



# Automated Detection of Poor-Quality Scintigraphic Images Using Machine Learning

Anil K. Pandey<sup>1</sup> Akshima Sharma<sup>2</sup> Param D. Sharma<sup>3</sup> Chandra S. Bal<sup>1</sup> Rakesh Kumar<sup>1</sup>

<sup>1</sup>Department of Nuclear Medicine, All India Institute of Medical Sciences, Ansari Nagar, New Delhi, India

<sup>2</sup>Department of Urology, All India Institute of Medical Sciences, Ansari Nagar, New Delhi, India

<sup>3</sup>Department of Computer Science, Sri Guru Tegh Bahadur Khalsa College, University of Delhi, Delhi, India

Address for correspondence Anil K. Pandey, PhD, Department of Nuclear Medicine, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India (e-mail: akpandeyaiims@gmail.com).

World J Nuclear Med 2022;21:276–282.

## Abstract

**Objective** In the present study, we have used machine learning algorithm to accomplish the task of automated detection of poor-quality scintigraphic images. We have validated the accuracy of our machine learning algorithm on <sup>99m</sup>Tc-methyl diphosphonate (<sup>99m</sup>Tc-MDP) bone scan images.

**Materials and Methods** Ninety-nine patients underwent 99mTc-MDP bone scan acquisition twice at two different acquisition speeds, one at low speed and another at double the speed of the first scan, with patient lying in the same position on the scan table. The low-speed acquisition resulted in good-quality images and the high-speed acquisition resulted in poor-quality images. The principal component analysis (PCA) of all the images was performed and the first 32 principal components (PCs) were retained as feature vectors of the image. These 32 feature vectors of each image were used for the classification of images into poor or good quality using machine learning algorithm (multivariate adaptive regression splines [MARS]). The data were split into two sets, that is, training set and test set in the ratio of 60:40. Hyperparameter tuning of the model was done in which five-fold cross-validation was performed. Receiver operator characteristic (ROC) analysis was used to select the optimal model using the largest value of area under the ROC curve. Sensitivity, specificity, and accuracy for the classification of poor- and good-quality images were taken as metrics for the performance of the algorithm.

**Result** Accuracy, sensitivity, and specificity of the model in classifying poor-quality and good-quality images were 93.22, 93.22, and 93.22%, respectively, for the training dataset and 86.88, 80, and 93.7%, respectively, for the test dataset.

**Conclusion** Machine learning algorithms can be used to classify poor- and good-quality images with good accuracy (86.88%) using 32 PCs as the feature vector and MARS as the classification model.

## Keywords

- ▶ multivariate adaptive regression splines
- ▶ number of principal components
- ▶ classification of images
- ▶ machine learning

published online  
September 5, 2022

DOI <https://doi.org/10.1055/s-0042-1750436>.  
ISSN 1450-1147.

© 2022. World Association of Radiopharmaceutical and Molecular Therapy (WARMTH). All rights reserved.

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

## Introduction

Scintigraphic procedures performed in nuclear medicine result in images having quality ranging from poor to good. The imaging scenarios involving poor uptake of radiopharmaceutical, administration of less than the optimum radiopharmaceutical dose, extravasations of radiopharmaceutical, deviation from the recommended acquisition protocol, technical issues in imaging devices, and so on may lead to poor-quality low-count images.<sup>1</sup>

Making diagnosis from these poor-quality images is challenging for nuclear medicine physicians because noise masks the visibility of small lesions. The study needs to be repeated if its image quality is so poor that no diagnosis can be made. Therefore, the images are reviewed by nuclear medicine physicians for its diagnostic quality before patient leaves the department after the completion of the scan.

During the review process, for each poor-quality image, nuclear medicine physician has two choices: (1) to use digital image processing techniques to transform poor quality image into diagnostic quality, or (2) to repeat the study. We hypothesize that the task of detection of images that could be improved using techniques of digital image processing to reach acceptable diagnostic quality can be automated. Machine learning algorithms are typically used in classification and regression analysis and existed for decades, but they have only recently become common in medical imaging due to technological advances.<sup>2,3</sup>

In this study, we have used a machine learning algorithm to accomplish the task of automated detection of poor quality scintigraphic images. We have validated the accuracy of our machine learning algorithm on 99mTc-MDP bone scan images.

## Materials and Methods

The whole-body Tc-99m MDP bone scan images included in the present study were acquired as the part of MSc thesis entitled "Image Enhancement Using Stochastic Resonance" which was aimed to enhance the bone scan images acquired with twice the routinely used scanning speed for bone scan acquisition. Enhancement of images was done using postprocessing technique based on stochastic resonance to make them approximately equal to the normal full time bone scan images. The study was approved by the ethics committee of our institute (Ref.No.IECPG-115/21.03.2018, RT-7/10.05.2018). Written consent was taken from all the patients before enrolment in the study.

Ninety-nine patients were scanned twice. First scan was acquired using standard imaging protocol (depending on the count rate in chest region the scanning speed was selected) and second scan was acquired at twice the scanning speed used in the first scan. The second scan had more statistical noise (approximately half the number of total counts in the first scan) compared with first scan. The acquisition protocol used was as follows: 7 to 11 MBq Tc-99m MDP per kg body weight was administered intravenously. Before the administration of Tc-99m MDP, patients were instructed to drink at

least 1 to 2 L of water and void their bladder frequently to reduce the radiation dose to patient, public and staff. After a waiting period of 3 to 4 hours, the whole body bone scan images (both anterior and posterior view) were acquired with low-energy, high-resolution (LEHR) collimator using dual head single-photon emission computerized tomography (SPECT) gamma camera (Symbia E, Siemens Medical Solutions USA, Inc.) in  $1,024 \times 256$  matrix. Therefore, we obtained 198 good-quality images and 198 poor-quality images in our image database.

Nuclear medicine physician visually compared the first and second scan images and verified that the second scan appears to be visually noisier than the first scan image. We labeled the first scan as good quality image and second scan as poor-quality image.

The image database was partitioned into training and test dataset in 60:40 ratio. The principal component analysis (PCA) of each image was performed and the first 32 components were retained as a feature vector of the image.

The automatic detection of poor-quality images was performed using multivariate adaptive regression splines (MARS) model (a machine learning algorithm). The MARS has been previously used for both regression and classification problem in machine learning. Using the feature vectors as input derived from each image using PCA, we trained MARS to classify good and poor-quality images.

We describe PCA and MARS in brief. For detail, interested reader can see the references<sup>4</sup> for PCA and for MARS.<sup>5</sup>

PCA is a statistical technique. Each column of the image is considered as variable and each row of the image is considered as an instance of these variables. PCA transforms this image data as a linear combination of the principal components (PCs) of image variables. The original dataset with maximum variability is represented with the first PC, whereas the data from the remaining dataset with the maximum variability are represented with the second PC. This process goes on consecutively as such with the data from the remaining dataset with the maximum variability being represented with the next PC. If  $m$  represents the number of all PCs and  $p$  represents the number of the significant PCs among all PCs, then  $p$  may be defined as the number of PCs of the  $m$ -dimensional dataset with the highest variance values. The number  $p$  should be chosen such that the data are represented at the very best.<sup>4</sup> In this study, we choose  $p = 32$ , this was based on our previous experience that it is possible to reconstruct the original image with  $p = 32$ .

There is an exhaustive list of classifiers in the literature, and any classifier could have been used. We did not have specific reasons for selecting MARS as classifier. The motivation for using MARS was our familiarity in working with MARS.

MARS is a nonparametric regression procedure that makes no assumption about the underlying functional relationship between the dependent and independent variables. It constructs this relationship from a set of coefficients and so-called basis functions that are entirely determined from the regression data.<sup>5,6</sup> MARS uses a pruning technique to limit the complexity of the model by reducing the number of

its basis functions.<sup>5</sup> The MARS model can be used for both regression and classification task. We have used this model for the classification task.

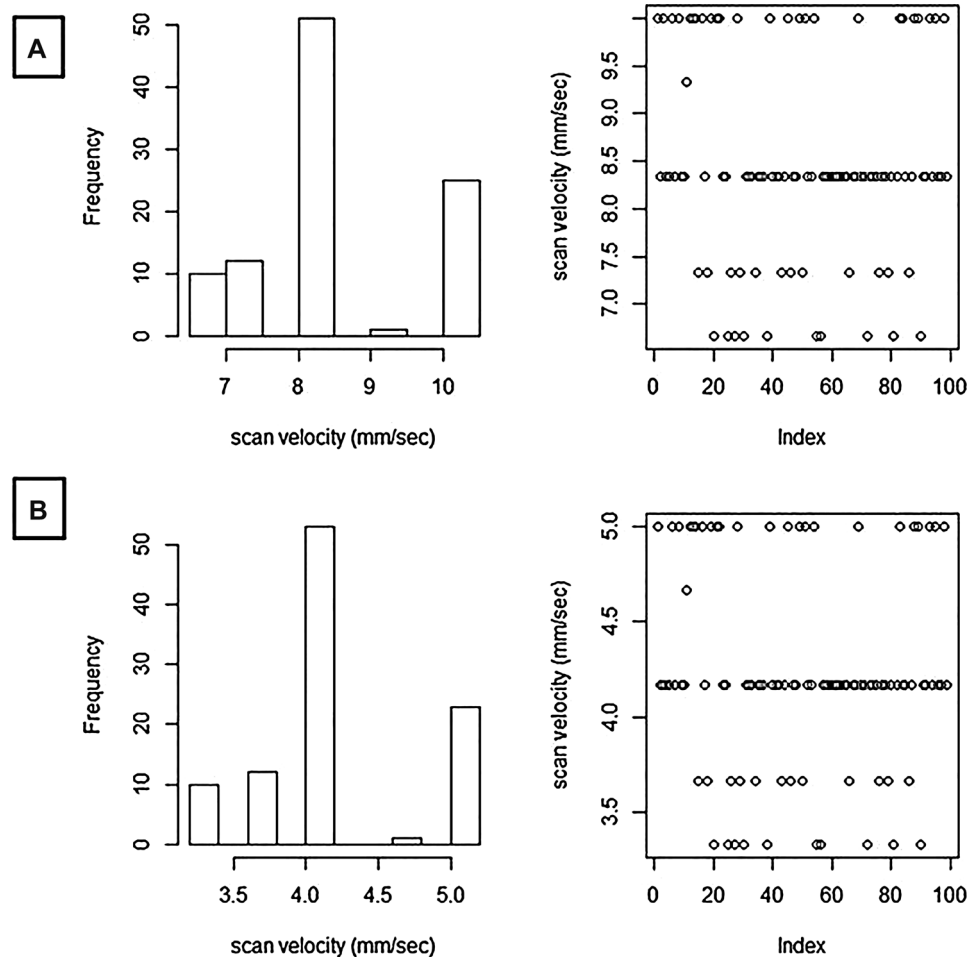
Training of MARS model was done using 32 PCs. Hyperparameter tuning of model was done in which five-fold cross-validation was performed. Receiver operator characteristic (ROC) analysis was used to select the optimal model using the largest value of area under ROC curve.

The performance of MARS model after hyperparameter tuning was further validated on test dataset for the classification of poor- and good-quality images.

All the analysis was performed in package “Caret” installed with free version of R (R version 3.5.1 [2018-07-02] nickname Feather Spray).<sup>7</sup> The motivation for using Caret package was as follows: (1) The Caret package (classification and regression training) contains numerous tools for developing predictive models, and (2) the package contains functions useful in the early stages of a project (e.g., data splitting and preprocessing). The implementation of MARS model provided in Caret package has been used in the study as the classification model.

## Results

The range of speed used in full-time scan and half-time scan was obtained, individual data have been plotted in ►Fig. 1. We have



**Fig. 1** (A) Histogram and plot of full-time scan velocity, (B) histogram and plot of half-time scan velocity.

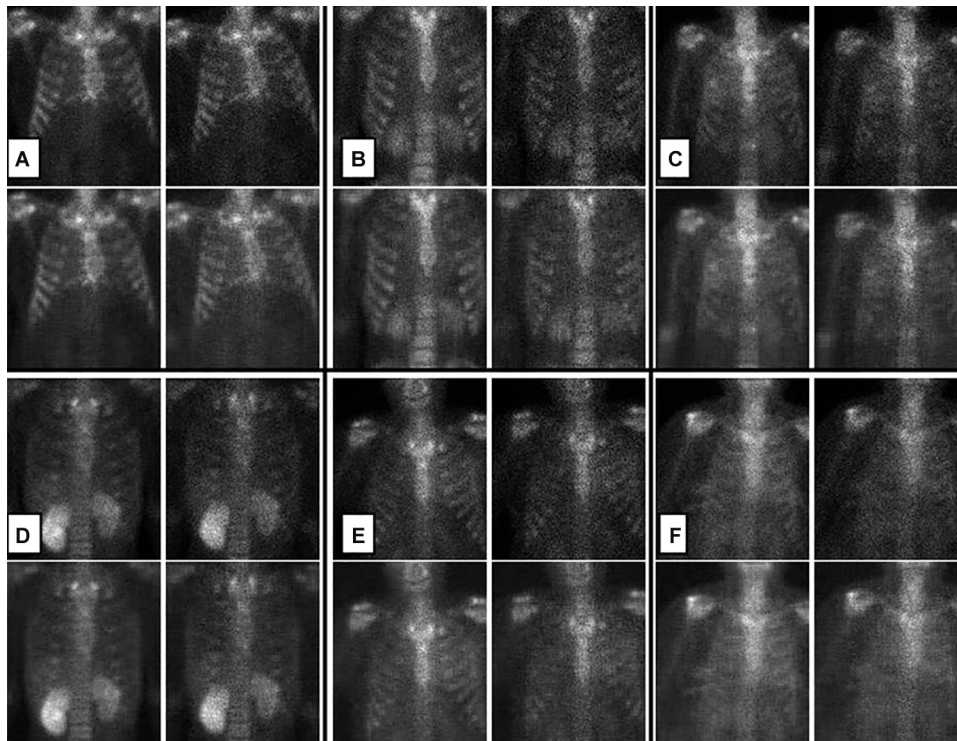
included few images from which reader can get some idea about the visual perception of good- and poor-quality images included in our image database. ►Fig. 2 displays representative images of six different patients acquired during full-time and half-time scans. It can be observed that half-time scan images were noisier than the full-time scan images. Also, the corresponding images reconstructed using 32 PCs are given in the figure.

## Training and Test Result

After pruning a fixed number of terms, the model accuracy in terms of area under ROC, sensitivity, and specificity is given in ►Table 1. For 32 PCs, pruning that resulted in best model performance was found to be two steps with an area under ROC of 0.967 and sensitivity and specificity of 92 and 93%, respectively. Graphs showing the number of predictors found important and the performance of model at different number of pruning steps are given in ►Fig. 3A and B, respectively

►Table 2 shows the confusion matrices of MARS model after hyperparameter tuning for training and test datasets. The confusion matrices show the actual number of true predictions of poor- and good-quality images by the MARS model in training and test datasets.

Accuracy: It shows how close the predicted value is to the actual value. The accuracy of our model in training dataset was 93.22% and test dataset was 86.88%.



**Fig. 2** (A–F) Images of six different patients. In each group, first row left image is the full-time image and first row right image is the half-time image, second row left image is the reconstructed full-time image with 32 PCs, right image is the reconstructed half-time image with 32 PCs.

**Sensitivity:** It is the percentage of actual positive class (good-quality images in our case) that was correctly predicted. In the present study, the sensitivity of model in classification of images in training dataset was 93.22% and in test dataset was 80%.

**Specificity:** Specificity is the proportion of actual negative class (poor-quality images) that are correctly predicted. In the present study, specificity was 93.22% for training dataset and 93.7% for test dataset.

In brief, the above results suggest that using 32 PCs as feature vectors and MARS as classifier for detection of poor and good quality images resulted in accuracy, sensitivity, and specificity greater than 80%.

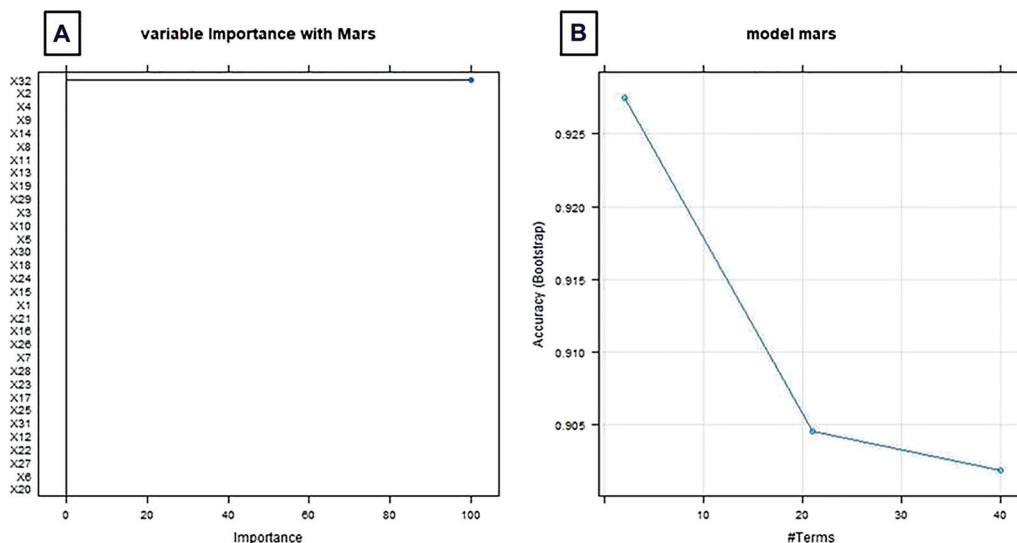
## Discussion

Objective of the present study was to develop an algorithm for automated detection of poor-quality whole-body  $^{99m}\text{Tc}$

**Table 1** Performance results after hyperparameter tuning of the MARS model, ROC was used to select the optimal model using the largest value

	32 PCs			
Number of samples	236			
Number of predictors	32			
Number of classes	2, "G," "P"			
Preprocessing	None			
Resampling	Cross-validated (five-fold)			
Resampling results across tuning parameters	nprune	ROC	Sensitivity	Specificity
	2	0.969	0.923	0.933
	11	0.951	0.933	0.889
	21	0.9215	0.881	0.932
	30	0.9328	0.863	0.924
	40	0.932	0.880	0.915
Final values used for model	nprune = 2, degree = 1			

Abbreviations: G, good-quality images; MARS, multivariate adaptive regression splines; P, poor-quality images; PC, principal component; ROC, receiver operating characteristics.



**Fig. 3** Training results (A) number of predictors found important at different number of pruning step, (B) performance of model at different number of pruning step. MARS, multivariate adaptive regression splines.

**Table 2** Confusion matrix of training and test dataset

Training data			
	Reference		
	G	P	Total
Prediction			
G	110	8	118
P	8	110	118
Total	118	118	236
Test data			
	Reference		
	G	P	Total
Prediction			
G	64	5	69
P	16	75	91
Total	80	80	160

Abbreviations: G, good-quality images; P, poor-quality images.

MDP bone scintigraphic images from a dataset having both poor- and good-quality images using machine learning. We performed the modeling using 32 PCs. By inspecting the input and reconstructed image, a reader can observe that 32 PCs was a good choice, as it has retained the information which is necessary for the reconstructing image very similar to the input image (→ Fig. 2). However, it has helped us in reducing the data from 262,144 data points ( $1024 \times 256 = 262,144$  pixels) to 32 data points (32 PCs). The accuracy, sensitivity, and specificity of model with 32 PCs were 86.88, 80, and 93.7%, respectively.

In medical field, the amount of data available for a given problem is very limited. Therefore, a classifier developed on the training dataset is applied to each sample in the test dataset for validation of the classifier's success. In other fields where sufficiently large data are available, one may use the

advantage of dividing data into three distinct subsets: training, test, and validation of data. However, dividing the available medical data in three ways would have further reduced the amount of data that would actually train the classifier.<sup>8,9</sup>

In our study, we divided the data into training and test dataset in the ratio 60:40 which means 236 samples for training and 160 samples for test datasets, respectively. In the literature, other authors have used various ratios for the splitting of dataset. Rosenwald et al split the dataset of 240 patients into 150:90 with about two-thirds of samples devoted to training set.<sup>10</sup> Boer et al split the data of 152 samples into 80/72, so that 53% of data were used to train the classifier.<sup>11</sup> In the study by Golub et al, the split for data was 40 for training and 32 for test set which means 56% of data were used to train the classifier.<sup>12</sup> Sun et al used 40-size



dataset for the training set and 91 for test set or 31% for training set.<sup>13</sup>

We have used PCA technique for feature extraction. It is a mathematical procedure that extracts relevant information from a large dataset. It has ability to identify relatively fewer numbers of features or components that as a whole represent the full image. Thus, PCs extracted by PCA implicitly represent all the features in the image. There are several other techniques for feature extraction which are based on spatial characteristics (statistical moments like mean and standard deviation), histogram (tonal distribution of digital image, standard deviation, skewness, kurtosis, etc.), texture (local binary patterns, gray-level cooccurrence matrix, the Fast Fourier Transform frequency domain features, etc.), and quality of images.<sup>14</sup> In nuclear medicine, previous work has been reported in compression of dynamic images using PCA. It also has a routine application in picture archiving and communication systems.<sup>15,16</sup> Recently, one study done by Cho et al<sup>17</sup> had shown the role of PCA–support vector machine (SVM) in classification of nuclear medicine images. In this study, the authors have performed 18F-florbetaben amyloid brain positron emission tomography (PET) scans, PCA was used for feature vectors extraction, and SVM was used as classification model whose accuracy was evaluated in classifying amyloid positive and amyloid negative scans. In all the reported work, number of studies have focused on the required number of feature vectors for image classification.

The classifier used in the present study is MARS which is a nonparametric regression procedure that makes no assumption about the underlying functional relationship between the dependent and independent variables. Various other techniques like linear models, generalized additive models, classification and regression tree (CART), artificial neural networks (ANN), logistic regression (LR), decision tree, and so on can be used as classifiers. Studies have been done previously to compare these classification techniques in different areas.<sup>18–23</sup> Moisen and Frescino<sup>18</sup> compared linear models, generalized additive models, CART, MARS, and ANN for mapping forest characteristics in the Interior Western United States using forest inventory field data and ancillary satellite-based information. Delen et al<sup>19</sup> compared LR, decision tree (C5), and ANN for predicting the survivability of diagnosed cases for breast cancer. Stärk and Pfeiffer<sup>20</sup> compared LR, classification tree algorithms (ID3, C4.5, CHAID, and CART), and ANN to solve classification problems in complex datasets in veterinary epidemiology. Colombet et al<sup>21</sup> evaluated the implementation and comparative performance of CART and ANN with an LR model to predict the risk of cardiovascular disease on a real database of patients. Chae et al<sup>22</sup> compared LR, CHAID, and C5.0 in predicting hypertension and provide policy information for hypertension management in a health-insurance domain. Ture et al<sup>23</sup> compared LR, flexible discriminant analysis (FDA), FDA with MARS (FDA/MARS), Chi-squared automatic interaction detector (CHAID), quick unbiased efficient statistical tree (QUEST), CART, radial basis function (RBF), and multilayer perceptron (MLP) to predict control and hypertension groups. We chose MARS for our study because we

wanted to start with no assumptions about possible relationships among feature set of the images.

Our study provides evidence that machine learning algorithms can classify poor- and good-quality images with good accuracy (86.88%) using 32 PCs as feature vector and MARS as classification model. It has a future clinical implication, as the model can help in selecting the poor-quality images automatically which can be further improved by applying suitable postprocessing techniques. This can accelerate the speed of NMPs in interpretation of scans. Moreover, this model can help in accelerating the research in the field of image enhancement in nuclear medicine because most difficult task in the research is the collection of data (poor quality images for image enhancement task). To improve poor quality images, first they need to be identified and separated from total pool of images. The most used and available identification technique is visual inspection which has to be performed with the help of NMPs, who are hard pressed for time. The methodology used in this study to classify the image into poor- and good-quality images can be used by the scientists working on image enhancement.

The limitation of our study is that the classifier model is validated for whole body 99mTc-MDP bone scan images only. In future, we would like to investigate or generalize the model for the classification of other scintigraphic images such as thyroid, iodine, PET, and SPECT images.

## Conclusion

Machine learning algorithms can be used to classify poor- and good-quality images with good accuracy (86.88%), using 32 PCs as feature vector and MARS as classification model.

### Conflict of Interest

There is no potential conflict of interest among authors to declare.

### Ethical Approval

Images included in the present study were acquired as the part of MSc thesis entitled, “Image Enhancement Using Stochastic Resonance” which was aimed to enhance the bone scan images acquired with twice the routinely used scanning speed for bone scan acquisition. Enhancement of images was done using postprocessing technique based on stochastic resonance to make them approximately equal to the normal full-time bone scan images. The study was approved by the ethics committee of our institute (Ref.No.IECPG-115/21.03.2018, RT-7/10.05.2018).

## References

- Cherry SR, Sorenson JA, Phelps ME. Image quality in nuclear medicine. In: Physics in Nuclear Medicine. Philadelphia, PA: Elsevier Health Sciences; 2012
- Uribe CF, Mathotaarachchi S, Gaudet V, et al. Machine learning in nuclear medicine: part 1—introduction. *J Nucl Med* 2019;60(04): 451–458
- Zukotynski K, Gaudet V, Uribe CF, et al. Machine learning in nuclear medicine: part 2—neural networks and clinical aspects. *J Nucl Med* 2021;62(01):22–29

- 4 Jolliffe IT. Principal components in regression analysis. In: *Principal Component Analysis*. New York, NY: Springer; 1986:129–155
- 5 Friedman JH. Multivariate adaptive regression splines. *Ann Stat* 1991;19(01):1–67
- 6 Steinberg D, Colla PL, Kerry M. *MARS User Guide*. San Diego, CA Salford Systems, 1999
- 7 Kuhn M. Building predictive models in R using the caret package. *J Stat Softw* 2008;28(05):1–26
- 8 Crowther PS, Cox RJ. A method for optimal division of data sets for use in neural networks. In *International conference on knowledge-based and intelligent information and engineering systems*. Springer, Berlin, Heidelberg: 2005;1–7
- 9 Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med Genomics* 2011;4(01):31
- 10 Rosenwald A, Wright G, Chan WC, et al; Lymphoma/Leukemia Molecular Profiling Project. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346(25):1937–1947
- 11 Boer JM, Huber WK, Sültmann H, et al. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res* 2001;11(11):1861–1870
- 12 Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286(5439):531–537
- 13 Sun L, Hui AM, Su Q, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* 2006;9(04):287–300
- 14 Campos GF, Mastelini SM, Aguiar GJ, Mantovani RG, de Melo LF, Barbon S. Machine learning hyperparameter selection for contrast limited adaptive histogram equalization. *EURASIP J Image Video Process* 2019;(01):59. Doi: 10.1186/s13640-019-0445-4
- 15 Chameroy V, Di Paola R. High compression of nuclear medicine dynamic studies. *Int J Card Imaging* 1990;5(04):261–269
- 16 Nandi D, Ashour AS, Samanta S, Chakraborty S, Salem MA, Dey N. Principal component analysis in medical image processing: a study. *International Journal of Image Mining*. 2015;1(01):65–86
- 17 Cho K, Kim WG, Kang H, et al. Classification of 18F-florbetaben amyloid brain PET image using PCA-SVM. *Biomed Sci Lett* 2019;25(01):99–106
- 18 Moisen GG, Frescino TS. Comparing five modelling techniques for predicting forest characteristics. *Ecol Modell* 2002;157(2,3):209–225
- 19 Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 2005;34(02):113–127
- 20 Stärk KD, Pfeiffer DU. The application of non-parametric techniques to solve classification problems in complex data sets in veterinary epidemiology—An example. *Intell Data Anal* 1999;3(01):23–35
- 21 Colombet I, Ruelland A, Chatellier G, Gueyffier F, Degoulet P, Jaulent MC. Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. In *Proceedings of the AMIA Symposium* 2000:156
- 22 Chae YM, Ho SH, Cho KW, Lee DH, Ji SH. Data mining approach to policy analysis in a health insurance domain. *Int J Med Inform* 2001;62(2,3):103–111
- 23 Ture M, Kurt I, Kurum AT, Ozdamar K. Comparing classification techniques for predicting essential hypertension. *Expert Syst Appl* 2005;29(03):583–588