# A Multiobjective Approach to Optimizing Compterized Detection Schemes

Mark A. Anastasio, *Student Member, IEEE*, Matthew A. Kupinski, Robert M. Nishikawa
and Maryellen L. Giger, *Member, IEEE*
Department of Radiology
The University of Chicago, Chicago, IL 60637

## Abstract

Computerized detection and classification schemes have the potential of increasing diagnostic accuracy in medical imaging by alerting radiologists to lesions that they initially overlooked and/or assisting in the classification of detected lesions. These schemes, generally referred to as computer-aided diagnosis (CAD) schemes, typically employ multiple parameters such as threshold values or filter weights to arrive at a detection or classification decision. In order for the system to have a high performance, the values of these parameters need to be set optimally. Conventional optimization techniques are designed to optimize a scalar objective function. The task of optimizing the performance of a CAD scheme, however, is clearly a multiobjective problem: we wish to simultaneously improve the sensitivity and reduce the false-positive rate of the system. In this work we investigate a multiobjective approach optimizing CAD schemes. In a multiobjective optimization, multiple objectives are simultaneously optimized, with the objective now being a vector-valued function. The multiobjective optimization problem admits a set of solutions, known as the *Pareto-optimal set*, which are equivalent in the absence of any information regarding the preferences of the objectives. The performances of the Pareto-optimal solutions can be interpreted as operating points on an optimal ROC or FROC curve, greater than or equal to the points on any possible ROC or FROC curve for a given dataset and given CAD classifier.

## I. INTRODUCTION

Computer-aided diagnosis schemes have the potential of substantially increasing diagnostic accuracy in radiological imaging [1]. Complicated image features, eye fatigue, and low conspicuity are factors that may cause a radiologist to miss a lesion in a mammogram or chest radiograph. One method for reducing the number of misses is to have two radiologists read the same image separately. This double-reading method is not performed routinely because of the added expense and logistical difficulties in clinical implementation. A CAD scheme can provide the advantage of having a second reader when one would otherwise be absent.

As with any complicated pattern recognition system, CAD schemes typically use multiple parameters such as threshold values, filter weights, and region of interest (ROI) sizes to arrive at a detection or classification decision. For the scheme to have a high performance (high sensitivity and a low false-positive rate), the values of these parameters need to be set optimally. In general, the optimal set of parameters may change when a component of the imaging chain is modified or changed. When the number of parameters is large, it becomes very difficult to manually determine the optimal choice of parameter values, because some of the values may be correlated with each other in some unknown manner. In response to these difficulties, we have previously proposed a genetic algorithm (GA)-based method for determining the set of parameter values that maximize the performance of the system [2].

There are problems, however, with using automated optimization techniques in this capacity. Conventional optimization techniques, including GAs, are designed to minimize (or maximize) a *scalar* objective function. The task of optimizing the performance of a diagnostic decision making system is, however, clearly a multiobjective problem: *we wish to simultaneously increase the sensitivity and reduce the false-positive rate of the system.* We are therefore forced to describe the two objectives by a single (scalar) objective function, which is commonly created by taking a weighted sum of the sensitivity and false-positive rate or using area under the receiver operating characteristic (ROC) curve. The optimization problem then becomes ambiguous because there are an infinite number of methods to map a (sensitivity, false-positive rate) pair to a single scalar number, and it is usually not clear *a priori* which mapping will result in an optimized set of clinically useful parameters.

Recently, we proposed a multiobjective approach for optimizing the performance of diagnostic classifiers [3]. In a multiobjective optimization, multiple objectives are simultaneously optimized, with the objective function now being a vector-valued function. This eliminates the need to find some *ad-hoc* function of the multiple attributes to yield a scalar objective function. It was demonstrated that such an approach has several important advantages over conventional scalar optimization approaches, and produces the ROC curve that best characterizes the classifier performance on the given training dataset.

In this paper, we employ a multiobjective approach for optimizing the performance of two rule-based CAD schemes that have been developed previously in our laboratory. One of the CAD schemes was designed to detect clustered microcalcifications (which are indicative of breast cancer) in digitized mammograms, while the other scheme was developed to detect breast masses. The performances of the solutions returned by the optimization of the detection schemes are observed to define the operating points on an optimal ROC

or free-response ROC (FROC) curve, describing the limiting performances achievable by the detection scheme on the available dataset.

## II. BACKGROUND

### A. Rule-Based CAD Schemes

Rule-based CAD schemes are generally comprised of three processing stages. In the first stage, candidate signals (here, a signal denotes the object one wishes to detect) are identified, often by using a preprocessing filter such as a matched filter. In the second stage, a set of features, denoted by the vector $\vec{x} = [x_1, x_2, \cdots, x_p]$ is calculated for each candidate signal. Lastly, the true signals (or abnormal classifications) are identified by subjecting the feature vectors $\vec{x}$ to the thresholding operation

$$C_a(\vec{w}) = \{\vec{x} \,:\, x_j \geq w_j \;\; \forall j \in \{1, \ldots, p\}\}, \qquad (1)$$

where the parameter vector $\vec{w} = [w_1, w_2, \cdots, w_p]$ represents the threshold value for each feature and $C_a(\vec{w})$ denotes the class of true (or abnormal) signal features. The class of false (or normal) signal features is given by $C_n(\vec{w}) = S - C_a(\vec{w})$ where $S$ is the space spanned by all feature vectors. In general, more complicated rules can be employed that, for example, threshold on a combination of features. The performance of the scheme may also implicitly depend on other parameters not explicitly used by the rules in Eqn. 1 such as ROI sizes.

The process of optimizing (or "training") the CAD scheme involves determining the parameter vector $\vec{w}$ that maximizes the performance of the scheme on an unknown independent dataset. In this work, we assume that the parameter vector $\vec{w}$ contains any system parameters (e.g., ROI sizes) that should be optimized in addition to the threshold values for each rule given in Eqn. 1. For a given $\vec{w}$, the performance of a detection scheme is described by its true-positive fraction $TPF(\vec{w})$ (sensitivity) and the average number of detected false-positive signals per-image $FPI(\vec{w})$ (false-positive rate). A useful construct for describing the overall performance of a detection scheme is the FROC curve [4]. An FROC curve is generated by varying one or more of the components of the parameter vector $\vec{w}$, and plotting the corresponding $(TPF(\vec{w}), FPI(\vec{w}))$ values. In this work, we demonstrate that an optimal FROC curve representing the best possible detection performances on the given dataset can be constructed by using the performances of the solutions returned by a multiobjective optimization of the detection scheme to define the curve.

For a given $\vec{w}$, the performance of a classification scheme is described by its $TPF(\vec{w})$ and false-positive fraction $FPF(\vec{w})$ (misclassification probability). To describe the overall classification performance an ROC curve can be generated by varying one or more of the components of the parameter vector $\vec{w}$, and plotting the corresponding $(TPF(\vec{w}), FPF(\vec{w}))$ values. In this work, we demonstrate the generation of an optimal ROC curve by performing a multiobjective optimization of a classification scheme.

### B. Multiobjective Optimization and ROC or FROC Curve Generation

The problem of simultaneously minimizing (or maximizing) the $n$ components of an objective vector function $\vec{f}(\vec{w}) = [f_1(\vec{w}), \cdots, f_n(\vec{w})]$ is called the multiobjective optimization problem. The solution to the multiobjective optimization problem is not a single vector $\vec{w}$, but rather a set of vectors $\{\vec{w}_i\}$ that are equivalent solutions to the problem in the absence of any *a priori* information about the relative merits of the different objectives. A potential solution to the multiobjective optimization problem is called *non-dominated* if there does not exist a solution superior to it in all of the objectives. Let $\vec{w}^*$ and $\vec{w}$ be two possible parameter vectors, with $\vec{g}^* = [g_1^*, g_2^*, \cdots, g_n^*] = \vec{f}(\vec{w}^*)$ and $\vec{g} = [g_1, g_2, \cdots, g_n] = \vec{f}(\vec{w})$. The parameter vector $\vec{w}$ is said to dominate $\vec{w}^*$ if $g_i \leq g_i^*$, $i = 1, \ldots, n$, and $g_i < g_i^*$ for at least one $i \in \{1, \ldots, n\}$. The solution to the multiobjective problem is the *Pareto-optimal set*, comprised of all non-dominated vectors $\{\vec{w}\}$.

When optimizing the performance of a CAD detection scheme the objective vector $\vec{f}(\vec{w})$ will be compromised of the two components $f_1(\vec{w}) = -TPF(\vec{w}; \mathcal{D})$ and $f_2(\vec{w}) = FPI(\vec{w}; \mathcal{D})$, where $\mathcal{D} \subset S$ denotes the set of features vectors $\{\vec{x}\}$ that are used to compute $\vec{f}(\vec{w})$. When optimizing the performance of a CAD classification scheme the objective vector $\vec{f}(\vec{w})$ will be compromised of the two components $f_1(\vec{w}) = -TPF(\vec{w}; \mathcal{D})$ and $f_2(\vec{w}) = FPF(\vec{w}; \mathcal{D})$. Without loss of generality, we have assumed the objective vector is to be minimized. The Pareto-optimal set will then contain the vectors $\{\vec{w}\}$ that provide the limiting tradeoffs between $TPF(\vec{w}; \mathcal{D})$ and $FPI(\vec{w}; \mathcal{D})$ or $TPF(\vec{w}; \mathcal{D})$ and $FPF(\vec{w}; \mathcal{D})$. In other words, for a given $FPI(\vec{w}; \mathcal{D})$ or $FPF(\vec{w}; \mathcal{D})$ value the Pareto-optimal solution will correspond to the solution possessing the largest possible $TPF(\vec{w}; \mathcal{D})$ value. *We thus observe that the performances of the solutions returned by a multiobjective optimization of a CAD scheme correspond to the operating points on an optimal FROC or ROC curve, indicating the best possible detection or classification performances achievable by the scheme on a given dataset.*

Conventionally, a CAD scheme is trained prior to ROC or FROC curve generation by using a conventional (scalar) optimization technique [2]. An ROC or FROC curve is then generated by varying one or more components of $\vec{w}$ and plotting the corresponding $TPF(\vec{w}; \mathcal{D})$ and $FPI(\vec{w}; \mathcal{D})$ or $FPF(\vec{w}; \mathcal{D})$ values, respectively. The curves generated by varying different sets of components of $\vec{w}$ will generally be distinct, representing the different classification or detection performances achievable by the scheme on the given dataset $\mathcal{D}$. What is usually desired, however, is the ROC or FROC curve describing the *best possible* performances achievable by the CAD scheme on the given dataset, which can be used to characterize the overall efficacy of the CAD scheme. Unfortunately, when the dimension of the parameter vector $\vec{w}$ becomes large, the total number of possible ROC or FROC curves increases as well and it soon becomes a computationally impractical task to compute them all.

Furthermore, a suboptimal conventionally-generated ROC or FROC curve may not be well-behaved in the sense that $TPF(\vec{w}; \mathcal{D})$ does not monotonically increase as $FPI(\vec{w}; \mathcal{D})$ or $FPF(\vec{w}; \mathcal{D})$ increases. It is well known that the sensitivity and false-positive rate of a detection task vary inversely, and thus it is not natural for the sensitivity to decrease as the false positive rate increases. The problem of generating the optimal ROC or FROC curve is circumvented when using a multiobjective optimization technique to train the CAD scheme because the optimal, well-behaved ROC or FROC curve is produced as a by-product of the optimization procedure.

## C. Multiobjective Genetic Algorithms (MOGAs)

Conventional optimization techniques, including gradient-based methods and simulated annealing, cannot directly solve the multiobjective problem because they are not designed with multiple solutions in mind. To apply these methods, an aggregating function must first be defined which projects $\vec{f}(\vec{w})$ to some scalar objective function.

Because genetic algorithms (GAs) maintain a population of candidate solutions (called chromosomes) at each iteration in the optimization, they have successfully been extended to handle multiobjective optimization problems, forming optimization algorithms that are collectively referred to as multiobjective GAs (MOGAs) [5]. In this work, we utilize a MOGA called the Niched Pareto GA (NP-GA), which is described in detail by Horn et al. [6]. We briefly review the salient features of this algorithm below. Readers not familiar with conventional GAs may consult reference [7].

The NP-GA differs from the conventional scalar GA (CS-GA) in that it uses a modified ranking scheme and tournament selection mechanism [7]. As with a CS-GA, the NP-GA begins by randomly generating a set of solutions $\{\vec{w}_i\}_{i=1}^{N_{pop}}$, where $N_{pop}$ denotes the population size. With a CS-GA, the fitness of each solution $\vec{w}_i$ would then be calculated by evaluating the scalar objective function $f_s(\vec{w}_i)$. Because the NP-GA seeks to minimize (or maximize) a vector objective function $\vec{f}(\vec{w}_i)$, it cannot use the objective function values directly to assess the fitness of the $\{\vec{w}_i\}_{i=1}^{N_{pop}}$, but rather utilizes *dominance-based ranking* to do so. In dominance-based ranking, the fitness of each solution $\vec{w}_i$ is made equal to the *rank* of $\vec{w}_i$, where the rank of $\vec{w}_i$ corresponds to the number of solutions in the current population by which it is dominated. Assigning the fitness in this way ensures that solutions with the same rank are treated as equally "good" potential solutions to the optimization problem.

As with the CS-GA, the NP-GA uses a tournament selection method [7] to select the subset of solutions from the current population that are to be put through the genetic operations and subsequently carried over to the future generation. The rank-based fitness strategy used by the NP-GA, based on the concept of dominance, incorporates the multiobjective nature of the problem into the selection mechanism. When the $t_{dom}$ solutions involved in the tournament have the same rank, there will not be a clear winner. In this case, a form of fitness sharing called *equivalence class sharing* [7] is employed where

the winner of a tied tournament is the solution that has the smallest *niche count*, which represents the density of solutions in objective space about the solution. The *niche radius* $\sigma_{share}$ represents the maximum distance between two solutions that will result in an increase in their niche counts. A discussion of the effect of the various NP-GA operating parameters is presented in reference [3].

## III. METHODS

We employed the multiobjective approach to optimize two CAD schemes that have been developed in our laboratory.

### A. CAD scheme for microcalcification detection

We first employed the NP-GA optimization method to optimize a CAD scheme for the detection of clustered microcalcifications which has been developed in our laboratory [8]. Our CAD scheme consists of three major steps. In the first step, the original mammogram is pre-processed by linear filtering to increase the signal-to-noise ratio of the microcalcifications. This is accomplished by use of a difference-image technique or the wavelet transform. The potential microcalcifications are identified in the second step by use of gray-level thresholding and morphological filtering. The third step involves extracting various features of the identified signals to reduce the number of false detections. The signals that remain after feature analysis are subjected to a clustering routine, which groups the detected microcalcifications into clusters.

Five parameters important for signal extraction ($w_1$) and false positive elimination ($w_2 \cdots w_5$) were identified in the signal extraction and feature analysis steps of our CAD scheme to be included in the optimization. The set of features $\mathcal{D}$, used to compute $TPF(\vec{w}; \mathcal{D})$ and $FPI(\vec{w}; \mathcal{D})$, was extracted from a database of 39 mammograms that was obtained with a conventional screen-film (Kodak Min R/OM) system. A standard single-point crossover with an application rate of 50% and standard mutation with an application rate of 10% were employed by the NP-GA as the genetic operations [7]. A $t_{dom}$ value of 3 and a $\sigma_{share}$ value of 0.35 were found (empirically) suitable for the problem.

One of the solutions returned by the NP-GA was chosen to represent a typical solution that might be returned by a CS-GA optimization [2]. An FROC curve was then generated by treating the $w_1$ component of $\vec{w}$ as an independent decision variable which was used to sweep out the curve. We hereafter refer to this curve as the "conventional FROC curve" because this was the method of FROC curve generation reported in a previous study [2]. Other decision variables have not previously been identified that produce better conventional FROC curves. The conventional FROC curve was compared to the FROC curve returned by the NP-GA.

### B. CAD scheme for mass detection

The computerized mass detection scheme currently being developed at the University of Chicago includes a preprocessing step of bilateral subtraction [9] followed by a classifier acting

as a post-processing step. With this CAD scheme, a series of potential lesion sites are located using the bilateral subtraction algorithm. Each potential lesion site is analyzed and features are extracted from it. These features are used by a classifier to determine whether that suspicious area is an actual lesion or a false-detection. This classification step should preferentially reduce the number of false detections returned by the initial detection algorithm while keeping most of the true detections [10].

We have implemented a rule-based classifier for use in the mass detection scheme. The optimal threshold values for this classifier have been determined using the NP-GA. A total of four features relating to the potential lesion contrast ($w_1$), edge sharpness ($w_2, w_4$), and circularity ($w_3, w_4$) were used in this study. The set of features $\mathcal{D}$, used to compute $TPF(\vec{w}; \mathcal{D})$ and $FPF(\vec{w}; \mathcal{D})$, was extracted from the 118 true-positive detected lesions and the 10,697 total false detections returned by the initial detection algorithm using a database of 60 cases (with 4 images per case). A standard single-point crossover with an application rate of 30% and standard mutation with an application rate of 5% were employed by the NP-GA as the genetic operations [7]. A $t_{dom}$ value of 3 and a $\sigma_{share}$ value of 0.1 were found (empirically) suitable for the problem.

Again, one of the solutions returned by the NP-GA was chosen to represent a typical solution that might be returned by a CS-GA optimization. An ROC curve was then generated by treating the $w_1$ component of $\vec{w}$ as an independent decision variable which was used to sweep out the curve. We hereafter refer to this curve as the "conventional ROC curve". Other decision variables have not previously been identified that produce better conventional ROC curves. The conventional ROC curve was compared to the ROC curve returned by the NP-GA.

## IV. RESULTS

### A. CAD scheme for microcalcification detection

The FROC curve produced by the NP-GA and the conventional FROC curve are shown in Fig. 1. The CS-GA optimization solution corresponded to the operating point $(TPF(\vec{w}; \mathcal{D}) = 0.85, FPI(\vec{w}; \mathcal{D}) = 0.21)$. It is seen that the NP-GA generated FROC curve is everywhere greater than or equal to the conventional FROC curve. Other conventional FROC curves generated using parameters $w_2$-$w_5$ as the independent decision variables produced curves that did not span both the low- and high-$FPI(\vec{w}; \mathcal{D})$ regions of FROC space and/or were lower than the $w_1$ generated curve.

### B. CAD scheme for mass detection

The ROC curve produced by the NP-GA and the conventional ROC curve are shown in Fig. 2. The CS-GA optimization solution corresponded to the operating point $(TPF(\vec{w}; \mathcal{D}) = 0.75, FPF(\vec{w}; \mathcal{D}) = 0.08)$. It is seen that the NP-GA generated ROC curve is everywhere greater than or equal to the conventional ROC curve. Note that the conventional ROC curve does not span both the low- and high-$FPF(\vec{w}; \mathcal{D})$ regions of ROC space. Other conventional
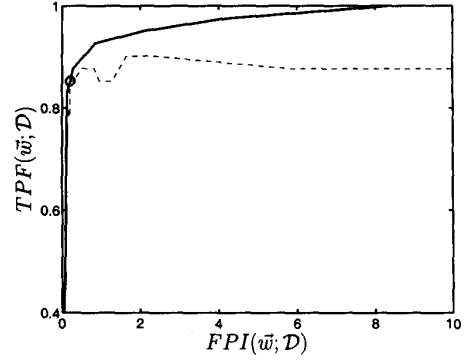


Fig. 1 Conventional FROC curve (dashed-lines) and the FROC curve produced by the NP-GA optimization (solid line). The CS-GA solution is represented by the circle. The conventional FROC curve was then generated by varying the $w_1$ component of the scalar optimized $\vec{w}$ as described in the text.
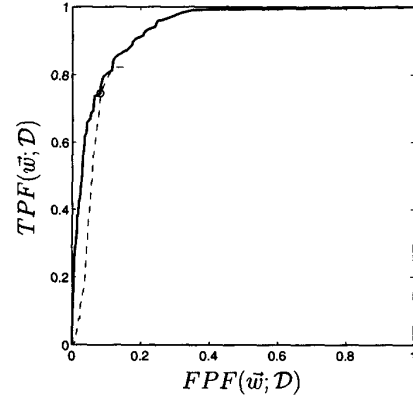


Fig. 2 Conventional ROC curve (dashed-lines) and the ROC curve produced by the NP-GA optimization (solid line). The CS-GA solution is represented by the circle. The conventional ROC curve was then generated by varying the $w_1$ component of the scalar optimized $\vec{w}$ as described in the text.

ROC curves generated using parameters $w_2$-$w_4$ as the independent decision variables also produced curves with this problem.

## V. DISCUSSION

As seen in Figs. 1 and 2, the NP-GA generated FROC and ROC curves are everywhere greater than or equal to the conventional FROC and ROC curves. Prior to generating the conventional FROC curves, a solution of a scalar optimization of the scheme was identified. The conventional FROC and ROC curves were then obtained by varying certain components of the optimized parameter vector $\vec{w}$. In general, the conventional curves generated in this manner will be suboptimal because the scheme was naturally trained to operate at only one point on the curves. The other operating points on the curves will generally correspond to $\vec{w}$ vectors that are dominated, and thus suboptimal. The NP-GA avoids this problem because each solution returned by the NP-GA corresponds to a non-dominated operating point on the FROC

or ROC curve, which produces an optimal FROC or ROC when the non-dominated set is Pareto-optimal.

In addition to being suboptimal, the conventional ROC curve shown in Fig. 2 covers only a restricted range of the $FPF(\vec{w}; \mathcal{D})$ axis, and thus does not provide a complete description of the classification performance. A similiar behavior was observed for the FROC curves of the detection scheme when the $w_2$-$w_5$ parameters were used as decision variables. This behavior is not surprising because a rule-based CAD scheme does not project the input data onto a single decision variable that can be thresholded to detect or classify a signal. Rather, the rules in Eqn. 1 are independently applied and the results logically combined to arrive at a detection or classification decision. When a subset of the parameter vector $\vec{w}$ is not used to generate the ROC or FROC curve, certain regions of the ROC or FROC space may remain inaccessible, as is clearly demonstrated by Fig. 2. The need to choose an appropriate decision variable which can sweep out the ROC or FROC curve is removed when using the NP-GA.

Note that in Fig. 1 the conventional FROC curve is not a monotonically increasing curve. We know, however, that the sensitivity and false-positive rate of a detection scheme should vary inversely, implying that the slope of the FROC curve should be positive. The solutions returned by the NP-GA are non-dominated, and hence, by definition, the FROC (or ROC) curve obtained by connecting the operating points of the NP-GA solutions necessarily has a positive slope at all points.

In this paper, we have addressed the issue of characterizing the performance of a CAD scheme on a training dataset. When the amount of training data is limited and/or is not representative of the global data distribution, the ROC or FROC curve generated using the training data may not be an accurate estimate of the general detection performance. To generate a "test" ROC or FROC curve using an independent dataset, the performances of the non-dominated solutions returned by the NP-GA optimization should be re-evaluated on the independent dataset. The test ROC or FROC curves generated in this way can be used by statistical tests, such as the jackknife test, to provide an estimate of the general system performance [2].

## VI. CONCLUSION

Conventional methods of CAD optimization optimize to a single operating point from which an ROC or FROC curve is generated by varying a subset (typically only one) of the CAD parameters. Using the NP-GA, we are able to optimize the entire ROC or FROC curve instead of just a point in ROC or FROC space. Assuming the optimization is complete, the curve returned by the NP-GA is optimal in the sense that no other ROC or FROC curve is better for the given dataset and detection scheme. This allows one to fully characterize the performance of a CAD scheme on a specified dataset.

## VII. REFERENCES

[1] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger, and K. Doi, "Improving breast cancer diagnosis with computer-aided diagnosis," Academic Radiology, 1998 (in press).

[2] M. A. Anastasio, H. Yoshida, R. Nagel, R. M. Nishikawa, and K. Doi, "A genetic algorithm-based method for optimizing the performance of a computer-aided diagnosis scheme for detection of clustered microcalcifications in mammograms," Medical Physics, vol. 25, pp. 1613–1620, 1998.

[3] M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating ROC curves," IEEE Transactions on Medical Imaging, 1998 (in review).

[4] P. C. Bunch, J. F. Hamilton, G. K. Sanderson, and A. H. Simmons, "A free-response approach to the measurement and characterization of radiographic-observer performance," Journal of Applied Photographic Engineering, vol. 4, pp. 166–171, 1978.

[5] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," Evolutionary Computation, vol. 3, no. 1, pp. 1–16, 1995.

[6] J. Horn and N. Nafpliotis, "Multiobjective optimization using the niched parto genetic algorithm," illiGAL report no. 93005, University of Illinois at Urbana-Champaign, July 1993.

[7] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Publishing Company, Inc., 1989.

[8] R. M. Nishikawa, M. L. Giger, K. Doi, C. Vyborny, and R. A. Schmidt, "Computer-aided detection of clustered microcalcifications on digital mammograms," Medical and Biological Engineering and Computing, vol. 33, pp. 174–178, 1995.

[9] F.-F. Yin, M. L. Giger, C. J. Vyborny, K. Doi, and R. A. Schmidt, "Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses," Investigative Radiology, vol. 28, pp. 473–481, 1993.

[10] M. Kupinski, M. L. Giger, P. Lu, and Z. Huo, "Computerized detection of mammographic lesions: Performance of artificial neural network with enhanced feature extraction," in SPIE, vol. 2434, pp. 598–605, 1995.