Learning a Nonlinear Channelized Observer for Image Quality Assessment

Jovan G. Brankov, *Member, IEEE,* Issam El-Naqa, *Student Member, IEEE*,
Yongyi Yang, *Senior Member, IEEE,* Miles N. Wernick, *Senior Member, IEEE*

***Abstract*-- We propose two algorithms for task-based image quality assessment based on machine learning. The channelized Hotelling observer (CHO) is a well-known numerical observer, which is used as a surrogate for human observers in assessments of lesion detectability. We explore the possibility of replacing the linear CHO with nonlinear algorithms that learn the relationship between measured image features and lesion detectability obtained from human observer studies. Our results suggest that both support vector machines and neural networks can offer improved performance over the CHO in predicting the human-observer performance.**

I. Introduction

IMAGES can be degraded in quality by a variety of factors such as blur, noise, and artifacts. Simple numerical measures, such as signal-to-noise ratio, are often not adequate for capturing the effects of such degradations [1], and may not effectively reflect the quality of the images when used to perform a specific task.

In clinical lesion-detection tasks, it is preferable to quantify image quality by means of detection performance. Ideally, this is accomplished by conducting a human-observer study; however, human-observer studies are costly and time-consuming, and thus impractical for routine use during preliminary stages of algorithm research or imaging-system design. As a result, there has been considerable interest in developing numerical observers that can predict human-observer performance.

A popular method for predicting human-observer performance is the channelized Hotelling observer (CHO) [2, 3], which is based on a linear discriminant applied to image features extracted from different bandpass channels, aimed to model the eye's receptive fields. More specifically, the Hotelling observer is the generalized likelihood ratio test for detection in the case of signal-independent Gaussian noise.

In most studies, the CHO is postulated as a model for the human observer, and it is hoped to correlate well with human-observer data. Indeed, the CHO has been demonstrated to be successful in predicting human-observer performance in medical detection applications [4-7]; however, there have also been occasions when the CHO does not perform as well, as illustrated by the results shown in this paper and in [8]. In these cases, it may be necessary to identify a better model for the human observer.

In this paper, we explore an alternative approach, in which we treat discovery of the human-observer model as a system identification problem. We explicitly estimate the relationship that maps images into detection performance by using nonlinear learning machines such as support vector machines and neural networks. This is illustrated in Fig. 1. To some extent our approach shares a similar philosophy with that in [9] where a linear template was estimated in a two-alternative, forced-choice detection task.

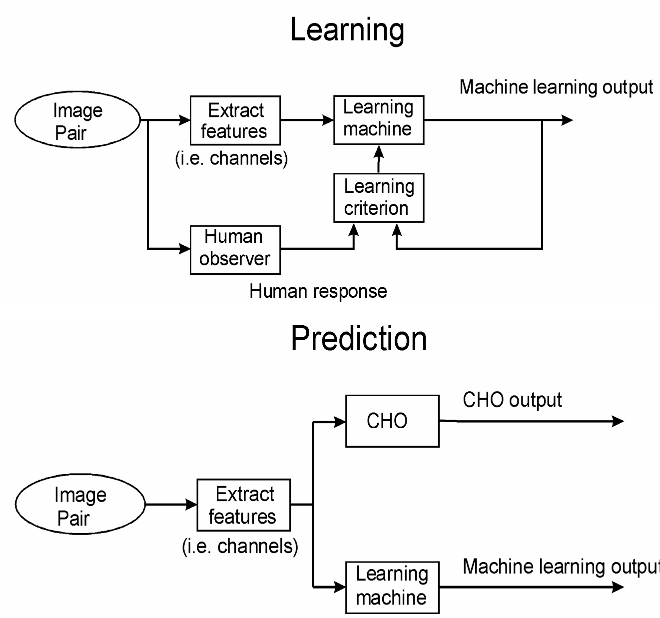


Figure 1. Illustration of proposed machine learning approach for modeling detection performance. Learning algorithm uses feedback in training stage to adaptively improve accuracy.

This work was supported by NIH/NHLBI Grant HL65425.

J. G. Brankov is with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 S. Dearborn St., Chicago, IL 60616 USA (telephone: 312-567-8819, e-mail: brankov@iit.edu).

I. El Naqa is with Department of Radiation Oncology Washington University School of Medicine, Box 8224, St. Louis, MO 63110 USA (telephone: 314-362-0129, e-mail: elnaqa@radonc.wustl.edu).

Y. Yang is with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, 3301 S. Dearborn St., Chicago, IL 60616 USA (telephone: 312-567-3423, e-mail: yy@ece.iit.edu).

M. N. Wernick is with the Department of Electrical and Computer Engineering and the Department of Biomedical Engineering, Illinois Institute of Technology, 3301 S. Dearborn St., Chicago, IL 60616 USA (telephone: 312-567-8818, e-mail: wernick@iit.edu).

II. Methodology

A common measure of observer detection performance is the receiver operating characteristic (ROC) curves. We have chosen in this initial study to train a learning algorithm to predict this measure. The regression model is

$$A_z = f(\mathbf{x}) + \varepsilon\,, \tag{1}$$

where $\mathbf{x}$ is a vector representing the image features, ε is the prediction error, and $f(\cdot)$ is the regression function modeling the relationship between the image features and the area under the ROC curve, A_z, obtained using human observers.

As in the conventional CHO, we use the frequency channels as the input features. Specifically, we assume the standard signal-known-exactly (SKE) channel model [2], that is,

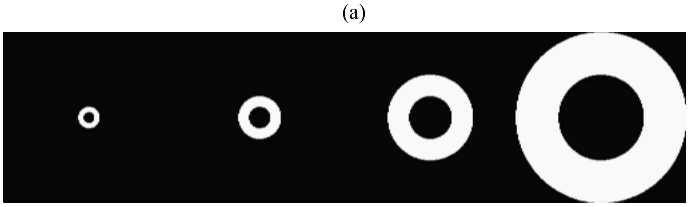
$$\mathbf{x} = \left[\left(\mathbf{U}\mathbf{f}_L\right)^T, \left(\mathbf{U}\mathbf{f}_B\right)^T \right]^T, \tag{2}$$

where $\mathbf{f}_L$ and $\mathbf{f}_B$ are vectors representing a pair of images, one with lesion present and one with lesion absent, respectively, $\mathbf{U}$ represents the channeling operation [2], and T denotes the transpose operation.

In the formulation in (1) the regression function $f(\mathbf{x})$ is used to characterize the detectability of the defect in $\mathbf{f}_L$, as compared to $\mathbf{f}_B$ (where the defect is absent). While $f(\mathbf{x})$ may vary with the noise level in the images, our goal is to train a learning machine so that $f(\mathbf{x})$ will approximate A_z, the detectability measure obtained using human observers.

In Fig. 2 we show the responses of the four constant-Q frequency-band filters used in this study. Note that these filters are spatially centered at the signal location.

(a)



(b)

Figure 2. Feature extraction channels used in the CHO: (a) spatial domain responses, and (b) frequency domain responses.

A. Learning algorithms

We applied and evaluated two learning machines—the support vector machine and the feedforward neural network—for estimation of the functional $f(\cdot)$. These methods are reviewed briefly in the following sections.

Support Vector Machine (SVM)

The SVM is a general learning procedure based on statistical learning theory [10]. SVM embodies the so-called *structural risk minimization* (SRM) principle, which has been shown to be superior [11] to the traditional empirical risk minimization principle. SRM minimizes an upper bound on expected prediction error, as opposed to the average error on the training data. Consequently, an SVM tends to generalize well beyond the training data.

The SVM regression function is characterized by a subset of the training samples known as *support vectors*, denoted by $\mathbf{s}_i$. Specifically, the approximating function is given by:

$$f_{SVM}(\mathbf{x}) = \sum_{i=1}^{l_s} w_i K(\mathbf{x}, \mathbf{s}_i) + w_0\,, \tag{6}$$

where $K(\cdot,\cdot)$ is a kernel function, l_s is the number of support vectors, and the coefficients w_i are determined by quadratic programming. In our experiments we used radial basis functions for $K(\cdot,\cdot)$. Fig. 3 illustrates one-dimensional regression using SVM.

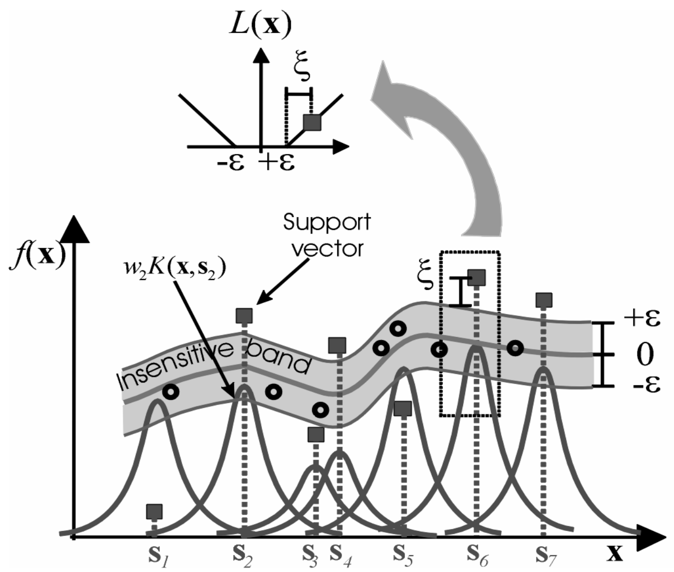


Figure 3. Illustration of regression using a support vector machine.

The SVM objective function is defined as follows:

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_n L_\varepsilon(\mathbf{x}_n) \tag{7}$$

where $\mathbf{w}$ is a vector formed by all the coefficients w_i, C is a regularization parameter that controls the trade-off between the model complexity and fitting accuracy, and $L_\varepsilon(\mathbf{x}_n)$ is the so-called ε*-insensitive* penalty term defined as

$$L_\varepsilon(\mathbf{x}_n) = \begin{cases} \left| f_{SVM}(\mathbf{x}_n) - y_n \right| - \varepsilon & \text{if } \left| f_{SVM}(\mathbf{x}_n) - y_n \right| \ge \varepsilon \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In (8) y_n denotes the desired quality measure, i.e., A_z for the human observers corresponding to the training sample $\mathbf{x}_n$.

Neural networks (FFNN)

For comparison, we also tested a three-layer feedforward neural network [12] (FFNN), which consisted of an input layer with eight neurons (corresponding to the four CHO channels for the image with lesion present and four for the image with lesion absent), a hidden layer with two neurons, and an output layer with a single neuron.

III. EXPERIMENTAL RESULTS

A. Human-observer data

A subset of a previously published human-observer study [7] was used in our development of the proposed methods. The study in [7] was based on images reconstructed using the ordered subsets expectation-maximization (OSEM) algorithm with one and five effective iterations. These images were low-pass filtered with three-dimensional Gaussian filters with different full-width at half-maximum (FWHM) of 0, 1, 2, 3, 4, or 5 pixels. Three human observers evaluated the defect visibility in a SKE environment for images under different algorithmic iterations and FWHM levels; for each setting a total of 130 noisy image realizations were used (65 with lesion present and 65 with lesion absent). The estimated area under the ROC curves, A_z , was calculated for each setting by using ROCKIT [13].

B. Learning setup

To evaluate the proposed algorithms the data were divided randomly into training and test sets. For each setting, (i.e., number of iterations and choice of smoothing), the training set contained 40 realizations were and the test set contained 25 realizations.

In SVM training, ε was chosen to be 0.05. The radial basis kernel width was 15 and C was set to be 10000. This choice led to 26.3 % of the data points as support vectors.

For FFNN training, the target mean-square error was set at 10^{-4} , with a maximum of 1000 epochs (iterations).

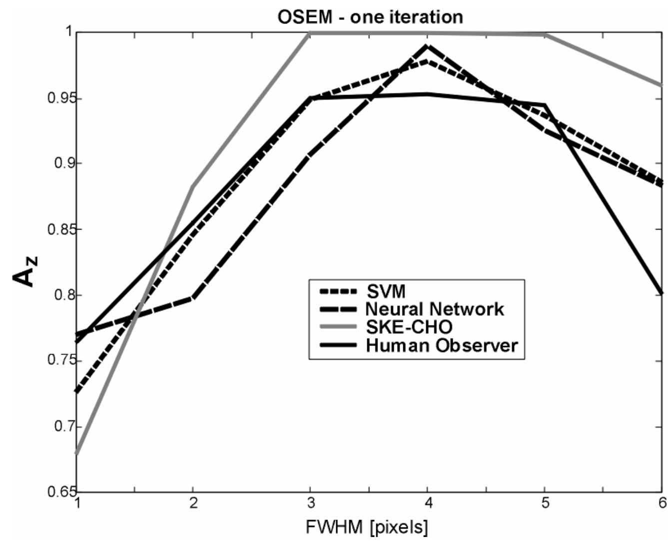
For comparison, the well-known SKE-CHO was used as a benchmark [2, 5]. This method can be viewed as a linear regression model, in which there is an internal observer-noise quantity. For best performance, we optimized this technique by exhaustive grid search over a wide range of internal observer-noise level so as to minimize prediction error for the training data set.

C. Prediction results

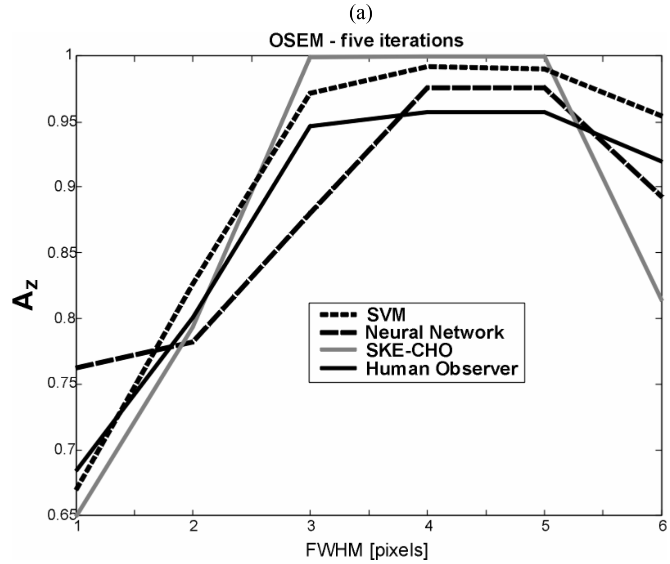
To reduce the influence of reconstruction noise, the prediction results were averaged over all the test images for the same number of iterations and smoothing. The prediction results are shown in Fig. 4 for the images reconstructed by one iteration of OSEM and those obtained by five iterations of OSEM.

As can be seen, the SKE-CHO performs somewhat poorly for this data set,. The best performance is produced by SVM. The overall mean-squared prediction error, relative to that of SKE-CHO, is shown in Table 1 for each method. It appears that nonlinear learning machines can outperform SKE-CHO in predicting human detection performance.

Some example cardiac images are shown in Fig. 5 at different level of smoothing (FWHM = 1, 3 and 5 pixel) for OSEM with one effective iteration. The arrow in each image indicates the lesion position. The lesion is fairly inconspicuous at all smoothing levels. At low smoothing levels, it is difficult to distinguish the lesion from noise artifacts. At high levels of smoothing, the low contrast between lesion and background makes it difficult for human observers to detect the lesion.



(a)



(b)

Figure 4. Comparison of actual and predicted A_z values vs. smoothing FWHM: (a) OSEM - one effective iteration; (b) OSEM - five effective iterations.

TABLE I: NORMALIZED PREDICTION ERROR (SKE-CHO=1)

	SVM	FFNN	SKE-CHO
Prediction error (relative to SKE-CHO)	0.505	0.855	1.000

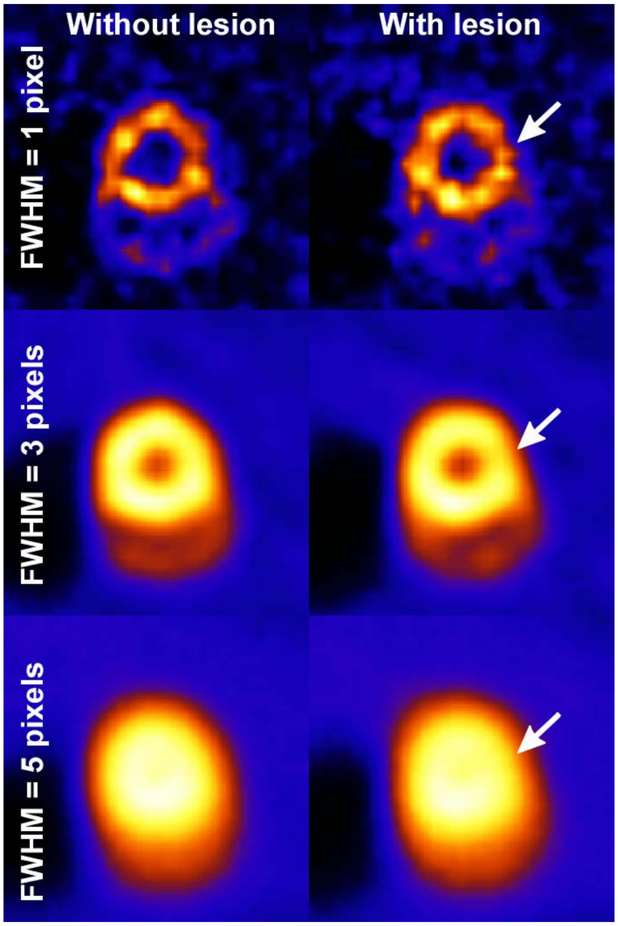


Figure 5. Example images of OSEM with one effective iteration. SKE-CHO predicts that the lesion is most detectable when FWHM=3 to 5 pixels.

IV. CONCLUSIONS

In this work we presented and tested two algorithms for task-based image quality assessment based on machine learning. Image quality was based on lesion detectability. We performed a comparison of the proposed methods with a well-known channelized Hotelling observer (CHO). The results indicate the possible benefit of replacing the linear CHO with nonlinear algorithms that learn the relationship between measured image features and lesion detectability obtained directly from human observers.

V. ACKNOWLEDGEMENT

We thank Manoj V. Narayanan and Michael A. King for generously providing the human-observer data used in the experiments.

VI. REFERENCES

[1] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 636-650, 2000.

[2] K. J. Myers, H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2447-2457, 1987.

[3] J. Yao and H. H. Barrett. "Predicting human performance by a channelized Hotelling model," *Proc. SPIE,* 1786, pp. 161-168, 1992.

[4] S. D. Wollenweber, B. M. W. Tsui, D. S. Lalush, E. C. Frey, K. J. LaCroix and G. T. Gullberg, "Comparison of radially-symmetric versus oriented channel: Models using channelized hotelling observers for myocardial defect detection in parallel-hole SPECT," *Proc. IEEE Nucl. Sci. Symp*., vol. 3, pp. 2090-2094, 1998.

[5] H. C. Gifford, R. G. Wells and M. A. King, "A comparison of human observer LROC and numerical observer ROC for tumor detection in SPECT images," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 4, , pp. 1032-1037, 1999.

[6] C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A*, vol. 18, no. 3, pp. 473-488, 2001.

[7] M. V. Narayanan, H. C. Gifford, M. A. King, P. H. Pretorius, T. H. Farncombe, P. Bruyant and M. N. Wernick, "Optimization of iterative reconstructions of 99m/Tc cardiac SPECT studies using numerical observers," *IEEE Trans. Nucl. Sci*, vol. 49, no. 5, pp. 2355-2360, 2002.

[8] T. K. Narayan and G. T. Herman, "Prediction of human observer performance by numerical observer: an experimental study," *J. Opt. Soc. Am. A*, vol. 16, no. 3, 1999.

[9] C. K. Abbey and M. Eckstein, "Estimates of human-observer templates for a simple detection task in correlated noise," *Proc. SPIE,* 398, pp. 70-77, 2000.

[10] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

[11] S. R. Gunn, M. Brown and K. M. Bossley, "Network performance assessment for neurofuzzy data modeling," *Lecture Notes in Computer Science*, 1280, pp. 313-323, 1997.

[12] S. Haykin, *Neural Networks*, New Jersey: Prentice Hall, 2nd Ed, 1998.

[13] C. E. Metz, "Statistical analysis of ROC data in evaluating diagnostic performance," in *Multiple Regression Analysis: Applications in the Health Sciences*, D. Herbert and R. Myers, Eds: American Institute of Physics, pp. 365-384, 1986.