NIH Public Access
Author Manuscript
J Struct Biol. Author manuscript; available in PMC 2012 March 1.

Published in final edited form as:
J Struct Biol. 2011 September ; 175(3): 348–352. doi:10.1016/j.jsb.2011.03.009.

A Clarification of the Terms Used in Comparing Semi-automated Particle Selection Algorithms in Cryo-EM

Robert Langlois[a] and **Joachim Frank**[a,b,c]

[a]Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032

[b]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032

Abstract

Many cyro-EM datasets are heterogeneous stemming from molecules undergoing conformational changes. The need to characterize each of the substrates with sufficient resolution entails a large increase in the data flow and motivates the development of more effective automated particle selection algorithms. Concepts and procedures from the machine-learning field are increasingly employed toward this end. However, a review of recent literature has revealed a discrepancy in terminology of the performance scores used to compare particle selection algorithms, and this has subsequently led to ambiguities in the meaning of claimed performance. In an attempt to curtail the perpetuation of this confusion and to disentangle past mistakes, we review the performance of published particle selection efforts with a set of explicitly defined performance scores using the terminology established and accepted within the field of machine learning.

Keywords

particle selection; cryo-EM; machine learning; false positive rate

Introduction

Single-particle reconstruction of samples imaged by cryo-electron microscopy (cryo-EM) has been established as a powerful method to visualize three-dimensional macromolecular complexes in structural biology. The trend of cryo-EM is moving toward large, automated collections of heterogeneous data to elucidate the spectrum of conformation states in equilibrium for a single macromolecular complex. Micrographs contain low-contrast two-dimensional projections of molecules captured by the transmission electron microscope. We refer to the projections visible on the micrograph as “particles.” Before the three-dimensional density map of the complex can be reconstructed, these particles must be located and extracted from their noisy background. This step remains a serious bottleneck in the single-particle reconstruction workflow.

© 2011 Elsevier Inc. All rights reserved.

[c]Corresponding Author: Dr. Joachim Frank, Howard Hughes Medical Institute, Columbia University, Dept. of Biochemistry and Molecular Biophysics, 650 West 168th Street, Black Building 2-221, New York, NY 10032., Phone: 212 305-9510, Fax: 212 305-9500, jf2192@columbia.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In situations where the contrast is not sufficient for the application of standard computer vision algorithms, the currently available state-of-the-art particle-picking tools highlight potential particles on the micrograph, requiring manual verification of each potential particle window (i.e., the usually square array of pixels containing the putative particle). While this procedure reduces the load on the manual picker, it falls far short of the promise to automate particle picking.

An automated particle-picking algorithm should both detect and recognize (i.e., classify) particles on a micrograph. With this view, current approaches can be divided into three categories: generative, discriminative, and unsupervised classifiers. A *generative* classifier recognizes particles by measuring their similarity to a reference; template-matching techniques best exemplify this approach (Chen and Grigorieff, 2007; Hall and Patwardhan, 2004; Huang and Penczek, 2004; Rath and Frank, 2004; Roseman, 2003; Sigworth, 2004; Volkmann, 2004; Wong et al., 2004; Yu and Bajaj, 2004). A *discriminative* classifier recognizes particles based on a decision boundary, which has been learned from a set of both positive (windows verified to contain a particle) and negative (windows verified not to contain a particle). Discriminative methods can be further subdivided into feature-based methods (Mallick et al., 2004; Sorzano et al., 2009) and neural networks (Ogura and Sato, 2004a), which implicitly derive features from the images themselves. An unsupervised method relies on the limited complexity of physical objects within an image in order to recognize particles on a micrograph. This category covers a broad range of algorithms including image segmentation (Adiga et al., 2005; Umesh Adiga et al., 2004), edge detection (Voss et al., 2009; Woolford et al., 2007a; Zhu et al., 2003), graph-based methods (Hall and Patwardhan, 2004; Singh et al., 2004), clustering (Shaikh et al., 2008), and reinforcement learning (Ogura and Sato, 2004b).

In this article we wish to address a problem that has arisen in the literature addressing particle selection due to inconsistent or nonstandard use of terminology. The remainder of this article is organized as follows. First, the proper terminology is introduced for common performance metrics and plots used in machine learning. Second, examples are presented where improper use of metric and plot terminology has led to ambiguous results. Finally, this paper concludes recommendations to ensure future papers will follow conventions established in the field of machine learning.

Performance Metrics

The particle selection problem is fundamentally a binary classification problem where particles form the *positive class* while “non-particles” such as contaminants and noise form the *negative class*. Since there are two classes, a learning algorithm can make two types of errors: one for each class. In addition, a cost can be associated with each type of error. Thus, a number of metrics have been proposed (Caruana and Niculescu-Mizil, 2006; Davis and Goadrich, 2006; Drummond and Holte, 2006; Sattar et al., 2006; Weng and Poon, 2008; Witten and Frank, 2005) to summarize the performance of a learning algorithm with regard to the type and cost of errors associated with a problem.

The confusion matrix, shown in Table 1, is a visual tool, which illustrates the two types of error: false positives and false negatives. The columns tabulate the number of samples in the actual class and the rows the predicted class. The two classes are commonly referred to as the positive class (or the class of interest) and the negative class. In the particle selection problem, windows that contain particles comprise the positive class whereas windows that do not the negative class.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

1. Hypothesis Testing

Machine learning is a discipline at the intersection of computer science and statistics, and much of the nomenclature has statistical roots. For example, a machine-learning algorithm can be viewed as finding an accurate hypothesis to explain a set of training data (Schapire and Singer, 1999). With this view, testing the validity of a learning algorithm is analogous to hypothesis testing where the concepts of type I and type II errors are analogous to the specificity (Eq. 1) and sensitivity (Eq. 2), respectively. Note that the abbreviations in the following equations follow the definitions in Table 1.

$$\text{Specificity} \quad \frac{TN}{FP+TN}=\frac{TN}{N} \tag{1}$$

$$\text{Sensitivity} \quad \frac{TP}{FN+TP}=\frac{TP}{P} \tag{2}$$

This terminology, specificity and sensitivity (also called recall, true positive rate, or hit rate) originated from medical diagnostics. The *specificity* estimates the ability of the classifier (*the test* in medical diagnostics) to correctly identify negative examples (or those people *without* the disease) by the fraction of true negatives (*TN*) over the total negatives (*N=FP +TN*). The *sensitivity* estimates the ability of the classifier to correctly identify positives (or those people *with* the disease) by the fraction of true positives (*TP*) to total positives (*P=FN +TP*). The sensitivity and specificity can also be expressed in terms of failure as measured by the *false positive rate* (Eq. 3) *and false negative rate* (Eq. 4), respectively. Often, it is useful to summarize both sensitivity and specificity in a single metric: the *accuracy* (Eq. 5) or (in terms of failure) the *error* (Eq. 6) of classification.

$$\text{False Positive Rate} \quad \frac{FP}{FP+TN}=\frac{FP}{N}=1-\text{Specificity} \tag{3}$$

$$\text{False Negative Rate} \quad \frac{FN}{FN+TP}=\frac{FN}{P}=1-\text{Sensitivity} \tag{4}$$

$$\text{Accuracy} \quad \frac{TP+TN}{FN+TP+TN+FN}=\frac{TP+TN}{P+N} \tag{5}$$

$$\text{Error} \quad \frac{FP+FN}{FN+TP+TN+FN}=\frac{FP+FN}{P+N} \tag{6}$$

A learning algorithm often produces a real-value output, *e.g.* correlation score from template matching, in which distinct thresholds produce different tradeoffs between specificity and sensitivity. This tradeoff can be visualized in a two-dimensional plot with a *receiver operating characteristic* (ROC) curve (see Fawcett (2004) for an in-depth review).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In the ROC curve, the sensitivity is plotted on the y-axis and 1-specificity on the x-axis; in ROC nomenclature, these are referred to as the true positive rate (TPR) and false positive rate (FPR), respectively. The area under the ROC curve serves as a useful measure to summarize the overall performance of a classifier.

2. Information retrieval (*i.e.* Particle selection)

A special case of classification occurs when the dataset is highly skewed, *e.g.* in information retrieval, the number of relevant documents is much smaller than the number of non-relevant documents. In this case, there is no agreed-upon negative set and, since the specificity cannot be estimated, the precision (or positive predictive value) (Eq. 7) is estimated in its place. The precision estimates the probability a classifier's positive prediction is actually positive as the fraction of true positives (*TP*) over the total predicted positives ($P'=FP+TP$). Note that the sensitivity can still be estimated, yet is usually referred to as *recall* (Eq. 8). One minus the precision is known as the *false discovery rate* (FDR), Eq. 9, (Benjamini and Hochberg, 1995). Finally, the f-measure (Eq. 10) summarizes both the precision and recall with the harmonic mean.

$$\text{Precision} \quad \frac{TP}{FP+TP}=\frac{TP}{P'} \tag{7}$$

$$\text{Recall} \quad \frac{TP}{FN+TP}=\frac{TP}{P} \tag{8}$$

$$\text{False discovery rate} \quad \frac{FP}{FP+TP}=\frac{FP}{P'}=1-\frac{TP}{P'} \tag{9}$$

$$\text{F}-\text{measure} \quad 2\cdot\frac{\text{precision}-\text{recall}}{\text{precision}+\text{recall}} \tag{10}$$

Similar to specificity and sensitivity, the tradeoff between precision and recall can also be visualized in two dimensions with the precision-recall curve. The precision-recall curve plots the precision on the y-axis against the recall (or sensitivity) on the x-axis. Unlike the ROC plot, the precision-recall curve must be interpolated to ensure it decreases monotonically.

Incongruent Terminology

The problem we wish to address originates with the terminology used to describe Eq. 9 as the *false positive rate* (Zhu et al., 2003). This occurred just prior to the rise of a renewed interest in the particle selection problem, which culminated in a particle selection bakeoff (Zhu et al., 2004). While the false positive rate is an accurate description of Eq. 9, which measures the number of false positives with respect to the number of discovered windows, this term had already be established as referring to Eq. 3 (Hand and Till, 2001; Kohavi, 1997; for a wider review see Drummond and Holte, 2006). The proper term for the quantity defined in Eq. 9 is the false discovery rate.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Prior to Zhu et al. (2003), the success of particle selection had been measured subjectively by the quality of class averages (Pascual-Montano et al., 2001) or the quality of a cluster analysis (Pascual et al., 2000) or, objectively, either by the accuracy, Eq. 5, (Harauz and Fong-Lochovsky, 1989) or by counting the number of true/false positives (Harauz and Fong-Lochovsky, 1989). For an in-depth review of work prior to 2001, see Glaeser (2004).

This divergence in terminology continued in a number of subsequent publications. The false positive rate continued to be described as referring to Eq. 9 in some studies (Huang and Penczek, 2004; Mallick et al., 2004; Singh et al., 2004; Woolford et al., 2007a) while others simply used the term without ever defining it (Chen and Grigorieff, 2007; Roseman, 2004; Sigworth, 2004; Volkmann, 2004; Wong et al., 2004; Zhu et al., 2004). It is the latter category of papers that has led to ambiguity in published results. With the publications resulting from the bakeoff (Roseman, 2004; Sigworth, 2004; Volkmann, 2004; Wong et al., 2004), it can be safely assumed their results use Eq. 9. This is, however, not a safe assumption for works coming much later (Chen and Grigorieff, 2007; Woolford et al., 2007b). Specifically, Woolford and coworkers (Woolford et al., 2007b) draw comparisons to prior work arriving at an FDR of 15%, yet bases this number on 2577 detected particles. It appears that when the same dataset was used, then the total number of confirmed particle windows should be 1042, which would yield an FDR of 63%.

Another consequence of the inconsistent use of terminology can be illustrated by the comparisons of the particle selection algorithms by means of a ROC plot (Mallick et al., 2004; Sorzano et al., 2009). As described in the previous section, a ROC curve compares TPR (Eq. 2) to FPR (Eq. 3) (Figure 1a); however, as previously pointed out, FDR (Eq. 9), rather than FPR, is actually being estimated (see Figure 1b) by most particle selection studies (Mallick et al., 2004; Sorzano et al., 2009). The standard curve closest to the “false discovery curve” in Figure 1b is the precision/recall curve (Figures 1c and 1d). These two curves, ROC and Precision/Recall, measure different aspects of an algorithm’s performance, and thus, have different properties and limitations. For problems such as particle selection where the number of negative examples far exceeds the number of positive, an empirical (i.e., without interpolation) precision-recall curve and its analog, the FDR curve have a distinct signature they do not necessarily increase monotonically like a ROC curve (Brodersen et al., 2010). For example, the ROC curve in Mallick et al. (2004) is actually an FDR curve, yet it does not exhibit the characteristics one would expect of an empirical FDR curve like the seesaw effect (Davis and Goadrich, 2006), see Brodersen et al. (2010) for an excellent illustration.

Other works avoided propagating the inconsistent terminology in a number of ways. Some compared the Fourier-shell correlation between reconstructions of particles windows chosen manually and with the particle selection algorithm (Hall and Patwardhan, 2004; Kumar et al., 2004; Roseman, 2003; Woolford et al., 2007a). Others reported the percentage of false positives (%FP) (Rath and Frank, 2004; Short, 2004; Umesh Adiga et al., 2004), which is rarely used in machine learning but arguably avoids the problems we have pointed out.

Concluding Remarks

Particle selection remains a significant bottleneck in automating single-particle reconstruction from data collected by cryo-EM. A proper evaluation of available tools for particle selection is crucial to their wider acceptance within the community. The particle selection problem is an information retrieval problem in that it has a highly skewed dataset with no well-defined set of negative (i.e., non-particle) samples; thus, the metrics and terminology should follow those established in information retrieval: precision and recall.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

It can be argued that the disputed term *false positive rate* has a clear meaning given the nature of the particle selection problem. Since the particle selection problem does not have a well-defined set of negative examples, the *false positive rate* can only measure the number of false positives with respect to the number of discovered samples. However, this incorrect use of terminology has clearly resulted in a number of inconsistencies including the subsequent incorrect use of related terminology, *e.g.* ROC curve, which, in turn, led to ambiguous results.

One step to alleviate future problems would be to require particle selection papers to explicitly state what quantities are precisely being measured. This would be analogous to the requirement of most experimental papers that units accompany measurements even if such units are obvious. For instance, the ambiguities we pointed out in the literature may have been avoided, even with the incorrect or nonstandard terminology, if the term false positive rate and related terms had been explicitly defined.

In sum, future advances in particle selection algorithms require not only a common benchmark as provided by Zhu et al. (2004) but also adherence to a consistent set of metrics as described in this work.

Acknowledgments

This work was supported by the Howard Hughes Medical Institute and NIH grants R37 GM 29169 and R01 GM 55440 to J. Frank. We would like to thank Melissa Thomas for assistance with the illustrations.

References

Adiga U, Baxter WT, Hall RJ, Rockel B, Rath BK, et al. Particle picking by segmentation: A comparative study with SPIDER-based manual particle picking. Journal of Structural Biology. 2005; 152:211–220. [PubMed: 16330229]

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57:289–300.

Brodersen, KH.; Ong, CS.; Stephan, KE.; Buhmann, JM. The Binormal Assumption on Precision-Recall Curves. Pattern Recognition (ICPR); 20th International Conference on; 2010. p. 4263-4266.p. 4263-4266.

Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning; ACM, Pittsburgh, Pennsylvania. 2006.

Chen JZ, Grigorieff N. SIGNATURE: A single-particle selection system for molecular electron microscopy. Journal of Structural Biology. 2007; 157:168–173. [PubMed: 16870473]

Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd international conference on Machine learning; ACM, Pittsburgh, Pennsylvania. 2006.

Drummond C, Holte R. Cost curves: An improved method for visualizing classifier performance. Machine Learning. 2006; 65:95–130.

Fawcett, T. HP Labs Tech Report. ROC Graphs: Notes and Practical Considerations for Researchers.

Glaeser RM. Historical background: why is it important to improve automated particle selection methods? Journal of Structural Biology. 2004; 145:15–18. [PubMed: 15065669]

Hall RJ, Patwardhan A. A two step approach for semi-automated particle selection from low contrast cryo-electron micrographs. Journal of Structural Biology. 2004; 145:19–28. [PubMed: 15065670]

Hand DJ, Till RJ. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning. 2001; 45:171–186.

Harauz G, Fong-Lochovsky A. Automatic selection of macromolecules from electron micrographs by component labelling and symbolic processing. Ultramicroscopy. 1989; 31:333–344. [PubMed: 2699113]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Huang Z, Penczek PA. Application of template matching technique to particle detection in electron micrographs. Journal of Structural Biology. 2004; 145:29–40. [PubMed: 15065671]

Kohavi, FPaTFaR. The Case Against Accuracy Estimation for Comparing Induction Algorithms. Proceedings of the Fifteenth International Conference on Machine Learning; Morgan Kaufmann; 1997. p. 445-453.p. 445-453.

Kumar V, Heikkonen J, Engelhardt P, Kaski K. Robust filtering and particle picking in micrograph images towards 3D reconstruction of purified proteins with cryo-electron microscopy. Journal of Structural Biology. 2004; 145:41–51. [PubMed: 15065672]

Mallick SP, Zhu Y, Kriegman D. Detecting particles in cryo-EM micrographs using learned features. Journal of Structural Biology. 2004; 145:52–62. [PubMed: 15065673]

Ogura T, Sato C. Automatic particle pickup method using a neural network has high accuracy by applying an initial weight derived from eigenimages: a new reference free method for single-particle analysis. Journal of Structural Biology. 2004a; 145:63–75. [PubMed: 15065674]

Ogura T, Sato C. Auto-accumulation method using simulated annealing enables fully automatic particle pickup completely free from a matching template or learning data. Journal of Structural Biology. 2004b; 146:344–358. [PubMed: 15099576]

Pascual A, Bárcena M, Merelo JJ, Carazo JM. Mapping and fuzzy classification of macromolecular images using self-organizing neural networks. Ultramicroscopy. 2000; 84:85–99. [PubMed: 10896143]

Pascual-Montano A, Donate LE, Valle M, Bárcena M, Pascual-Marqui RD, et al. A Novel Neural Network Technique for Analysis and Classification of EM Single-Particle Images. Journal of Structural Biology. 2001; 133:233–245. [PubMed: 11472094]

Rath BK, Frank J. Fast automatic particle picking from cryo-electron micrographs using a locally normalized cross-correlation function: a case study. Journal of Structural Biology. 2004; 145:84–90. [PubMed: 15065676]

Roseman AM. Particle finding in electron micrographs using a fast local correlation algorithm. Ultramicroscopy. 2003; 94:225–236. [PubMed: 12524193]

Roseman AM. FindEM--a fast, efficient program for automatic selection of particles from electron micrographs. Journal of Structural Biology. 2004; 145:91–99. [PubMed: 15065677]

Sattar, A.; Kang, B-H.; Sokolova, M.; Japkowicz, N.; Szpakowicz, S. AI 2006: Advances in Artificial Intelligence. Vol. 4304. Springer Berlin / Heidelberg; 2006. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation; p. 1015-1021.p. 1015-1021.

Schapire RE, Singer Y. Improved Boosting Algorithms Using Confidence-rated Predictions. Mach Learn. 1999; 37:297–336.

Shaikh TR, Trujillo R, LeBarron JS, Baxter WT, Frank J. Particle-verification for single-particle, reference-based reconstruction using multivariate data analysis and classification. Journal of Structural Biology. 2008; 164:41–48. [PubMed: 18619547]

Short JM. SLEUTH--a fast computer program for automatically detecting particles in electron microscope images. Journal of Structural Biology. 2004; 145:100–110. [PubMed: 15065678]

Sigworth FJ. Classical detection theory and the cryo-EM particle selection problem. Journal of Structural Biology. 2004; 145:111–122. [PubMed: 15065679]

Singh V, Marinescu DC, Baker TS. Image segmentation for automatic particle identification in electron micrographs based on hidden Markov random field models and expectation maximization. Journal of Structural Biology. 2004; 145:123–141. [PubMed: 15065680]

Sorzano COS, Recarte E, Alcorlo M, Bilbao-Castro JR, San-Martín C, et al. Automatic particle selection from electron micrographs using machine learning techniques. Journal of Structural Biology. 2009; 167:252–260. [PubMed: 19555764]

Umesh Adiga PS, Malladi R, Baxter W, Glaeser RM. A binary segmentation approach for boxing ribosome particles in cryo EM micrographs. Journal of Structural Biology. 2004; 145:142–151. [PubMed: 15065681]

Volkmann N. An approach to automated particle picking from electron micrographs based on reduced representation templates. Journal of Structural Biology. 2004; 145:152–156. [PubMed: 15065682]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Voss NR, Yoshioka CK, Radermacher M, Potter CS, Carragher B. DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. Journal of Structural Biology. 2009; 166:205–213. [PubMed: 19374019]

Weng, CG.; Poon, J. In: Roddick, JF., et al., editors. A New Evaluation Measure for Imbalanced Datasets; Seventh Australasian Data Mining Conference (AusDM 2008); ACS, Glenelg, South Australia. 2008. p. 27-32.p. 27-32.

Witten, IH.; Frank, E. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann; 2005.

Wong HC, Chen J, Mouche F, Rouiller I, Bern M. Model-based particle picking for cryo-electron microscopy. Journal of Structural Biology. 2004; 145:157–167. [PubMed: 15065683]

Woolford D, Hankamer B, Ericksson G. The Laplacian of Gaussian and arbitrary z-crossings approach applied to automated single particle reconstruction. Journal of Structural Biology. 2007a; 159:122–134. [PubMed: 17490891]

Woolford D, Ericksson G, Rothnagel R, Muller D, Landsberg MJ, et al. SwarmPS: Rapid, semi-automated single particle selection software. Journal of Structural Biology. 2007b; 157:174–188. [PubMed: 16774837]

Yu Z, Bajaj C. Detecting circular and rectangular particles based on geometric feature detection in electron micrographs. Journal of Structural Biology. 2004; 145:168–180. [PubMed: 15065684]

Zhu Y, Carragher B, Mouche F, Potter CS. Automatic particle detection through efficient Hough transforms. Medical Imaging, IEEE Transactions on. 2003; 22:1053–1062.

Zhu Y, Carragher B, Glaeser RM, Fellmann D, Bajaj C, et al. Automatic particle selection: results of a comparative study. Journal of Structural Biology. 2004; 145:3–14. [PubMed: 15065668]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

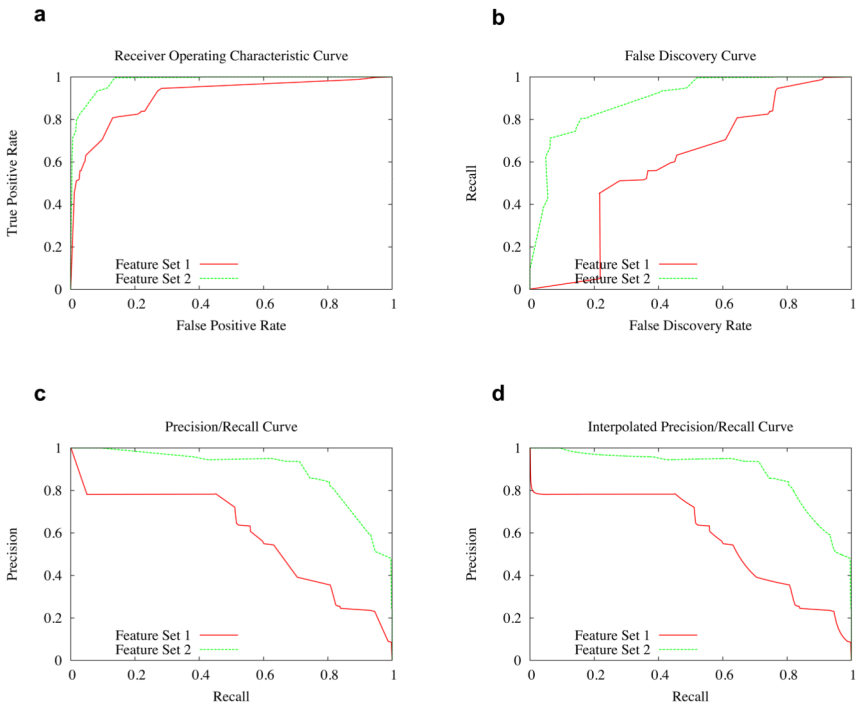

Figure 1.

Metric curves comparing two classifiers on the Keyhole Limpet Hemocyanin dataset, a publicly available dataset used in the 2004 particle selection bakeoff (Zhu et al., 2004). The title “false discovery curve” (b) was coined for lack of a better term. Note that the curve shown in (d) depicts a proper linear interpolation, rather than the non-linear interpolation, which would produce a smooth curve similar to a ROC plot.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

A confusion matrix summarizing the symbols and terms used to define the performance metrics. Each column of the matrix tabulates the number of instances in the actual class and each row the number of instances in the predicted class.

	Actual Class		**Total**
Predicted Class	True Positive (TP)	False Positive (FP)	Total Predicted Positives (P′)
	False Negative (FN)	True Negative (TN)	Total Predicted Negatives (N′)
Total	Total Actual Positives (P)	Total Actual Negatives (N)	