


Diagnostic accuracy of an artificial neural network compared with statistical quantitation of myocardial perfusion images: a Japanese multicenter study

Kenichi Nakajima¹  · Takashi Kudo² · Tomoaki Nakata³ · Keisuke Kiso⁴ · Tokuo Kasai⁵ · Yasuyo Taniguchi⁶ · Shinro Matsuo¹ · Mitsuru Momose⁷ · Masayasu Nakagawa⁸ · Masayoshi Sarai⁹ · Satoshi Hida¹⁰ · Hirokazu Tanaka¹¹ · Kunihiro Yokoyama¹² · Koichi Okuda¹³ · Lars Edenbrandt¹⁴

Received: 13 August 2017 / Accepted: 6 September 2017 / Published online: 26 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract

Purpose Artificial neural networks (ANN) might help to diagnose coronary artery disease. This study aimed to determine whether the diagnostic accuracy of an ANN-based diagnostic system and conventional quantitation are comparable.

Methods The ANN was trained to classify potentially abnormal areas as true or false based on the nuclear cardiology expert interpretation of 1001 gated stress/rest ^{99m}Tc-MIBI images at 12 hospitals. The diagnostic accuracy of the ANN was compared with 364 expert interpretations that served as the gold standard of abnormality for the validation study.

Conventional summed stress/rest/difference scores (SSS/SRS/SDS) were calculated and compared with receiver operating characteristics (ROC) analysis.

Results The ANN generated a better area under the ROC curves (AUC) than SSS (0.92 vs. 0.82, $p < 0.0001$), indicating better identification of stress defects. The ANN also generated a better AUC than SDS (0.90 vs. 0.75, $p < 0.0001$) for stress-induced ischemia. The AUC for patients with old myocardial infarction based on rest defects was 0.97 (0.91 for SRS, $p = 0.0061$), and that for patients with and without a history of revascularization based on stress defects was 0.94 and 0.90 ($p = 0.0055$ and $p < 0.0001$ vs. SSS, respectively). The SSS/SRS/SDS steeply increased when ANN values (probability of abnormality) were >0.80 .

Conclusion The ANN was diagnostically accurate in various clinical settings, including that of patients with previous myocardial infarction and coronary revascularization. The ANN could help to diagnose coronary artery disease.

✉ Kenichi Nakajima
nakajima@med.kanazawa-u.ac.jp

- ¹ Kanazawa University Hospital, Kanazawa, Japan
- ² Nagasaki University Hospital, Nagasaki, Japan
- ³ Hakodate Goryoukaku Hospital, Hakodate, Japan
- ⁴ National Cerebral and Cardiovascular Center, Suita, Japan
- ⁵ Tokyo Medical University Hachioji Medical Center, Hachioji, Japan
- ⁶ Hyogo Brain and Heart Center, Himeji, Japan
- ⁷ Tokyo Women's Medical University, Tokyo, Japan
- ⁸ Akita City Hospital, Akita, Japan
- ⁹ Fujita Health University Hospital, Toyoake, Japan
- ¹⁰ Tokyo Medical University Hospital, Tokyo, Japan
- ¹¹ Tokyo Medical University Ibaraki Medical Center, Ibaraki, Japan
- ¹² Public Central Hospital of Matto Ishikawa, Hakusan, Japan
- ¹³ Kanazawa Medical University, Uchinada, Kahoku, Japan
- ¹⁴ University of Gothenburg, Gothenburg, Sweden

Keywords Artificial intelligence · Diagnostic imaging · Coronary artery disease · Nuclear cardiology · Computer-aided diagnosis

Introduction

Myocardial perfusion images have been interpreted based on an integrated understanding of myocardial perfusion distribution at stress and rest, and the difference between these two conditions is used to differentiate stress-induced ischemia and infarction. Although visual interpretation is always the first step, nuclear cardiology trainees need to become competent in reading single-photon emission

computed tomography (SPECT) images [1, 2]. On the other hand, quantitative analysis has usually consisted of myocardial segmentation into walls and segments, and the extent and severity of perfusion defects are expressed in terms of the size of an abnormal area and defect scores [3, 4]. Most nuclear cardiology software has taken this type of quantitative approach, and it has aided the visual evaluation of images. One advantage of quantitative analysis is that it has led nuclear cardiologists and nuclear medicine physicians to standardize interpretations among physicians and hospitals [5]. Perfusion defects and ischemia are quantified not only for diagnosis, but for the prognostic evaluation of future cardiac events [6].

Artificial intelligence including artificial neural networks (ANN), deep learning and machine learning has recently been applied to nuclear cardiology [7–11]. Artificial intelligence learns how to appropriately interpret images based on a large number of SPECT images with definitive diagnoses based on nuclear cardiology expert interpretations as the teacher. The ANN approach includes an appropriate combination of features of abnormalities, and it differs from statistical approaches using regional count distribution, which is typically based on means and deviations. The ANN approach simulates how experts learn the art of interpretation. Therefore, stenosis of coronary arteries is not a truth or a standard; the integrated judgment of abnormalities by experts is the standard. Based on such repetitive learning processes, artificial intelligence could provide support for decision-making [12, 13].

The ANN system trained in Sweden using Swedish databases has been applied to Swedish and Japanese MPI SPECT studies related to an ANN diagnosis from myocardial perfusion images (MPI) [9, 10]. Although the diagnostic accuracy was comparable to that of defect scoring methods, additional clinical experience has revealed that slight abnormalities seem to be underestimated, and thus improvements in diagnostic accuracy are needed.

The purpose of this multicenter project was, therefore, to retrain the ANN system using a new Japanese MPI database and to develop a high-performance diagnostic system for coronary artery disease. The diagnostic ability of the ANN system was further compared with expert readings of another dataset from a multicenter validation database.

Methods

Patients for creating a training database

Data from 1001 patients who underwent stress/rest MPI were accumulated from 12 institutions in Japan. The entry

criteria included confirmed and suspected coronary artery disease, and a confirmed diagnosis of the presence or absence of myocardial ischemia and/or infarction. The exclusion criteria comprised patients aged <20 years, arrhythmia causing inappropriate electrocardiographic gating, left bundle branch block (to avoid false positive defects), idiopathic and other types of severe cardiomyopathy, as well as moderate or severe valvular heart diseases. Patient data included age, sex, height, weight, risk factors, results of coronary angiography or coronary CT angiography (CCTA), and a history of percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG). We tried to include either stress-induced ischemia or infarction in at least 50% of the patients when the MPI data were collected to train the ANN system.

Patients for validation study

After the initial creation of the database, additional patients were registered using criteria similar to that of the training database. Data from 364 patients collected from nine hospitals served as the validation dataset.

Myocardial perfusion imaging

Datasets were derived from all patients using exercise or adenosine stress MPI with electrocardiographic gating and non-gating (Table 1). ^{99m}Tc -labeled hexakis-2-methoxyisobutylisonitrile (MIBI) was used in all participating institutions, and the second injection dose was 2- to 3-fold higher than the first, resulting in a total administered dose of 740 to 1110 MBq. All institutional protocols for stress and image acquisition followed standard exercise and adenosine stress MPI protocols [1, 14]. All image data were anonymized in the "Digital Imaging and Communication in Medicine" (DICOM) format at each institution and uploaded into a custom-designed server. The only patient information in the image database in the server was serial number, age and sex.

Training the artificial neural network

The left ventricle (LV) was segmented using a three-dimensional heart-shaped active-shape model reconstructed with short-axis slice images [15, 16]. A specific software algorithm was used for this analysis (cardioREPO software, FUJIFILM RI Pharma Co. Ltd., Tokyo, Japan; EXINI Diagnostics AB, Lund, Sweden). Areas of possible perfusion abnormalities in stress and rest images (stress and rest defects, respectively) were segmented using a method that mimicked analyses by physicians who visually interpret defects. Candidate regions with abnormalities were identified using deformable models of circular

Table 1 Patient demographics in the training and validation databases

Items	Training database: value, %, means \pm SD (range)	Validation database: value, %, means \pm SD (range)	p
Number of participants	1001	364	–
Age (years)	69 \pm 10	71 \pm 10	0.0011
Male (%)	75%	73%	0.48
Height (male, cm)	165 \pm 7	165 \pm 7	0.98
Weight (male, kg)	66 \pm 12	67 \pm 11	0.48
Body mass index (male, kg/m ²)	24 \pm 4	24 \pm 3	0.64
Height (female, cm)	151 \pm 7	151 \pm 6	0.12
Weight (female, kg)	54 \pm 11	55 \pm 10	0.18
Body mass index (female, kg/m ²)	24 \pm 4	24 \pm 4	0.46
Pharmacological stress (%)	70%	82%	0.0001
Number of vessel stenosis \geq 75% (1-, 2- and 3-vessel disease)	391 (156, 123, 112)	225 (78, 82, 65)	0.0001
Hypertension (%)	73%	75%	0.40
Diabetes mellitus (%)	47%	39%	0.019
Dyslipidemia (%)	65%	66%	0.79
History of myocardial infarction (%)	27%	31%	0.17
History of PCI (%)	38%	39%	0.87
History of CABG (%)	4%	3%	0.53
ANN stress defect	–	0.63 \pm 0.37	–
Presence of stress defect (%)	71%	73%	0.54
ANN ischemia	–	0.51 \pm 0.34	–
Presence of ischemia (%)	59%	59%	1.00
ANN rest defect	–	0.54 \pm 0.38	–
Presence of rest defect (%)	57%	56%	0.76
Summed stress score	9.5 \pm 9.9 (0–53)	9.5 \pm 9.8 (0–52)	1.00
Summed rest score	7.0 \pm 8.6 (0–45)	7.1 \pm 8.8 (0–49)	0.85
Summed difference score	3.3 \pm 3.9 (0–26)	3.1 \pm 3.3 (0–21)	0.38
Rest end-diastolic volume (mL)	105 \pm 38 (38–325)	107 \pm 46 (34–341)	0.42
Rest end-systolic volume (mL)	38 \pm 29 (5–250)	37 \pm 29 (3–246)	0.57
Rest ejection fraction (%)	67 \pm 13 (19–97)	63 \pm 14 (25–92)	0.0001

Abbreviations: ANN, artificial neural network (probability of abnormality in this table); CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention

topology based on an active contour framework, and fit to the image based on edge and intensity information [17]. Stress images were subtracted from rest images to identify stress-induced ischemia. Appropriate features were extracted using the characteristics of shape, extent, location, count, perfusion homogeneity, regional motion, wall thickening and sex. The best combination of features was determined for stress, rest and difference datasets, and these features served as inputs for the ANN analysis.

Nuclear cardiology experts at each institution judged all regions in MPI data as abnormal (true) or normal (false). The software algorithm identified 5685 candidate regions, but whether or not an abnormality is true was determined by nuclear cardiologists and nuclear physicians. The interpreters used original myocardial stress and rest images and polar maps, but not quantitative data,

such as defect scores for detecting abnormalities. After all data were accumulated at the core center, six nuclear cardiology experts reconfirmed the appropriateness of the judgment without clinical information. Two experts interpreted the data during this process and, if necessary, modified the judgment with referral to other expert opinions to reach consensus. The final judgments of the candidate regions made at the core center served as the gold standard with which to train the ANN system.

Based on the initial ANN training, discordance between visual and ANN findings were again selected and used for appropriately adjusting diagnostic thresholds. In the final output of the software, the ANN value indicated the probability of an abnormality under conditions of 0.0 (definitely normal), 0.5 (borderline) and 1.0 (definitely abnormal) for stress defects, rest defects and induced ischemia.

Processing by the artificial neural network in the software package

Four DICOM stress/rest gated and ungated images were transferred from the nuclear medicine dedicated computer system to an on-line connected (off-line acceptable) Windows computer with the aid of a custom-made launcher. After selecting a patient dataset, all processing steps including LV segmentation, polar map generation, detection of abnormal regions, judgment of the abnormality with the ANN system, scoring based on Japanese Society of Nuclear Medicine (JSNM) working group normal databases [18, 19], LV functional analysis including wall motion, volume, ejection fraction and diastolic function, and phase dyssynchrony analysis were performed within 10 s (64-bit operating system, Windows 10) [10, 16, 19–21]. This processing was automatic without operator interaction, but LV contour detection can be modified if required. The ANN system displays a list of regions including the location, extent, severity and ischemia/infarct with the probability of an abnormality (ANN value) in each region.

Defect scoring

Myocardial perfusion defects were scored using a standard 17-segment model, and abnormalities were defined as normal, slightly, moderately and severely decreased, and defective (scores 0, 1, 2, 3, and 4, respectively) in each segment [3, 4]. Scores were validated in automated analysis using cardioREPO software, and the correlation was good compared with that determined using QPS (Cedars Sinai Medical Center, Los Angeles, CA, USA) [10]. The maximum score was 68 points for summed stress/rest/difference scores (SSS/SRS/SDS). Normal MPI databases were constructed according to the JSNM working group database that includes stress/rest normal SPECT data with non-attenuation-corrected Anger camera images [18, 19].

Statistics

Data are shown as means \pm standard deviation (SD). Differences between groups were assessed using the one-way analysis of variance, Student's *t* tests and F tests. A contingency table was analyzed with Fisher exact tests. Receiver operating characteristics (ROC) were analyzed, and the area under the curve (AUC) was calculated. The cutoff was determined using either an ANN value (probability of abnormality) of 0.50 or the highest value of (sensitivity + specificity - 1) depending on the analysis. The statistics software was JMP version 12 (SAS Institute Inc., Cary, NC, USA), and Mathematica 12 (Wolfram Research Inc., Champaign, IL, USA) was also used for some mathematical calculations. A *p* value <0.05 was considered significant.

Results

Training database

Table 1 summarizes the characteristic of the training databases that included stress defects (71%), rest defects (57%) and ischemia (59%). Old myocardial infarction (OMI) and revascularization were found in 27% and 42% of the patients, respectively. The AUC of the training database calculated to obtain results concordant with those of the experts were 0.912 (sensitivity 86%, specificity 77%, accuracy 82%) for stress defects, 0.834 (sensitivity 83%, specificity 74%, accuracy 78%) for rest defects and 0.888 (sensitivity 82%, specificity 77%, accuracy 80%) for ischemia.

Validation database

Background conditions and associated diseases did not significantly differ between the validation and the training databases for most of the variables (Table 1). Rates of stress defects, rest defects and ischemia were similar. However, the validation database had a higher incidence of pharmacological stress and of coronary stenosis, and a lower ejection fraction than the training database.

Interpretation of the ANN system

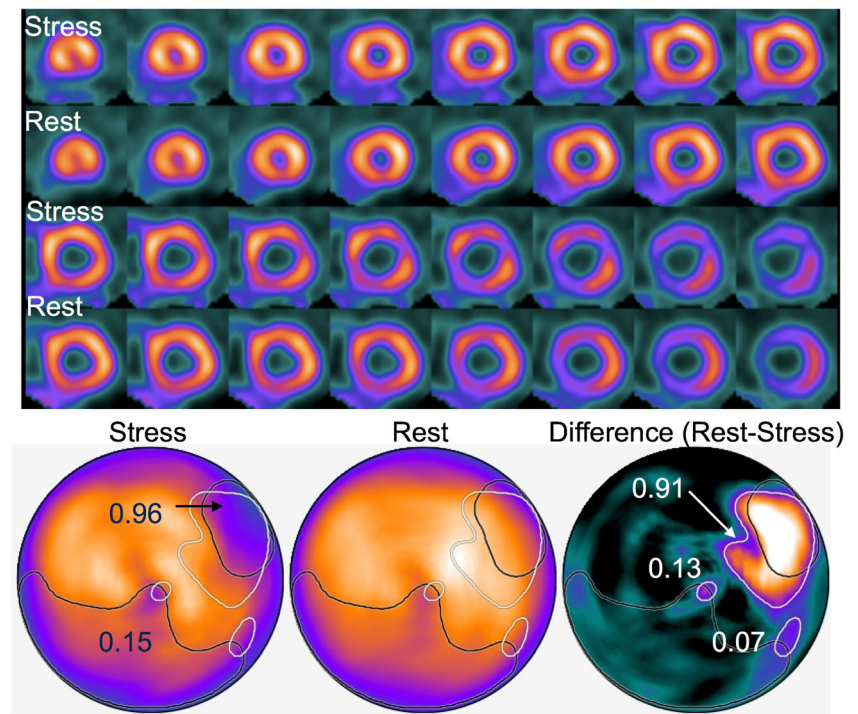
Figure 1 shows the standard output generated using the ANN. This patient had 99% stenosis of the circumflex coronary artery (#11). Six months after stenting he developed chest pain, which was assessed by MPI. The ANN judged the high lateral region as abnormal in both stress and subtraction images with high probability.

ROC analysis for scores and the ANN

The AUC values of the stress defect were 0.815 and 0.916 ($\chi^2 = 34.4$, $p < 0.0001$) for the scoring system and the ANN, respectively (Fig. 2). Similarly, the AUC was 0.857 vs. 0.926 ($\chi^2 = 21.9$, $p < 0.0001$) for rest defects, and 0.754 vs. 0.895 ($\chi^2 = 34.2$, $p < 0.0001$) for ischemia, respectively, indicating a higher AUC for the ANN compared with the scoring method.

Average scores for stress defects were 13.2 ± 10.6 vs. 3.9 ± 4.3 ($p < 0.0001$, F ratio = 99) in patients with and without stress defects, respectively (Fig. 3). The values for stress defects determined using the ANN in patients with and without stress defects were 0.84 ± 0.24 vs. 0.31 ± 0.29 ($p < 0.0001$, F ratio = 347). Similarly, although both scores and ANN values under conditions of rest, stress and ischemia significantly differed between positive and negative expert readings ($p < 0.0001$), the F ratios were higher when determined by the ANN than the scoring method.

Fig. 1 Myocardial perfusion study and artificial neural network (ANN) analysis of 70-year-old man after percutaneous coronary intervention to the left circumflex coronary artery. Numbers indicate probability of abnormality. Basal lateral ischemia is evident in short-axis images (upper panel), whereas the ANN system identified abnormality in stress (probability) and subtraction (probability) images with probabilities of 0.96 and 0.91, respectively. Other regions with probability of <0.5 were considered insignificant



ROC analysis in subgroups of patients

The ability of the ANN to detect abnormalities in patient subgroups with and without OMI or revascularization was

generally good, with an AUC of ≥ 0.89 (Table 2). The AUC values for the ANN to detect stress defects in patients with and without OMI were 0.976 and 0.892, respectively. In patients with OMI, the AUC values to detect rest defects using SRS and

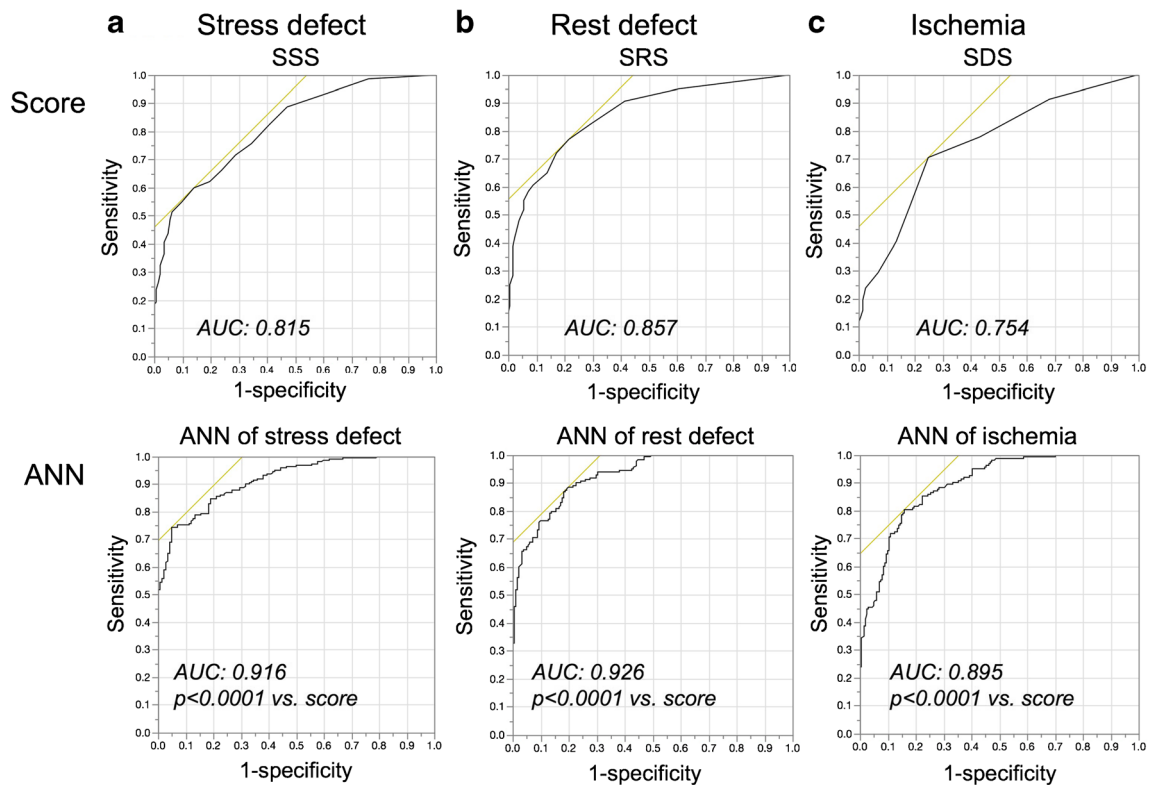


Fig. 2 Receiver operating characteristics (ROC) analysis of stress defect (a), rest defect (b) and ischemia (c) using the scoring method (upper panel) and the artificial neural network (ANN; lower panel). All areas under ROC curves (AUC) were higher for the ANN ($p < 0.0001$)

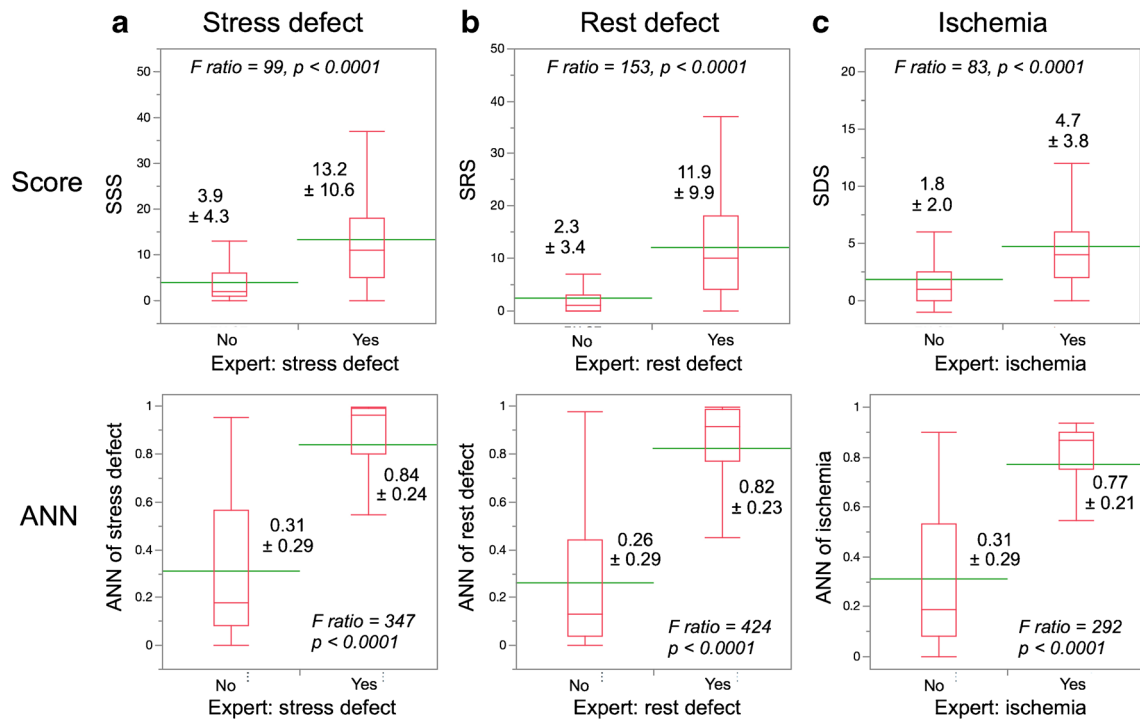


Fig. 3 Comparison of the scoring method (upper panel) and artificial neural network (ANN; lower panel) based on expert judgments. Positive and negative judgments significantly differed in all comparisons of stress defects (a), rest defects (b) and ischemia (c)

the ANN were 0.905 and 0.974, respectively. The AUC values for detecting stress defects with and without a history of revascularization were 0.939 and 0.900, respectively. The AUC in all subgroups was higher for the ANN than the scoring method.

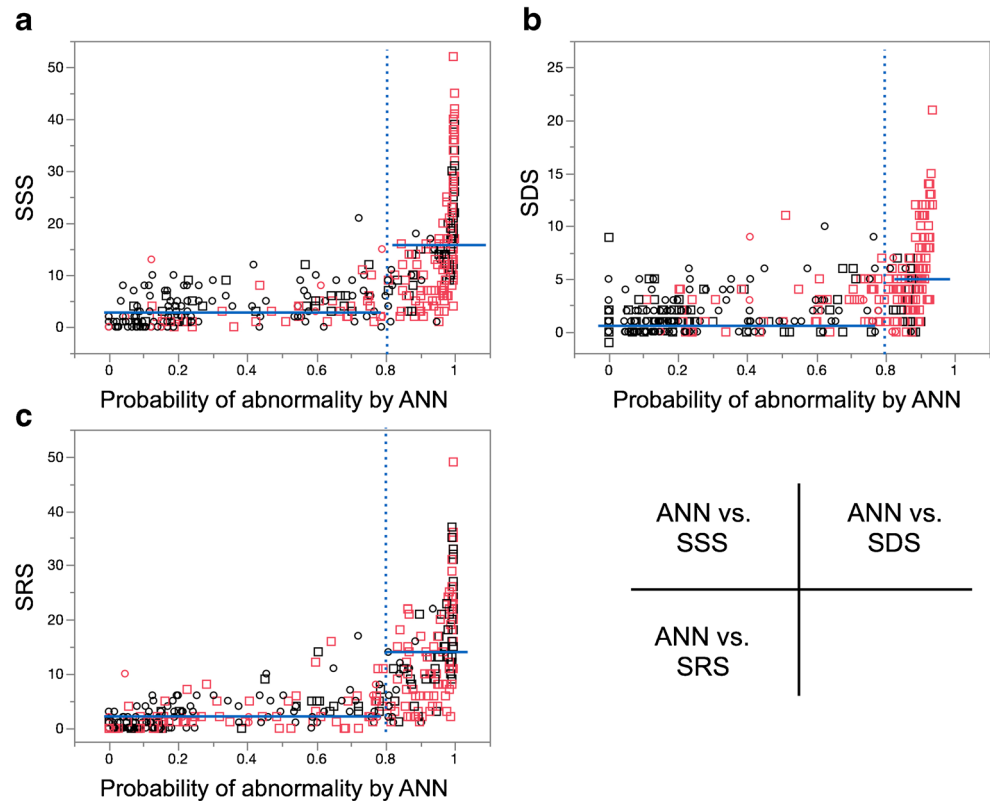
According to $\geq 75\%$ of coronary stenosis, we divided the patients into groups according to their having 0–1-, or 2–3-vessel disease. The AUC values were 0.895 and 0.785 when determined using the ANN and SSS, respectively ($\chi^2 = 24.2$,

Table 2 Receiver operating characteristics (ROC) analysis of subgroups in the validation study

	ANN/score	AUC	Standard error	Lower 95%	Upper 95%	p vs. score
Stress defect with and without OMI						
Without OMI	ANN	0.892	0.020	0.846	0.925	<0.0001
	SSS	0.771	0.030	0.708	0.824	
With OMI	ANN	0.976	0.014	0.929	0.992	0.0036
	SSS	0.890	0.033	0.807	0.940	
Rest defect						
With OMI	ANN	0.974	0.014	0.925	0.991	0.0061
	SRS	0.905	0.029	0.831	0.949	
Stress defect with and without history of revascularization						
No revascularization	ANN	0.900	0.020	0.854	0.932	<0.0001
	SSS	0.781	0.030	0.716	0.835	
Revascularization	ANN	0.939	0.019	0.889	0.967	0.0055
	SSS	0.863	0.031	0.790	0.913	
Ischemia with and without history of revascularization						
No revascularization	ANN	0.898	0.020	0.853	0.931	<0.0001
	SDS	0.771	0.032	0.703	0.827	
Revascularization	ANN	0.889	0.027	0.823	0.932	0.0002
	SDS	0.727	0.042	0.636	0.802	

Abbreviations: ANN, artificial neural network; AUC, area under the curve; OMI, old myocardial infarction; ROC, receiver-operating characteristics; SDS, summed difference score; SRS, summed rest score; SSS, summed stress score

Fig. 4 Relationship between scoring methods and probability of abnormality judged by the ANN. Dotted vertical lines indicate probability of 0.8, and blue lines indicate mean values for probabilities of <0.8 and ≥ 0.8 . Squares and circles denote positive and negative stress defect, respectively by expert interpretations. Red and black marks denote positive and negative ischemia, respectively by expert interpretations



$p < 0.0001$) for patients with 0–1-vessel disease and 0.945 and 0.838, respectively, for patients with 2–3-vessel disease ($\chi^2 = 11.0$, $p = 0.0009$), indicating that the ANN had better diagnostic accuracy.

Relationship between scores and ANN values

Figure 4 shows a non-linear relationship between scores and the ANN values in the validation group. Although the threshold of abnormality using the ANN was a probability of 0.5, the scores steeply increased when the ANN probability was ≥ 0.8 . The SSS was 3.4 ± 3.5 in the range of ANN probability <0.80 , but 15.8 ± 10.3 when in the probability range of ≥ 0.80 ($p < 0.0001$) determined by the ANN. The SDS values were 1.8 ± 4.7 and 4.7 ± 3.8 when the ANN probability was <0.8 and ≥ 0.8 , respectively, and the SRS values were 2.3 ± 3.0 and 14.2 ± 9.8 when the ANN probability was <0.8 and ≥ 0.8 , respectively ($p < 0.0001$).

Diagnostic accuracy of the ANN compared with expert interpretation

Diagnostic accuracy using criteria of significant coronary stenosis of $\geq 75\%$ with either coronary angiography or CCTA was examined in a subgroup of patients ($n = 220$), excluding those after coronary revascularization. Using these criteria, 0, 1-, 2- and 3-vessel disease was found in 98 (45%), 44 (20%), 41

(19%) and 37 (17%) patients, respectively. The AUC values using stress defects for detecting 1-, 2- and 3-vessel disease were 0.69, 0.67 and 0.69 by expert interpretation, and those by the ANN were 0.68, 0.66 and 0.72, respectively, showing an overall AUC of 0.66 (95% CI: 0.60–0.72) by expert interpretation and 0.65 (0.58–0.72) by ANN ($p = 0.74$) for detecting at least one coronary stenosis.

Discussion

This multicenter study aimed to develop an ANN-based diagnosis for myocardial perfusion stress/rest gated and ungated SPECT images. The diagnostic ability of the ANN became comparable to that of nuclear cardiology expert interpretation and was better than conventional semi-quantitative defect scoring. Concordance was reasonable between experts and the ANN even for subgroups of patients with infarction and revascularization because the nuclear cardiologist interpretation taught the ANN.

Comparison with conventional methods

Basic approaches of artificial intelligence-based diagnostic systems differ from the diagnostic steps based on conventional statistical methods [13]. Normal or standard databases are created from individuals with a low likelihood

of coronary artery disease for statistical analysis, and the limits of normal distribution are determined for all myocardial data points. Combinations of the average and the deviation or ratio (%) of threshold counts are used to measure degrees of abnormalities, which are expressed as defect scores. Such statistical analysis provides consistent diagnosis using common databases fitted for populations [22]. It also serves as an aid to clinical diagnosis and has, therefore, been used in most quantitation software packages [5, 23–25]. Quantitative diagnosis using a scoring method is convenient for measuring amounts of ischemia, and for guiding appropriate indications for coronary revascularization [6].

Here, we scored defects using a JSNM working group database of myocardial SPECT images for this purpose [18, 19], which was compared with ANN-based diagnosis. On the other hand, the ANN uses a completely different approach [9, 10, 12]. The software in this study first identified candidate areas of abnormality from >5000 candidate regions based on a specific algorithm, and learned the expert interpretation. Based on the notion that the expert diagnosis is true, the most appropriate features were determined for stress, rest and difference images. In the clinical setting, MPI diagnosis is not simply based on average counts on polar maps, but on the integration of myocardial counts, size and shape of defects, wall motion and systolic thickening, gender and other parameters. Although experts might have unintentionally taken these factors into consideration, the ANN was capable of learning this comprehensive diagnostic process.

Artificial intelligence such as a support vector machine (SVM) and machine learning technologies has generated promising results for the diagnosis and prognosis of patients with coronary artery disease. The AUC for SVM (0.92) is significantly better than total perfusion deficit (0.90) when the SVM algorithm is applied, and its diagnostic accuracy is comparable to the overall accuracy of visual readers [8]. Machine learning that included a combination of clinical and CCTA data predicted 5-year all-cause mortality more effectively than existing clinical or CCTA metrics alone [11]. Since artificial intelligence considers a greater number and complexity of variables than humans, it might be applied for wider indications in the near future.

The meaning of true abnormality

Since the purpose of this project was to create an ANN with diagnostic ability that was similar to that of experts, true abnormality was defined by consensus among expert interpretations. Popular methods for evaluating diagnostic accuracy to date have used coronary stenosis as the gold standard by which sensitivity and specificity are calculated. However, all three coronary territories might not be

judged as abnormal in patients with three-vessel disease. If the most severe stenosis is a culprit lesion and only one territory appears abnormal, even experts might judge only one coronary territory as abnormal. Moreover, some patients who have been diagnosed with myocardial infarction might have recovered perfusion to some extent. Others might have insufficient coronary blood flow supply even from the revascularized coronary artery. Recent indications for MPI include not only patients who are untreated, but also those with many complex pathophysiological conditions such as myocardial ischemia, fibrosis and modifications after revascularization and medical therapies. In fact, 30% of patients included in the present study had OMI and 40% had a history of revascularization. Although we did not use coronary stenosis as the gold standard, the diagnostic ability of the ANN was significantly better than scoring, and comparable to that of experts, indicating a good foundation for further clinical application.

However, even when coronary stenosis of $\geq 75\%$ was used as a diagnostic criterion, the diagnostic accuracy using ROC analysis was comparable between the expert interpretation and ANN. The reason for excluding patients with history of coronary revascularization was that “truth” regarding the presence of remaining ischemia or defects was not confirmed in these patients. Although this analysis was not the goal of the present study, strict criteria of ischemia using coronary stenosis and fractional flow reserve could be evaluated in further studies.

Training and validation databases

The interpretation of myocardial perfusion images during ANN training could affect final diagnostic accuracy. The diagnostic accuracy of the ANN in a previous study of a Swedish database was 92% for abnormalities at stress and 87% for ischemia [9]. In addition, the same ANN generated an AUC of ≥ 0.88 in a Japanese population [10]. After accumulating clinical experience with this software, we found that a minor degree of ischemia was overlooked in a Japanese population. We therefore tried to judge even a slight degree of abnormality as abnormal, which is in agreement with our clinical practice. The ANN learned this tendency of expert interpretation during this project. However, the probability of an abnormality, namely the calculated ANN value in regions with a minor degree of ischemia, might be somewhat lower than that of a definite perfusion defect or ischemia. Overall, since a common stress/rest protocol using ^{99m}Tc -MIBI was applied, we suppose that the software can be applied to routine clinical investigations at many hospitals. The applicability of the ANN-based system to diagnostic ^{201}Tl and other SPECT studies using equipment such as cadmium-zinc-telluride detectors should be separately investigated.

Value of ANN and scoring methods

The non-linear relationship between ANN values and defect scoring was a notable feature that seems in agreement with human decision-making processes. Nuclear cardiologists do not usually judge findings with defect scores <4 as abnormal, but will definitely do so when scores are ≥ 8 . Thus, the likelihood of an abnormal probability steeply increasing between 4 and 8 is understandable. Considering this non-linear relationship, the scoring method cannot be replaced by the ANN. Although 10% ischemia of the left ventricle is used as a guide to coronary revascularization [6, 26], the ANN does not provide a comparable threshold of ischemic severity. However, the ANN can emphasize a possible abnormal region with a degree of probability, which can suggest a diagnosis. Avoidance of overlooking positive findings and over-diagnosing minor abnormalities might be helpful for nuclear cardiology trainees. Since cardiologists and radiologists are not always specialists in nuclear cardiology, an ANN-based suggestion might help enhance confidence in a diagnosis. Even some nuclear cardiology specialists tend to interpret findings with high sensitivity (active reading for abnormality), and others with high specificity (modest reading). The average reading provided by the ANN could thus be a second opinion for such specialists.

Lastly, all perfusion processing, defect scoring and functional analysis are automatically calculated by the software within 10 s. Therefore, if the data transfer protocol from a nuclear-dedicated computer to Windows PC and connection to an institutional picture archiving and communication system (PACS) are appropriately constructed, the ANN software system could function in any institution.

Limitation

The territories of the three coronary arteries were not included in the training process. Although the coronary territory and polar maps might roughly correspond, such as the anteroseptal region to the left anterior descending coronary artery, strict correspondence in border zones of coronary territories is difficult to determine. Although we could not determine coronary-based accuracy in this study due to its algorithm, fusion images with CCTA and perfusion maps with their probability of ischemia will effectively integrate information from stenosis and perfusion. As discussed above, we could not generate a severity score indicating 10% ischemia. To develop an ANN that can assess ischemic severity comparable to that of severity scoring, new large-scale training will be required. The judgment suggested by the ANN system was based on average readings of the MPI studies at 12 institutions. The variety of conditions is a limitation from the viewpoint of specific readings at these institutions, but this could conversely provide the advantage of averaged interpretation.

Conclusion

Based on a new multicenter database trained by nuclear cardiology specialists, the diagnostic ability of the ANN was good with an AUC of >0.9 overall, including subgroups with and without MI and coronary revascularization. Further studies of large patient cohorts are indicated based on these promising outcomes.

Acknowledgements For help collecting the databases and providing valuable comments regarding this study, we appreciate Taishiro Chikamori (Tokyo Medical University Hospital, Tokyo), Chisato Kondo (Tokyo Women's Medical University, Tokyo, at the time of study), Takaaki Isshiki (Ageo Central General Hospital, Ageo) and Akira Yamashina (Tokyo Medical University Hospital, Tokyo) in Japan. We also thank Karin Toth (EXINI Diagnostic AB, Lund, Sweden) for the software development, Takehiro Ishikawa (FUJIFILM RI Pharma, Tokyo) for assistance in database collection and Norma Foster for editorial assistance.

Funding This study was partly supported by JSPS Grants-in-Aid for Scientific Research (C) in Japan (PI: K. Nakajima, no. 15 K09947).

Compliance with ethical standards

Disclosure of potential conflict of interest K. Nakajima collaborates with FUJIFILM RI Pharma, Tokyo, Japan to develop software and has received speaker honorarium from FUJIFILM RI Pharma, Tokyo, Japan (FRI).

K. Okuda, T. Kudo and T. Kasai have collaborations with FRI.

T. Nakata, T. Kasai and S. Matsuo have received speaker honoraria from FRI.

L. Edenbrandt is employed part time at EXINI Diagnostics AB, Lund, Sweden.

K. Kiso, Y. Taniguchi, M. Momose, M. Nakagawa, M. Sarai, S. Hida, H. Tanaka and K. Yokoyama have no conflicts of interest to declare.

Ethical approval All procedures involving human participants complied with the ethical standards enshrined in the Declaration of Helsinki (1964) and its later amendments or comparable ethical standards.

The institutional ethics committee at Kanazawa University approved this multicenter study as the core laboratory, and institutional review boards or ethics committees at all involved hospitals approved participation in this study.

Informed consent All ethics committees waived the requirement for informed consent from individual patients as data collection was retrospective. The opportunity for patients to opt out of participation in the study was presented by public notification of this project online.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Henzlova MJ, Duvall WL, Einstein AJ, Travin MI, Verberne HJ. ASNC imaging guidelines for SPECT nuclear cardiology procedures: stress, protocols, and tracers. *J Nucl Cardiol*. 2016;23:606–39.

2. Klocke FJ, Baird MG, Lorell BH, Bateman TM, Messer JV, Berman DS, et al. ACC/AHA/ASNC guidelines for the clinical use of cardiac radionuclide imaging—executive summary: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (ACC/AHA/ASNC Committee to revise the 1995 guidelines for the clinical use of cardiac radionuclide imaging). *J Am Coll Cardiol*. 2003;42:1318–33.
3. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging Committee of the Council on clinical cardiology of the American Heart Association. *Circulation*. 2002;105:539–42.
4. Berman DS, Abidov A, Kang X, Hayes SW, Friedman JD, Sciammarella MG, et al. Prognostic validation of a 17-segment score derived from a 20-segment score for myocardial perfusion SPECT interpretation. *J Nucl Cardiol*. 2004;11:414–23.
5. Germano G, Kavanagh PB, Slomka PJ, Van Kriekinge SD, Pollard G, Berman DS. Quantitation in gated perfusion SPECT imaging: the cedars-Sinai approach. *J Nucl Cardiol*. 2007;14:433–54.
6. Hachamovitch R, Hayes SW, Friedman JD, Cohen I, Berman DS. Comparison of the short-term survival benefit associated with revascularization compared with medical therapy in patients with no prior coronary artery disease undergoing stress myocardial perfusion single photon emission computed tomography. *Circulation*. 2003;107:2900–7.
7. Guner LA, Karabacak NI, Akdemir OU, Karagoz PS, Kocaman SA, Cengel A, et al. An open-source framework of neural networks for diagnosis of coronary artery disease from myocardial perfusion SPECT. *J Nucl Cardiol*. 2010;17:405–13.
8. Arsanjani R, Xu Y, Dey D, Fish M, Dorbala S, Hayes S, et al. Improved accuracy of myocardial perfusion SPECT for the detection of coronary artery disease using a support vector machine algorithm. *J Nucl Med*. 2013;54:549–55.
9. Johansson L, Edenbrandt L, Nakajima K, Lomsky M, Svensson SE, Tragardh E. Computer-aided diagnosis system outperforms scoring analysis in myocardial perfusion imaging. *J Nucl Cardiol*. 2014;21:416–23.
10. Nakajima K, Matsuo S, Wakabayashi H, Yokoyama K, Bunko H, Okuda K, et al. Diagnostic performance of artificial neural network for detecting ischemia in myocardial perfusion imaging. *Circ J*. 2015;79:1549–56.
11. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2016;38:500–507.
12. Lomsky M, Gjerdtsson P, Johansson L, Richter J, Ohlsson M, Tout D, et al. Evaluation of a decision support system for interpretation of myocardial perfusion gated SPECT. *Eur J Nucl Med Mol Imaging*. 2008;35:1523–9.
13. Garcia EV, Klein JL, Taylor AT. Clinical decision support systems in myocardial perfusion imaging. *J Nucl Cardiol*. 2014;21:427–39. quiz 40
14. Tamaki N. JCS Joint Working Group. Guidelines for Clinical Use of Cardiac Nuclear Medicine (Japanese Circulation Society 2010). <https://www.j-circorjp/guideline/pdf/JCS2010tamaki.pdf> (English digest Version in https://www.jstage.jstgo.jp/article/circj/76/3/76_CJ-88-0019/_pdf). 2010.
15. Lomsky M, Richter J, Johansson L, Hoiland-Carlson PF, Edenbrandt L. Validation of a new automated method for analysis of gated-SPECT images. *Clin Physiol Funct Imaging*. 2006;26:139–45.
16. Nakajima K, Okuda K, Nystrom K, Richter J, Minarik D, Wakabayashi H, et al. Improved quantification of small hearts for gated myocardial perfusion imaging. *Eur J Nucl Med Mol Imaging*. 2013;40:1163–70.
17. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *Int J Comput Vis*. 1988;1:321–31.
18. Nakajima K. Normal values for nuclear cardiology: Japanese databases for myocardial perfusion, fatty acid and sympathetic imaging and left ventricular function. *Ann Nucl Med*. 2010;24:125–35.
19. Nakajima K, Matsumoto N, Kasai T, Matsuo S, Kiso K, Okuda K. Normal values and standardization of parameters in nuclear cardiology: Japanese Society of Nuclear Medicine working group database. *Ann Nucl Med*. 2016;30:188–99.
20. Yoneyama H, Nakajima K, Okuda K, Matsuo S, Onoguchi M, Kinuya S, et al. Reducing the small-heart effect in pediatric gated myocardial perfusion single-photon emission computed tomography. *J Nucl Cardiol*. 2016;24:1378–1388.
21. Nakajima K, Okuda K, Matsuo S, Kiso K, Kinuya S, Garcia EV. Comparison of phase dyssynchrony analysis using gated myocardial perfusion imaging with four software programs: based on the Japanese Society of Nuclear Medicine working group normal database. *J Nucl Cardiol*. 2017;24:611–21.
22. Nakajima K, Okuda K, Kawano M, Matsuo S, Slomka P, Germano G, et al. The importance of population-specific normal database for quantification of myocardial ischemia: comparison between Japanese 360 and 180-degree databases and a US database. *J Nucl Cardiol*. 2009;16:422–30.
23. Garcia EV, Faber TL, Cooke CD, Folks RD, Chen J, Santana C. The increasing role of quantification in clinical nuclear cardiology: the Emory approach. *J Nucl Cardiol*. 2007;14:420–32.
24. Ficaro EP, Lee BC, Kritzman JN, Corbett JR. Corridor4DM: The Michigan method for quantitative nuclear cardiology. *J Nucl Cardiol*. 2007;14:455–65.
25. Slomka PJ, Nishina H, Berman DS, Akincioglu C, Abidov A, Friedman JD, et al. Automated quantification of myocardial perfusion SPECT using simplified normal limits. *J Nucl Cardiol*. 2005;12:66–77.
26. Authors/Task Force members, Windecker S, Kolh P, Alfonso F, Collet JP, Cremer J, et al. 2014 ESC/EACTS guidelines on myocardial revascularization: the task force on myocardial revascularization of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS) developed with the special contribution of the European Association of Percutaneous Cardiovascular Interventions (EAPCI). *Eur Heart J*. 2014;35:2541–619.