Clinical Physiology, Sahlgrenska Academy at the University of Gothenburg SE 413 45, Gothenburg, Sweden.
E-mail: may.sadik@vgregion.se

value in many applications in the field of diagnostic imaging. The concept of CAD is to assist the physician by combining his or her competence and knowledge with the capability of the computer in detecting lesions in medical images.

We have recently developed an automated CAD system for the interpretation of bone scans, and the results showed a sensitivity of 90% at a specificity of 74% (*15*). The results were encouraging, but further improvement of the system is needed for it to be used in the daily clinical setting. Therefore, on the basis of improved image processing and artificial neural network techniques and a large database of whole-body bone scans, the purpose of the present study was to develop a completely automated CAD system for the interpretation of bone scans to determine the presence or absence of metastases.

MATERIALS AND METHODS

Patients

Training Group. A training group was used in the process of developing the CAD system described in this article. The CAD system consists of image-processing techniques and an artificial neural network that learns by example. We selected a little more than 800 bone scans (which is the recommended number of bone scans a physician should interpret during specialist training) for the training group. Cases that could be misleading for the CAD system during the training process, for example, patients with a urine catheter, large bladder, sternotomy, or fracture, were excluded from the training group. These types of cases with high radiotracer activity are generally easy for the physician to interpret but difficult for a computer method that learns by example, because only few cases with a similar pattern are present even in a large training group.

We retrospectively included 971 consecutive patients who had undergone whole-body bone scintigraphy with a dual-detector γ-camera because of suspected bone metastatic disease during the period January 1999 to June 2002. Only patients with a complete set of technically sufficient images were included. A total of 51 cases with images that could be misleading were excluded.

We made the exclusions without any knowledge of the test patients. The patients from the test group and their follow-up examinations during the period of study were also excluded from the training group (110 patients). The final training group consisted of 810 patients (Table 1).

Test Group. Patients who had undergone whole-body bone scintigraphy with a dual-detector γ-camera because of suspected bone metastatic disease and who also had at least 1 follow-up bone scan were retrospectively selected. The reason for including the follow-up examinations was that these images could improve the accuracy of the gold standard interpretation. The patients were selected from the period August 1999 to January 2001 at Sahlgrenska University Hospital.

TABLE 1
Study Population

Parameter	Training group	Test group
No. of patients	810	59
Female (%)	35	31
Mean age (y)	66 (range, 25–92)	65 (range, 43–86)
Prevalence of bone metastases (%)	34	36

To avoid skewed material, the patients were selected to reflect the spectrum of pathology found in everyday clinical work, that is, patients with breast or prostate cancer coming for either their first bone scan or a follow-up scan. We recently studied this type of patient group and found that approximately one third of the cases had clear-cut benign findings (estimated probability of metastases, 0.05 or lower), one third were difficult cases with an intermediate probability of metastases (estimated probability of metastases, 0.06–0.94), and one third were clear-cut cases with obvious metastases (estimated probability of metastases, 0.95 or higher) (*15*). We aimed to achieve approximately the same relation between the 3 groups when selecting the bone scans for the present study. Patients in the 3 groups, "benign findings," "difficult cases," and "obvious metastases," were included consecutively until the one-third quota was fulfilled for each group.

The final test group consisted of 59 patients with a diagnosis of breast or prostate cancer (Table 1). The test group had previously been used in our nationwide survey, in which interpretations by physicians of bone scans were studied (*7*).

Bone Scintigraphy. Bone scans were obtained approximately 3 h after an intravenous injection of ^{99m}Tc-methylene diphosphonate (600 MBq). Whole-body images—anterior and posterior views (scan speed, 10 cm/min; matrix, 256 × 1,024)—were obtained with a γ-camera equipped with low-energy, high-resolution parallel-hole collimators (Maxxus; GE Healthcare) and stored on a computer system (Star Cam RMX; GE Healthcare). Energy discrimination was provided by a 15% window centered on the 140-keV peak of ^{99m}Tc.

Gold Standard

Training Group. The gold standard classification of the patients in the training group for presence or absence of bone metastases was based on the clinical reports and the bone scan images. Clinical data such as medical condition, localization of bone pain, and previous history of trauma were available to the reporting physicians at the time of the clinical reporting. These reports and the corresponding bone scans were reevaluated by a trained technologist together with an experienced physician who estimated the probability of bone metastases on an analog scale from 0 to 1. In difficult cases, reports from other diagnostic examinations that the patient had undergone, for example, follow-up scans, MRI scans, radiographs, or CT scans, were considered in the reevaluation.

A cutoff value of 0.5 was chosen to provide the CAD system with the binary classification of "bone metastases" or "no bone metastases." All patients with values below 0.5 were assigned to the no-bone-metastases group and patients with values equal to or above 0.5 were assigned to the bone-metastases group.

Test Group. The gold standard classification for the patients in the test group for presence or absence of bone metastases was based on the final clinical assessments made by the same experienced physician who reevaluated the patients of the training group.

These clinical assessments were based on all bone scan images, including the follow-up scans; the patients' medical records, including the results of laboratory tests; and available diagnostic images from MRI scans, CT scans, and radiographs. Biopsy was available in 1 case. The follow-up scans were used to observe

whether hot spots had disappeared, remained unchanged, or decreased or increased in size and intensity.

The following diagnostic criteria were applied for the final clinical assessment:

- Grade 1: absence of bone metastases (The scintigraphic pattern is normal or shows hot spots typical of degenerative changes or fractures; there are no clinical or radiographic data indicating bone metastases.)
- Grade 2: bone metastases cannot be ruled out with certainty (There are one or more visible hot spots that have disappeared, remained unchanged, or decreased in size and intensity on the follow-up scan. The patients in this group all received cancer therapy between the first and the follow-up scans, and hot spots could be healed fractures or degenerative changes or could be metastases. Radiographic modalities, when available, in the suspected regions did not favor a diagnosis of malignancy, and the gathered clinical judgment leaned toward a low probability of bone metastases.)
- Grade 3: bone metastases probable (Visible hot spots with localization, distribution, and intensity not typical of degenerative changes or fractures are demonstrated. Scintigraphic follow-up shows no substantial changes. Radiographic findings are equivocal, but the overall clinical judgment indicates probable bone metastases.)
- Grade 4: definite presence of bone metastases (Scintigraphic or radiographic patterns are typical of bone metastases; the medical record states bone metastases as a secondary diagnosis.)

The follow-up scans and the computerized medical record were updated until May 2006, resulting in a mean follow-up duration of 4.8 y (range, 11–81 mo). Twenty-two patients died during the follow-up period.

In the final clinical assessments, 32 patients were classified as grade 1, 6 as grade 2, 0 as grade 3, and 21 as grade 4. The 38 patients classified as grade 1 or 2 were considered as having no bone metastases, and the 21 patients classified as grade 3 or 4 were considered as having bone metastases in the calculations of sensitivity, specificity, and accuracy.

CAD System

The automated method for interpretation of bone scans developed in this study was based on the experience from our initial work in this field (*15*). New algorithms for segmentation, hot spot detection, and feature extraction were developed, and a larger training group was used. A more precise segmentation of the skeleton makes it possible to use different algorithms for detection of hot spots in different parts of the image and to present information in greater detail regarding the localization and distribution of hot spots to the artificial neural networks. The methods used for segmentation, hot spot detection, and feature extraction were developed using the MATLAB (The MathWorks Inc.) programming language. The training and validation of the artificial neural networks were performed using customized software. The methods developed will be incorporated in a software package developed by EXINI diagnostics AB (http://www.exini.com).

Image Segmentation

The active-shape model method was used to segment the entire skeleton, except for the distal parts of the arms and legs, in both anterior and posterior views. The distal extremities are often not acquired completely in routine bone scanning. The active-shape model was developed to find statistical models of objects in images, and the method has been successfully applied in many application areas including medical images (*16*). In the segmentation process, the skeleton was divided into 4 separate parts for both the anterior and the posterior images (head and spine, proximal arms and clavicles, chest, and pelvis and proximal legs) (Fig. 1). The search for and delineation of a specific part of the skeleton were based on a corresponding model, which was adjusted to optimize the fit with the image data in an iterative process. The model contained statistical information about the shape variation of that part of the skeleton from a set of training images.

The first step in the development of the segmentation method was to select several training images that included different shapes of the skeleton. In these cases, an operator manually delineated the shape of the skeleton by selecting different landmarks in the images. Each landmark point was placed on a particular anatomic part of the skeleton. This procedure was performed separately for each part of the skeleton. The shapes from all the training images (between 13 and 32 training cases for the different parts of the skeleton) were then aligned to a common coordinate frame. This was achieved by scaling, rotating, and translating the training shapes so that they corresponded as closely as possible to each other. The resulting shape model contained the typical mean shape and its variation observed in the training set. The resulting 8 shape models (1 for each of the 4 anterior and 4 posterior parts of the skeleton) were then used to segment new images.

FIGURE 1. Segmentation of skeleton divided into 4 separate parts for both anterior and posterior images (head and spine, proximal arms and clavicles, chest, and pelvis and proximal legs).

Automatic segmentation of new images was performed in 3 steps. The first step was to find a start position for the model of the first part of the skeleton (head and spine) in the anterior image. The most cranial part of the head proved to be a robust start position, and it was easily located by examining the intensity in the superior part of the image. In the second step, the landmarks of the model were adjusted to better fit with the actual pixel values of the image. This was achieved by examining a small area around each landmark to find changes in intensity that could, for example, correspond to the border of the skeleton. In the third step, the shape of the model was adjusted on the basis of the new positions of the different landmarks and the allowable shape variation of the model. The second and third steps were repeated in an iterative process until no significant change occurred between 2 iterations. After the segmentation of the first part of the skeleton, that segmentation was used to define the start positions of the other models.

Hot Spot Detection

Hot spots were detected using a region-specific threshold algorithm based on the mean and SD of all pixel count values from a specific region. Clusters of pixels with count values above this threshold and with a cluster size of at least 13 pixels were regarded as potential hot spots. The localization of each potential hot spot could be obtained on the basis of the result of the segmentation process. Hot spots corresponding to the bladder and the kidneys were excluded on the basis of location and size.

Feature Extraction

Forty-five features were used to describe each hot spot. The features were selected to describe both the hot spot itself and its relation to other hot spots (e.g., symmetric uptake in the shoulders) and the surrounding region. The size, shape, intensity, and localization of a hot spot and the intensity characteristics of the region in which the hot spot was located were calculated. The hot spot features are presented in Table 2.

Artificial Neural Networks

Artificial neural networks were used both to assess the likelihood that a specific hot spot represented a metastasis and to classify the complete bone scan examination (i.e., both the anterior and the posterior images of a patient) as having signs of metastases or not. A more general description of artificial neural networks can be found elsewhere (*17–19*). An ensemble of 30 single artificial neural networks was used for each classification task. The individual members of the ensemble were standard multilayer perceptrons (*20*) with 1 input, 1 hidden, and 1 output layer.

TABLE 2
Hot Spot Features

Hot spot feature	No. of features	Example
Geometry	8	Area, width, height
Pixel values	13	Maximal value, SD
Skeletal region	8	Spine, pelvis
Skeletal region features	8	Area of region, maximal value
Before classification	5	Symmetry, bladder
Combined features	3	Area ratio of hot spot to region
Total	45	

The neural networks classifying single hot spots consisted of 45 nodes in the input layer, 1 for each of the hot spot features. The hidden layer contained 20 nodes. The output node encoded whether the hot spot was classified as a metastasis or not. The output of a neural network ensemble was computed as the mean of the outputs of the individual members of the ensemble. The optimization of neural network parameters was performed using a 6-fold cross-validation scheme on the training group.

The neural networks classifying the complete bone scan examination consisted of 26 nodes in the input layer (Table 3), 10 nodes in the hidden layer, and 1 output node encoded as to whether the patient had metastases. The input nodes were fed with 26 features describing the hot spots found in the anterior and posterior images. The 4 hot spots in each of the images with the highest outputs from the hot spot networks were used as inputs to the bone scan networks. The neural networks presented an output between 0 and 1 for each test case. The output value reflects the assessment of the neural network of likelihood for the patient having bone metastases, and one approach is to present that value (e.g., 0.23) as the CAD advice to the physician. In clinical routine, however, physicians generally use phrases such as "cannot be ruled out," "probable," or "definite metastases" to report likelihood and, therefore, we decided to categorize the network output by the use of threshold values. Patient examinations with output values above a threshold in the interval between 0 and 1 were classified as having bone metastasis. The threshold was selected to achieve a sensitivity of 95% in the training group. Two other thresholds, 1 below and 1 above the first threshold, were used to categorize the network classification into the same 4-grade scale as was used as the gold standard. The distribution of the gold standard classifications in the test group showed that 90% of the cases (53/59) were classified as grade 1 or 4, and 10% (6/59) were classified in the 2 middle categories. Therefore, the thresholds were selected to classify 5% of the training cases as grade 2 and 5% as grade 3. After the training processes of both sets of neural networks and the selection process of thresholds, the CAD system was applied to the images of the test set (Figs. 2 and 3).

Statistical Methods

The percentage agreement (PA) and the κ-coefficient (which measures agreement beyond that expected by chance) were calculated. The classifications made by the CAD system for the 59 patients in the test group were compared with the gold standard.

Disagreement between the CAD classifications and the gold standard could be systematic or random. To quantify the disagreement between paired ordered categoric classifications, a method reported by Svensson et al. (*21,22*) was used. Two types of systematic variation are possible; the first is due to overestimation or underestimation of the classifications, and the second is due to concentration of the classifications. Systematic overestimation or underestimation occurs when the CAD classifies cases as being more, or less, abnormal than does the gold standard. Systematic concentration occurs, for example, when the CAD uses the middle part of the 4-point scale ("cannot be ruled out" or "probable") more often than does the gold standard, which uses the grades "absence of bone metastases" or "definitely bone metastasis" more often. Overestimation or underestimation is reflected by the variable relative position (RP), and concentration is exhibited by the variable relative concentration (RC). The possible values for RP and RC range from −1 to 1, with a value of 0 indicating that no systematic disagreement

TABLE 3
Bone Scan Features

Bone scan feature	No. of features	Example
Highest neural network outputs of hot spots	8	4 highest-output hot spots in both anterior and posterior views
Ratio of number of hot spots with high neural network output in region to total number of hot spots	14	Head, shoulder, arm, spine, ribs, pelvis, and lower limb in anterior and posterior views
Ratio of total area of hot spots to body area	2	Anterior and posterior
Ratio of total area of hot spots with high neural network output to body area)	2	Anterior and posterior
Total	26	

is present. A positive RP value reflects systematic overestimation of the classifications, and a negative RP value reflects a systematic underestimation. The RC value is positive if systematic concentration to the middle of the 4-point scale is present, whereas a negative RC value reveals systematic concentration to the extremity.

The pattern of random differences was quantified using the variable relative rank variance (RV). The possible values for RV are between 0 and 1, with 0 indicating no random contribution.

Confidence intervals for PA, κ, RP, RC, and RV were calculated using 10,000 bootstrap replicates (*23*). The exact (Clopper–Pearson) test was used to calculate the 95% confidence interval for sensitivity, specificity, and accuracy.

RESULTS

The CAD system made correct classifications for 19 of the 21 patients with bone metastases, showing a sensitivity of 90%. True-negative interpretations were made for 34 of the 38 patients classified as not having bone metastases by the gold standard, resulting in a specificity of 89%. Two of 4 false-positive cases had fractures, one in the rib-costal cartilage regions and the other in the rib-vertebra region of the thoracic spine. The other 2 had degeneration with high-intensity hot spots, one in the lower lumbar spine and the other in the acetabulum region.

A comparison of the classifications by the CAD system and the gold standard is shown in Table 4; the resulting PA and κ-values were 76% and 0.58, respectively (Table 5). The contribution of systematic variations in position (RP) and concentration (RC) was small; that is, CAD, compared with the gold standard, did not over- or underestimate the classifications, nor did CAD concentrate the classifications to a certain part of the 4-point scale (Table 5). The main reason for the disagreement between the CAD system and the gold standard was because of random errors. As shown in Table 4, the CAD system classified 2 patients as being 3 grades more pathologic than did the gold standard (grade 4 vs. grade 1) and 2 patients as 3 grades less pathologic than did the gold standard (grade 1 vs. grade 4).

DISCUSSION

Main Findings

A completely automated CAD system based on image-processing techniques and artificial neural networks can be used to classify bone scans for the presence or absence of metastases. Our results showed a high detection rate, with a sensitivity of 90% in the test group at a specificity of 89%. Our current results are an improvement, compared with our previously presented method, which showed a specificity of 74% at the same level of sensitivity. In the present study, new algorithms for skeleton segmentation, hot spot detection, feature extraction, and neural networks were developed, and a larger training group was used in the training process, which enabled the program to better

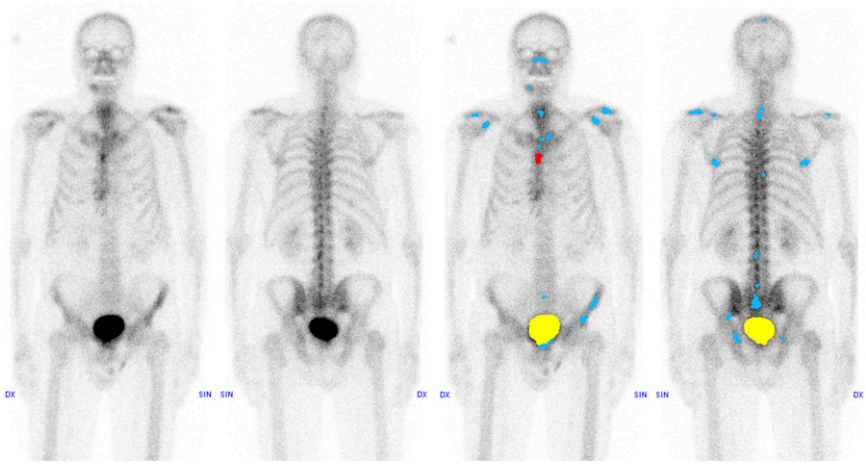

FIGURE 2. Images showing 70-y-old man with prostate cancer. Increased radiotracer uptake can be seen in right part of mandible, most probably because of bad teeth, and in sternum secondary to sternotomy. Artificial neural networks classifying hot spots separately indicate that uptake in sternum could be metastasis, but neural networks classifying complete examination, considering all hot spots, correctly report ''absence of bone metastases.'' Suggestive metastases are marked in red, and symmetric or benign radiotracer uptake in blue; bladder is yellow.

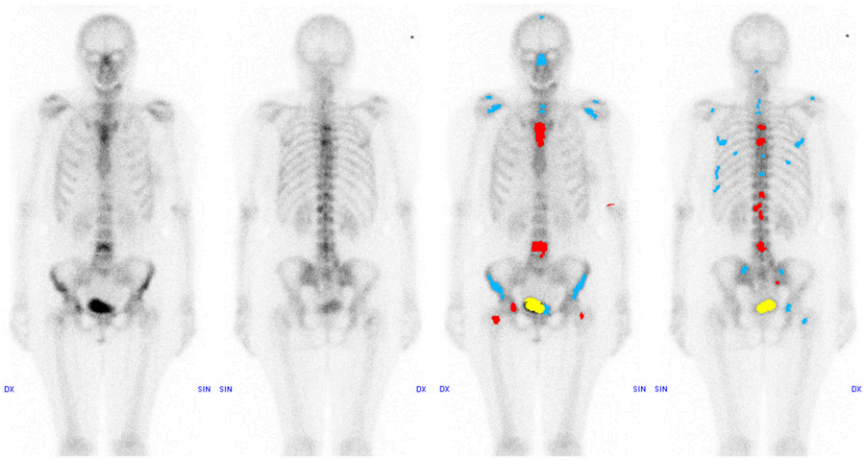

FIGURE 3. Images showing 56-y-old woman with breast cancer. Multiple focal increases of pathologic radiotracer uptake can be seen. CAD system correctly reports whole-body bone scan examination as showing "definite presence of bone metastases." Suggestive metastases are marked in red, and symmetric or benign radiotracer uptake in blue; bladder is yellow.

differentiate normal or benign uptake patterns from pathologic processes.

Our sensitivity and specificity were also higher than those found by Sajn et al. (*24*) (79.6% and 85.4%, respectively), who presented an automatic method for analysis of whole-body bone scans. An explanation for this could be that we included a larger number of patients in the training process and used different techniques in the development of the CAD system.

In our previous, nationwide survey, 37 observers interpreted the same 59 bone scans as used in the present study and showed a mean sensitivity of 77% (range, 52%−100%) at an average specificity of 96% (*7*). The main problem in the interpretations of bone scans was the false-negative errors. The advantage with a CAD system would be to increase the detection rate of the physician and consequently minimize the risk that abnormal findings would be overlooked. For this reason, we tried to adjust the CAD system to achieve a high sensitivity. The agreement for the CAD system, compared with the final clinical assessment, expressed as PA and κ, was 76% and 0.58, respectively, whereas the variation of the observers was higher (PA and κ, 64% and 0.48, respectively) (*7*). A CAD system, therefore, has the potential to decrease interobserver variation.

When the findings of the 37 observers were compared with the final clinical assessment, there was some contribution of systematic overestimation or underestimation of the classifications, but the main reason for the disagreement was that the observers concentrated more on the middle of the 4-point scale (grades 2 and 3), in contrast to the gold standard. In general, the established guidelines that aim for structured reporting state that the final report of an investigation should possess clarity and emphasize whether a study is "normal" or "abnormal." The middle categories, such as "cannot be ruled out" or "probable," should be used as infrequently as possible but may allow for communication of interpretive uncertainty (*25*). Therefore, we constructed the CAD system to use the middle of the 4-point scale in only 10% of the cases and use the 2 categories that are most useful to the referring physician (i.e., "absence of bone metastases" or "definite presence of bone metastases") in 90% of the cases.

TABLE 4
Frequency Table Showing Classifications Made by CAD System, Compared with Gold Standard

	Gold standard				
CAD	Grade 1	Grade 2	Grade 3	Grade 4	Total
Grade 4	2	1	0	18	21
Grade 3	0	1	0	1	2
Grade 2	3	0	0	0	3
Grade 1	27	4	0	2	33
Total	32	6	0	21	59

PA = 76%; κ = 0.58; RP = −0.007; RC = −0.02; RV = 0.013.

Clinical Implications

The concept with CAD systems in general is to take into account equally the role of the physician and that of computers by combining the competence of the physician in interpreting medical images with the high capability of the computer for detecting abnormalities. In mammography, investigators have shown an increase in the detection

TABLE 5
Performance of CAD System, Compared with Gold Standard

Parameter	CAD vs. gold standard
PA	76% (64–86)
κ	0.58 (0.40–0.75)
Systematic difference	
RP	−0.007 (−0.103–0.087)
RC	−0.02 (−0.148–0.099)
Random difference (RV)	0.013 (0.002–0.04)
Sensitivity	90% (70%−99%)
Specificity	89% (75%−97%)
Accuracy	89% (82%−98%)

95% confidence interval is given in parentheses.

rate of breast cancer with CAD (*10–12*). Freer et al. found that 19.5% more cancers were detected when physicians interpreted the images with the advice of the computer (*12*). Cupples et al. reported a 164% increase in the detection of small breast cancers and, in addition, a reduction of 5.3 y in the mean age at the time of detection when CAD was used (*11*). Similar CAD systems in other imaging fields have also been shown to increase the sensitivity by assisting less experienced physicians in the interpretation of CT colonography (*13*) and to improve the quality of image reporting in myocardial perfusion imaging (*14*).

A CAD system for bone scintigraphy, therefore, has the potential to increase the physician's performance in finding bone metastases and also reduce interobserver variation. To investigate whether physicians benefit from the advice of our CAD system, we have invited the 37 observers who participated in our nationwide survey (*7*) to make a second interpretation of the same bone scans, this time with the CAD system. Previous studies in myocardial scintigraphy have shown decreased interobserver variation among readers when CAD was used in the classification of the images (*14,26*). CAD systems can be used to shorten the learning curve needed to achieve high-quality reports and minimize errors due to reading fatigue or interruptions during interpretations at a busy practice.

To improve diagnostic accuracy in patients in whom whole-body bone scans fail to demonstrate metastases, other imaging modalities such as SPECT/CT or MRI can be of value. The CAD system presented here is, at the current stage, designed only to assist physicians in the interpretation of whole-body bone scans. Future CAD systems may be able to analyze combinations of image series from the same patient, such as a whole-body bone scan and a SPECT/CT study of the pelvic region.

Study Limitations

Ideally, the databases used for these types of studies should be at least on the order of hundreds of cases, including representative cases found in clinical routine, and have an accurate and independent gold standard method. These features are, however, difficult to combine in 1 study. A limitation of the present study is that we used the final clinical assessment of an experienced physician as the gold standard, based on, in addition to the bone scans, the follow-up scans, the patient's computerized medical record including the results of laboratory tests, and available diagnostic images (MRIs, CTs, or radiographs) for almost 5 y of follow-up for the test group. Histologic verification for each hot spot found in the bone scans would have been a more accurate gold standard, but this type of gold standard is difficult to obtain.

A retrospectively acquired database from only 1 hospital was used for the present work. An advantage of this approach was that the gold standard classifications of the images in the large database could be based on a follow-up duration of almost 5 y. More recent data or prospectively selected patients would have shortened the follow-up duration or delayed the study considerably. Studies including cases from multiple centers would be of value to evaluate if the CAD system showed the same performance on images acquired with different γ-cameras or different protocols, but that was not within the scope of the present study. A multicenter study will also address issues such as differences in interpretive style at different clinics and differences in incidence of metastatic disease.

Our test database was selected from patients with at least 1 follow-up scan. This introduces a risk of selection bias, and we tried to minimize that by selecting the same relation between "benign findings," "difficult cases," and "obvious metastases" as was found in a previous study of consecutive cases.

CONCLUSION

A completely automated CAD system can be used to detect metastases in bone scans. Application of the method as a clinical decision support tool appears to have significant potential.

ACKNOWLEDGMENTS

The study was approved by the Research Ethics Committee at Gothenburg University. This study was supported by grants from the Swedish Research Council (2007-2488), Stockholm, Sweden. Lars Edenbrandt and Mattias Ohlsson are shareholders in EXINI Diagnostics AB, which owns a CAD system for the interpretation of bone scans.

REFERENCES

1. Sergieva S, Kirova G, Dudov A. Current diagnostic approaches in tumor-induced bone disease. *J BUON.* 2007;12:493–504.
2. Abuzallouf S, Dayes I, Lukka H. Baseline staging of newly diagnosed prostate cancer: a summary of the literature. *J Urol.* 2004;171:2122–2127.
3. Bombardieri E, Gianni L. The choice of the correct imaging modality in breast cancer management. *Eur J Nucl Med Mol Imaging.* 2004;31(suppl):S179–S186.
4. Myers RE, Johnston M, Pritchard K, Levine M, Oliver T, and the Breast Cancer Disease Site Group of the Cancer Care Ontario Practice Guidelines Initiative. Baseline staging tests in primary breast cancer: a practice guideline. *CMAJ.* 2001;164:1439–1444.
5. Rossing N, Munck O, Nielsen SP, Andersen KW. What do early bone scans tell about breast cancer patients? *Eur J Cancer Clin Oncol.* 1982;18:629–636.
6. Peters AM, Bomanji J, Costa DC, et al. Clinical audit in nuclear medicine. *Nucl Med Commun.* 2004;25:97–103.
7. 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;35:1464–1472.
8. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph.* 2007;31:198–211.
9. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology.* 2000;215:554–562.
10. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. *Radiology.* 2005;236:451–457.
11. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. *AJR.* 2005;185:944–950.
12. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology.* 2001;220:781–786.
13. Baker ME, Bogoni L, Obuchowski NA, et al. Computer-aided detection of colorectal polyps: can it improve sensitivity of less-experienced readers? Preliminary findings. *Radiology.* 2007;245:140–149.

14. Tägil K, Bondouy M, Chaborel JP, et al. A decision support system improves the interpretation of myocardial perfusion imaging. *Eur J Nucl Med Mol Imaging*. 2008;35:1602–1607.
15. Sadik M, Jakobsson D, Olofsson F, Ohlsson M, Suurkula M, Edenbrandt L. A new computer-based decision-support system for the interpretation of bone scans. *Nucl Med Commun.* 2006;27:417–423.
16. Cootes TF, Hill A, Taylor CJ, Haslam J. Use of active shape models for locating structures in medical images. *Image Vis Comput J.* 1994;12:355–366.
17. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet.* 1995;346:1075–1079.
18. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet.* 1995;346:1135–1138.
19. Dybowski R, Gant V. Artificial neural networks in pathology and medical laboratories. *Lancet.* 1995;346:1203–1207.
20. Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing*. Cambridge, MA: MIT Press; 1986:1–2.
21. Svensson E, Holm S. Separation of systematic and random differences in ordinal rating scales. *Stat Med.* 1994;13:2437–2453.
22. Svensson E, Starmark JE, Ekholm S, von Essen C, Johansson A. Analysis of interobserver disagreement in the assessment of subarachnoid blood and acute hydrocephalus on CT scans. *Neurol Res.* 1996;18:487–494.
23. Iwi G, Millard RK, Palmer AM, Preece AW, Saunders M. Bootstrap resampling: a powerful method of assessing confidence intervals for doses from experimental data. *Phys Med Biol.* 1999;44:N55–N62.
24. Sajn L, Kononenko I, Milcinski M. Computerized segmentation and diagnostics of whole-body bone scintigrams. *Comput Med Imaging Graph.* 2007;31:531–541.
25. Hendel RC, Wackers FJ, Berman DS, et al. American Society of Nuclear Cardiology consensus statement: reporting of radionuclide myocardial perfusion imaging studies. *J Nucl Cardiol.* 2003;10:705–708.
26. Lindahl D, Lanke J, Lundin A, Palmer J, Edenbrandt L. Improved classifications of myocardial bull's-eye scintigrams with computer-based decision support system. *J Nucl Med.* 1999;40:96–101.