

Comparison of Classification Algorithms on MSTAR Data Using Risk-Based Empirical Statistics

1st Lieutenant Seth B. Wagenman, USAF*

Air Force Operational Test & Evaluation Center (AFOTEC) Detachment 4, Peterson AFB, CO 80914

Lieutenant Colonel Steven N. Thorsen, USAF†

U.S. Air Force Academy, CO 80840

Major David M. Kaziska, USAF‡

and Dr. Mark E. Oxley§

Air Force Institute of Technology, Wright-Patterson AFB, OH 45431

This article is approved for public release, distribution is unlimited.

AFOTEC Public Affairs Release Number 2009-09-22.

Imagine a warfighter analyzing the battlespace. He implements a system that classifies remotely sensed objects according to a set of 200 different possible labels. Unknown to the warfighter or the command element, numerous enemy forces have infiltrated the battlespace, and so the classification system reports a large group of enemy tanks in exactly the same spot where friendly forces were previously stationed—i.e., targets that must be destroyed immediately. The warfighter has never seen such a large concentration of enemy forces in this particular battlespace and begins to question the results, failing to immediately order the destruction of those objects. Time ticks by, and it becomes apparent to the command element and the warfighter that the objects are dangerous enemies after they retreat beyond weapons range, and the window of opportunity to act decisively shrinks away.

If briefed beforehand that the classification system could minimize risk based on the classification cost and battlespace information provided by the command element, the warfighter would have had more confidence in the classification system, and might have quickly taken decisive action.

When comparing classification systems to one another via Receiver Operating Characteristic (ROC) analysis, some comparison methods do not consider the whole picture—i.e., costs and class prevalences along with the class-conditional probabilities. Because the volume under a ROC surface in a 200-class case would be a 39,800-dimensional object, concepts such as Volume Under the Surface (VUS) become rather cumbersome. Most attempts to generalize geometric concepts to the general n -class case choose to ignore either the class prevalences or the costs. The concept of risk allows a much more robust form of ROC analysis to take place, one which considers many more of the characteristics of the operating environment in which the receiver of information resides.

Recent research into comparison methods for classification systems explores continuous fixed-support joint distributions of class prevalences as weighting functions to deal with classification domains about whose class prevalences one has limited or no knowledge. Using empirical statistical methods, a warfighter can calculate the probability and expected

*Test Analyst, Detachment Test Support, 625 Suffolk Street

†Assistant Professor, Department of Mathematical Sciences, 2354 Fairchild Drive Suite 6D158

‡Assistant Professor of Statistics, Department of Mathematics and Statistics, 4950 Hobson Way

§Professor of Mathematics, Department of Mathematics and Statistics, 4950 Hobson Way

cost or loss appropriate to each type of classification decision, assuming costs to be subjectively fixed, and that acceptable estimates for class-conditional probabilities exist. As the sum of the products of cost and probability for all types of classification decisions, total classification risk for a classification system is easily calculated. Empirical risk data produced by statistical simulation of the battlespace lends itself to statistical description of total classification risk for comparison with other classification systems.

An example of a joint distribution for class prevalences over a standard simplex is provided by way of a multivariate triangular distribution. Families of classification systems are created using Probabilistic Neural Nets (PNN) acting on the Moving and Stationary Target Acquisition and Recognition (MSTAR) mixed targets data set. The spread parameter of the PNNs serves as one threshold distinguishing the PNN classification systems from one another, and a second parameter is a cropping proportion used in processing the image data. Using computer simulation, a warfighter can choose a threshold that minimizes risk under the assumption of temporarily fixed costs.

Nomenclature

AER	Actual Error Rate
AUC	Area Under the ROC Curve
p_j	Class Prevalence
c_{ij}	Classification Cost
\mathbf{C}	Classification Cost Matrix $[\mathbf{C}]_{ij} = c_{ij}$
$\pi_{ij}(A)$	Classification Probability
$\mathbf{\Pi}(A)$	Classification Probability Matrix $[\mathbf{\Pi}]_{ij} = \pi_{ij}$
$R_{\mathbf{C}}(A)$	Classification Risk
$A: E \rightarrow L$	Classification System
$q_{ij}(A)$	Class-conditional Probability
$\mathbf{Q}(A)$	Conditional Probability Matrix $[\mathbf{Q}(A)]_{ij} = q_{ij}(A)$
ERRT	Empirical ROC Risk Threshold
\langle, \rangle_F	Frobenius Inner Product
\odot	Hadamard Product Operator
\mathbf{P}	Prevalence Matrix: each row identical to transposed vector \mathbf{p}^T of Class Prevalences $\{p_j\}_{j=1}^n$
PNN	Probabilistic Neural Net
ROC	Receiver Operating Characteristic
RRF	ROC Risk Functional
θ	Threshold Parameter
Θ	Threshold Set
VUS	Volume Under the ROC Surface
<i>Subscripts</i>	
i	Refers to label $\ell_i \in L$
j	Refers to class $\mathcal{E}_j \subset E$
$i j$	Label $\ell_i \in L$ given class \mathcal{E}_j

Note: Familiarity with mathematical symbols such as \in , \cup , \cap , \subset , \sum , \equiv , T , $|$, and \forall , is recommended.

I. Introduction

The process of classification requires an algorithm known as a *classification system*. Given a sample space E of possible outcomes or *events*, along with a finite set $L = \{\ell_1, \ell_2, \ell_3, \dots, \ell_n\}$ of distinct *labels*, a function $A: E \rightarrow L$ is a classification system. An “n-class system” has exactly n labels ($n = 1, 2, 3 \dots$); for example: an Identification, Friend or Foe (IFF) 3-class system that labels objects as “friendly,” “unfriendly,” or “unrecognized.”

A classification system may have a *threshold parameter* θ , selected from a finite-dimensional *threshold set* Θ of parameters that may influence classification. For example, when classifying adults into men and women

based on height alone, a single-dimensional threshold parameter θ might be a point between the median men's and women's heights, where if $\text{Height}(\text{adult}) > \theta$, we label the adult as "man"; otherwise, "woman". If classification entails using a measuring tape, then a two-dimensional threshold set $\Theta = \Theta_1 \times \Theta_2$, might contain ordered pairs $\theta = (\theta_1, \theta_2)$, where θ_1 is the height chosen above and θ_2 the tension on the tape.

When facing a decision of where to set threshold parameters, Bayesian decision theory suggests applying the concept of risk.^{2,5} To calculate the risk $R_{\mathbf{C}}(A_{\theta})$ of a classification system A_{θ} (that is, a general classification system A , specified by a particular choice of θ from the threshold set), we must know each possible outcome and its associated cost and probability. We rely on the formula for classification risk R (suppressing notational dependence on θ), given by³

$$R_{\mathbf{C}}(A) \equiv \langle \mathbf{\Pi}(A), \mathbf{C} \rangle_F \quad (1)$$

where $\mathbf{\Pi}(A)$ is a matrix of classification probabilities, \mathbf{C} is a matrix of classification costs (assumed temporarily fixed), and $\langle \cdot, \cdot \rangle_F$ represents the action of the Frobenius inner product, given by⁵

$$\langle \mathbf{U}, \mathbf{V} \rangle_F \equiv \sum_{i=1}^s \left(\sum_{j=1}^r u_{ij} v_{ij} \right) \quad (2)$$

where $[\mathbf{U}]_{ij} = u_{ij}$ and $[\mathbf{V}]_{ij} = v_{ij}$ are any two matrices of the same size $s \times r$. For an n -class system, with exactly n^2 types of possible classification decisions, $\mathbf{\Pi}(A)$ and \mathbf{C} are both $n \times n$.

We now define $\mathbf{\Pi}(A)$ and \mathbf{C} in precise mathematical terms, beginning with necessary preliminaries.

A. Conditional Probability Matrix

The elements $[\mathbf{Q}(A)]_{ij} = q_{ij}(A)$ of a *conditional probability matrix* are *class-conditional probabilities*. Point estimates of class-conditional probabilities for a classification system are calculated by means of a *confusion matrix* (the most accurate and reliable being one produced using Lachenbruch's holdout procedure^{1,4}). To illustrate, consider a 2×2 contingency matrix of results from a classification experiment where we have explicit knowledge of the number of items in each population. A matrix such as that shown in Table 1 is a simple tally of the numbers of each type of classification decision, both correct and incorrect, with correct decisions on the diagonal and columns corresponding to truth. Here, class 1 is *positive*, and class 2 *negative*; hence, the true positive count TP is how many items of class 1 were correctly labeled, and the false negative count FN is how many were not, and so forth.^{3,6}

Table 1. Two-Class Contingency Matrix.

Contingency Matrix	Actual Class: 1	Actual Class: 2
Labeled Class: 1	TP	FP
Labeled Class: 2	FN	TN

From this matrix, form estimates of the class-conditional probabilities by dividing each element in a column by the number of items in the class corresponding to truth for that column; with M_1 and M_2 items from Classes 1 and 2, respectively, estimates of class-conditional probabilities appear in Table 2.

Table 2. Two-Class Confusion Matrix.

Confusion Matrix	Actual Class: 1	Actual Class: 2
Labeled Class: 1	$\frac{TP}{M_1}$	$\frac{FP}{M_2}$
Labeled Class: 2	$\frac{FN}{M_1}$	$\frac{TN}{M_2}$

The result is a transpose stochastic *confusion matrix*, such that the sum of each column is one; therefore, the information contained in a 2×2 confusion matrix may be represented by an ordered pair comprised of one entry from each column, which may then be plotted on a unit square. As the convex hull of plotted

points from many such classification systems, a curve known as a Receiver Operating Characteristic (ROC) curve is formed and may be analyzed through measures such as the Area Under the ROC Curve (AUC), or Volume Under the ROC Surface (VUS) for classification systems with more than two classes; however, such traditional methods of ROC analysis do not inherently facilitate the calculation of classification risk.⁶

Given a classification system $A: E \rightarrow L$, along with a probability measure $P: \mathcal{E} \rightarrow [0, 1]$ defined on a σ -field \mathcal{E} over E containing all pre-images $A^{\natural}[\{\ell_i\}] \subset E$ of singleton label subsets $\{\ell_i\} \subset L$ (where the Becuadro, \natural , denotes the *set function* $A^{\natural}: L \rightarrow \mathcal{E}$ with pre-images of A in the codomain) and all classes in the partition $\bigcup_{j=1}^n (\mathcal{E}_j) = E$ induced by L on E , the *class-conditional probability* $q_{ij}(A)$ is the *conditional probability* that $A(e) = \ell_i$, given that $e \in \mathcal{E}_j$, and is given by⁵

$$\begin{aligned} q_{ij}(A) &\equiv P\left(A(e) = \ell_i \mid e \in \mathcal{E}_j\right) \\ &= P\left(e \in A^{\natural}[\{\ell_i\}] \mid e \in \mathcal{E}_j\right) \\ &= \frac{P\left(A^{\natural}[\{\ell_i\}] \cap \mathcal{E}_j\right)}{P(\mathcal{E}_j)}, \quad i, j = 1, 2, 3, \dots, n \end{aligned} \quad (3)$$

when class \mathcal{E}_j has $P(\mathcal{E}_j) \neq 0$. For a class \mathcal{E}_j with prior probability $P(\mathcal{E}_j) = 0$, all class-conditional probabilities conditioned on \mathcal{E}_j are given by $q_{ij}(A) = 0$, $\forall i = 1, 2, 3, \dots, n$. A class-conditional probability may take on any value in $[0, 1]$; thus, for each i and j , $q_{ij}(A)$ is a well-defined probability measure with $\sum_{i=1}^n q_{ij}(A) = 1$, $\forall j = 1, \dots, n$ (i.e., each *column* of \mathbf{Q} sums to one).

B. Prevalence Matrix

The elements $[\mathbf{P}]_{ij} = p_j$ of a *prevalence matrix* are *class prevalences*. If we multiply (3) above by the probability $p_j \equiv P(\mathcal{E}_j)$ of being in class \mathcal{E}_j , or the class prevalence of \mathcal{E}_j , the probability of the classification system labeling an outcome $e \in \mathcal{E}_j$ with ℓ_i immediately results; therefore, to make the calculation of such probabilities as simple as possible, the *prevalence matrix* \mathbf{P} for an n -class system is given by⁶

$$\mathbf{P} \equiv \begin{bmatrix} \mathbf{p}^T \\ \vdots \\ \mathbf{p}^T \end{bmatrix}_{n \times n} = \begin{bmatrix} p_1 & \cdots & p_n \\ \vdots & & \\ p_1 & \cdots & p_n \end{bmatrix}_{n \times n} \quad (4)$$

where \mathbf{p}^T is the transposed vector of class prevalences $\{p_j\}_{j=1}^n$ to which each row of \mathbf{P} is identical. Note that since $\bigcup_{j=1}^n (\mathcal{E}_j)$ is a partition of E , $\sum_{j=1}^n p_j = 1$. In other words, each *row* of the stochastic matrix \mathbf{P} sums to one and is identical to all other rows. It is therefore not necessary to label an element of \mathbf{P} with two subscripts as usual, so we subscript elements of \mathbf{P} according to the *column* in which they reside.

C. Classification Probability Matrix

The elements $[\mathbf{\Pi}(A)]_{ij} = \pi_{ij}(A)$ of a *classification probability matrix* are *classification probabilities* given by $\pi_{ij}(A) \equiv q_{ij}(A) p_j = P\left(A^{\natural}[\{\ell_i\}] \cap \mathcal{E}_j\right)$; in other words, the probability that $A(e \in \mathcal{E}_j) = \ell_i$. Since all possible outcomes are accounted for, the *elements* of $\mathbf{\Pi}(A)$ sum to one.

The Hadamard product operator is given by

$$[\mathbf{U} \odot \mathbf{V}]_{ij} \equiv u_{ij} v_{ij} \quad (5)$$

where $[\mathbf{U}]_{ij} = u_{ij}$ and $[\mathbf{V}]_{ij} = v_{ij}$ are any two matrices of the same size. A classification probability matrix is therefore given by $\mathbf{\Pi}(A) \equiv \mathbf{Q}(A) \odot \mathbf{P}$, and so (1) becomes $R_{\mathbf{C}}(A) \equiv \langle \mathbf{\Pi}(A), \mathbf{C} \rangle_F = \langle [\mathbf{Q}(A) \odot \mathbf{P}], \mathbf{C} \rangle_F$.

D. Classification Cost Matrix

The elements $[\mathbf{C}]_{ij} = c_{ij}$ of a *classification cost matrix* are numerical *classification costs* associated with the classification decisions whose probabilities appear in the corresponding positions of $\mathbf{\Pi}(A)$. A general rule of thumb is that correct decisions have a cost of zero (i.e., the diagonal elements of \mathbf{C} are all zero) and incorrect decisions have a positive cost value; however, the framework presented above does not require any such restrictions on the methodology used to apply the cost concept.

Costs are assumed to be authoritative and based on a subjective assessment of the inherent losses (according to some relative numerical scale) associated with making each possible kind of classification decision. For example, one may assume there is no cost associated with classifying a woman as a woman or a man as a man, but the cost associated with classifying a woman as a man may not be exactly equal with that of classifying a man as a woman. Costs may change, but as the framework we present allows for near-realtime re-calculation of classification risk whenever cost structure changes, we assume them temporarily fixed.

To calculate classification risk, simply add all products of cost and probability together, as in (1) above.

II. Method

The calculation of classification risk is very simple if all quantities involved are constants; however, depending on the environment in which classification occurs, there may be significant variability in the classification probabilities. Note that elements of the matrices \mathbf{C} , \mathbf{P} , $\mathbf{Q}(A)$, and $\mathbf{\Pi}(A)$ are actually random variables over the sample space E ; however, for the purposes of this paper, we assume costs c_{ij} to be *constant* random variables. The variables that may have the greatest effect on the classification process are the class prevalences, since they are part of the definition of the class-conditional probabilities and there is statistical dependence between p_j and q_{ij} in many cases.⁶ In addition, the class prevalences are a function of the environment in which classification occurs, and if this is the physical world, such variables may tend to be extremely unpredictable. However, limited knowledge based on expert opinion is better than no knowledge at all.

Recent work on the subject of calculating classification risk in an uncertain classification environment illustrates the framework for constructing a joint distribution of class prevalences, with the restriction that no class may have zero population density. The work employs classical statistical methods to select a classification system based on a point estimate of classification risk via the ROC Risk Functional (RRF)^{5,6}

$$\begin{aligned} \arg \min_{A_\theta \in \mathbb{A}_\Theta} \left\{ E [R_{\mathbf{C}}(A_\theta)] \right\} &\equiv \arg \min_{A_\theta \in \mathbb{A}_\Theta} \left\{ E \left[\left\langle [\mathbf{Q}(A_\theta) \odot \mathbf{P}], \mathbf{C} \right\rangle_F \right] \right\} \\ &\approx \arg \min_{A_\theta \in \mathbb{A}_\Theta} \left\{ \left\langle \widehat{\mathbf{Q}}(A_\theta), \mathbf{C} \odot \mathbf{E}[\mathbf{P}] \right\rangle_F \right\} \end{aligned} \quad (6)$$

where a family $\mathbb{A}_\Theta = \{A_\theta : \theta \in \Theta\}$ of classification systems is defined over a threshold set Θ of parameters. The RRF relies heavily on assumptions of statistical independence to allow a quick calculation, but these assumptions do not hold up to scrutiny.⁶ Therefore, we propose the Empirical ROC Risk Threshold (ERRT)

$$\theta^* \equiv \arg \min_{\theta \in \Theta} \left\{ E [R_{\mathbf{C}}(A_\theta)] + D [R_{\mathbf{C}}(A_\theta)] \right\} \approx \arg \min_{\theta \in \Theta} \left\{ E \left[\left\langle \widehat{\mathbf{\Pi}}(A_\theta), \mathbf{C} \right\rangle_F \right] + D \left[\left\langle \widehat{\mathbf{\Pi}}(A_\theta), \mathbf{C} \right\rangle_F \right] \right\} \quad (7)$$

where $\widehat{\mathbf{\Pi}}(A_\theta)$ is an acceptable estimate of $\mathbf{\Pi}(A_\theta)$, and where $E [R_{\mathbf{C}}(A_\theta)]$ and $D [R_{\mathbf{C}}(A_\theta)]$ are robust measures of central tendency and dispersion for risk, respectively. Stated simply, we choose any threshold parameter θ^* such that the quantity $E [R_{\mathbf{C}}(A_{\theta^*})] + D [R_{\mathbf{C}}(A_{\theta^*})]$ is a minimum over all $\theta \in \Theta$.

With no assumptions regarding the nature of statistical distributions for the variables involved (except for costs, as mentioned above), we use statistical simulation to produce and compare values of the *comparative risk quantity* $E [R_{\mathbf{C}}(A_\theta)] + D [R_{\mathbf{C}}(A_\theta)]$ for as many choices of the parameter θ as suit our needs (i.e., we define the threshold set Θ to be of convenient size). The ERRT has the advantage of considering dispersion, or variability, in addition to a measure of central tendency for the classification risk, unlike the RRF. This allows a given threshold parameter with a slightly higher measure of central tendency for risk to still compete against other threshold parameters if its statistical dispersion of risk is smaller. Threshold parameters with

extremely high statistical dispersion of risk are essentially eliminated from the competition, which tends to result in the selection of a system with smaller statistical dispersion.

To illustrate, we classify the 8-class MSTAR Mixed Targets data set, using between six and seven thousand total data points, varying by image cropping factor. We use a Probabilistic Neural Net (PNN) classifier trained on standardized data containing the first two principal components from a set of processed data.⁵ Processed data for an MSTAR image include the eccentricity of an ellipse fit to the convex hull of edges detected using MATLAB[®] in a mathematically transformed, cropped, noise-reduced version of the image (images with no suitable edges are excluded). Each data point also contains numerical values from the binary file header characterizing the target image, and which were gathered simultaneously with the synthetic aperture radar image pixels, such as bandwidth and dynamic range. The first principal component of the processed data set, heavily loaded against the “X_Velocity” data from the file header, accounts for 99.98% of overall variability, and the second is loaded against eccentricity as described above. The data are standardized for use with the PNN classifier, which employs a common “spread” parameter as the standard deviation of a multivariate normal marginal distribution about each training data point, as illustrated in Figure 1 for a two-dimensional data set. Standardization ensures experimentation with spreads greater than one will not produce different results than a spread of one when implementing the PNN classifier.

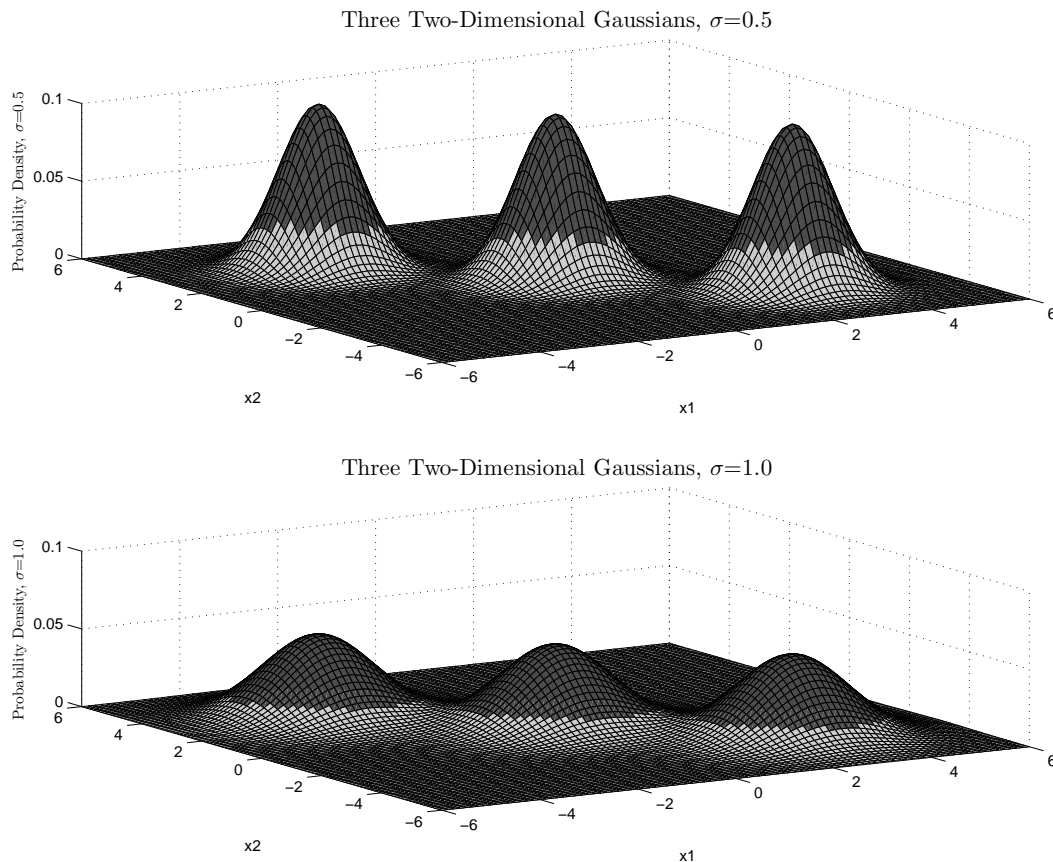


Figure 1. Effects of Differing Spread Parameters for a PNN classifier.

For the purpose of this illustration, we use a two-dimensional threshold set. The chosen PNN spread is along the first parameter axis, and the second parameter axis is a proportion, namely, the ratio of the small side of the remaining image to the original square image side length after cropping appropriately rotated images with a golden section rectangle during the first step in data processing. Proportions of the golden section are used for cropping rectangles due to the rectangular nature of the image targets (mostly tanks or other such vehicles), with the rectangles placed as close as possible to the exact center of the images.

Given target image types BTR-60, 2S1, BRDM-2, D7, T62, ZIL-131, ZSU-23/4, and SLICY, we label these classes one through eight, respectively, and simulate class prevalences for the first seven by means of a jointly triangular distribution over a standard 7-simplex.⁶ The triangular marginal distributions for classes 1 through 7 each have support on $[0, 1]$, with modes evenly spaced every tenth from $\frac{2}{10}$ to $\frac{8}{10}$, respectively. The prevalence of the eight class (SLICY, a fabricated control target) is then given by the difference between one and the sum of the first seven class prevalences previously drawn randomly from their joint distribution.⁶ We simulate four random draws of all eight theoretical class prevalences in this way to account for variability of the class prevalences, and for each of these four random draws we further simulate another four random selections of individual data points from the standardized principal component data according to the theoretical prevalences drawn, to account for variations within the MSTAR data. Although all classes must have non-zero population density for a classification label to have meaning, a classification system operating in the “real world” may not encounter any items of a certain class over a finite time period. We disallow random theoretical class prevalence draws that round to an actual draw of zero population for any class, to more aptly illustrate application of the risk-based comparison theory.

III. Results

We employ three different classification cost matrices to illustrate the effect of cost on risk, with the median and the median absolute deviation from the median as measures of central tendency and dispersion for risk, respectively. A standard cost matrix, such as the one appearing in Table 3, has zeroes on the diagonal and ones everywhere else, and risk calculations with this matrix yield the Actual Error Rate (AER) of the classification system when the Lachenbruch holdout procedure is used for classifier training and validation.^{1,4}

Table 3. Standard Cost Matrix.

Standard Cost Matrix	Actual: 1	Actual: 2	Actual: 3	Actual: 4	Actual: 5	Actual: 6	Actual: 7	Actual: 8
Labeled: 1	0	1	1	1	1	1	1	1
Labeled: 2	1	0	1	1	1	1	1	1
Labeled: 3	1	1	0	1	1	1	1	1
Labeled: 4	1	1	1	0	1	1	1	1
Labeled: 5	1	1	1	1	0	1	1	1
Labeled: 6	1	1	1	1	1	0	1	1
Labeled: 7	1	1	1	1	1	1	0	1
Labeled: 8	1	1	1	1	1	1	1	0

A “low” cost matrix chosen to illustrate the effect of cost on risk appears in Table 4. Figures 2 and 3 show that the same ordered threshold parameter pair $(\frac{1}{3}, \frac{1001}{1000})$ yields minimal comparative risk for both the standard and “low” cost matrices.

Table 4. Low Cost Matrix.

Low Cost Matrix	Actual: 1	Actual: 2	Actual: 3	Actual: 4	Actual: 5	Actual: 6	Actual: 7	Actual: 8
Labeled: 1	0	1	1	1	1	1	1	3
Labeled: 2	1	0	1	1	1	1	1	3
Labeled: 3	1	1	0	1	1	1	1	3
Labeled: 4	1	1	1	0	1	2	1	3
Labeled: 5	1	1	1	1	0	1	1	3
Labeled: 6	1	1	1	3	1	0	1	3
Labeled: 7	1	1	1	1	1	1	0	3
Labeled: 8	5	5	5	5	5	5	5	0

A “high” cost matrix chosen to illustrate the inherent flexibility of the definition of cost appears in Table 5. Decision-makers who provide guidance on cost are free to choose numbers in any way that suits them. For instance, ones along the diagonal (for correct classification decisions) could indicate that there is a financial cost incurred due to the act of classification, regardless of the outcome, and so costs for incorrect decisions

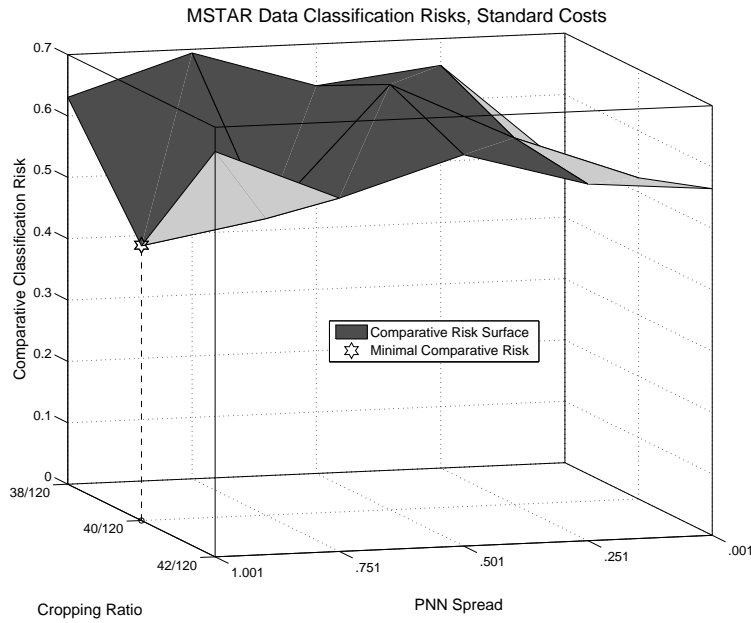


Figure 2. Comparative Risk Surface Over a Two-Dimensional Threshold Set, Standard Costs.

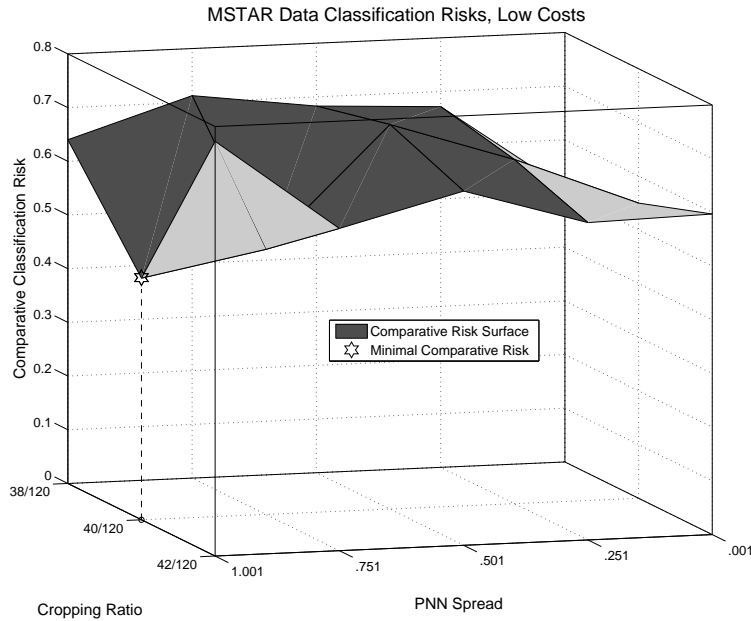


Figure 3. Comparative Risk Surface Over a Two-Dimensional Threshold Set, Low Costs.

might be scaled according to “cost” of a correct decision. Figure 4 illustrates how the entire surface is generally elevated above those of Figures 2 and 3 as a result.

Note that for the “high” cost matrix, an ordered threshold parameter pair $(\frac{1}{3}, \frac{751}{1000})$, distinct from the pair $(\frac{1}{3}, \frac{1001}{1000})$ selected when using either the standard and “low” cost matrices, yields minimal comparative risk. Although the example may be extreme, it does serve to illustrate that cost can be an important factor in risk-based comparison of classification systems. However, even if standard cost matrices are always used, thereby producing only the AER of a classification system, the empirical statistical method presented still presents a sound basis for decision-making.

Table 5. High Cost Matrix.

High Cost Matrix	Actual: 1	Actual: 2	Actual: 3	Actual: 4	Actual: 5	Actual: 6	Actual: 7	Actual: 8
Labeled: 1	1	10	10	10	10	10	10	10
Labeled: 2	2	1	2	2	2	2	10	2
Labeled: 3	2	2	1	2	2	2	10	2
Labeled: 4	2	2	2	1	2	2	10	2
Labeled: 5	2	2	2	2	1	2	10	2
Labeled: 6	2	2	2	2	2	1	10	2
Labeled: 7	2	2	2	2	2	2	1	2
Labeled: 8	2	2	2	2	2	2	10	1

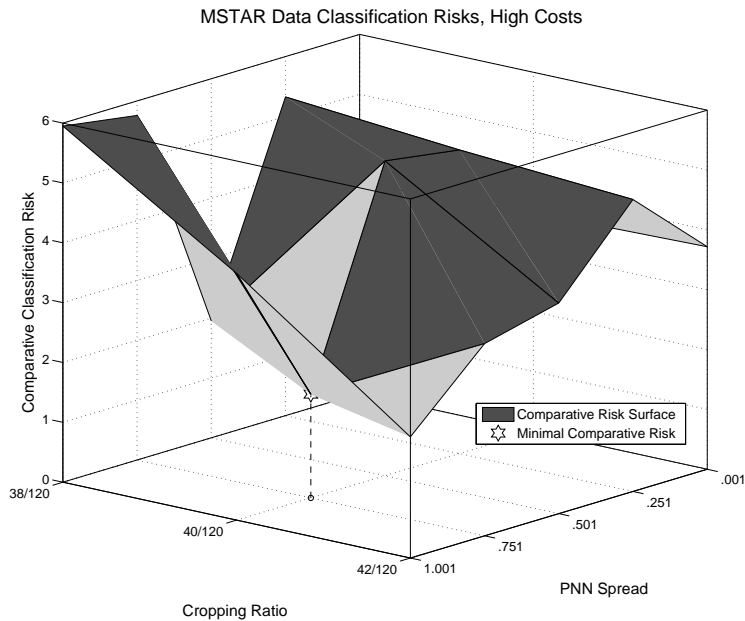


Figure 4. Comparative Risk Surface Over a Two-Dimensional Threshold Set, High Costs.

IV. Conclusion

The result of the theory presented is that risk can be calculated very quickly once the computationally intensive statistical simulation of the classification environment (i.e., the “battlespace”) is completed. The assumption that costs are fixed is acceptable because calculation of the Frobenius inner product of any two matrices (even one whose dimension is 200×200 , for example) can be performed in near-real time by any computer or capable pocket calculator. This allows end users of a classification algorithm to have confidence in decisions made by the system, even (and perhaps especially) when those decisions are surprising, because all possible outcomes, and their associated costs and probabilities, are inherently considered and accounted for by a risk-based approach.

References

- ¹Bauer, K. W., *Air Force Institute of Technology Class Lectures, OPER 685 and OPER 785*, Fall 2007 and Winter 2008, respectively.
- ²Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*, John Wiley and Sons, Inc., New York, 2001 (second ed.).
- ³Egan, J. P., *Signal Detection Theory and ROC Analysis*, Academic Press, New York, 1975.

⁴Lachenbruch, P. A. and Mickey, M. R., *Estimation of Error Rates in Discriminant Analysis*, Technometrics, Volume 10, 1968.

⁵Thorsen, S. N., Wagenman, S. B., Oxley, M. E., and Bauer, K. W. *A Risk-Based Comparison of Classification Systems*, Proceedings of Signal Processing, Sensor Fusion, and Target Recognition XVIII, SPIE Volume 7336, April 2009.

⁶Wagenman, S. B., *Risk-Based Comparison of Classification Systems*, Master's thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, 2008.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 22 SEP 2009	2. REPORT TYPE Conference Proceeding	3. DATES COVERED 16-04-2009 to 14-01-2010			
4. TITLE AND SUBTITLE Comparison of Classification Algorithms on MSTAR Data Using Risk-Based Empirical Statistics		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S) Seth Wagenman; Steven Thorsen; David Kaziska; Mark Oxley		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Operational Test & Evaluation Center, Detachment 4 (AFOTEC DET 4), 625 Suffolk Street, Peterson AFB, CO, 80914-1198		8. PERFORMING ORGANIZATION REPORT NUMBER AFOTEC PA Rel 2009-09-22			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Proceeding from 6th AIAA Biennial National Forum on Weapon System Effectiveness, Tucson, Arizona, 22-24 September 2009.					
14. ABSTRACT As the sum of the products of cost and probability for all types of classification decisions, total classification risk for a classification system is easily calculated. Empirical risk data produced by Monte Carlo simulation of the battlespace lends itself to statistical description of total classification risk for comparison with other classification systems. Families of classification systems are created using Probabilistic Neural Nets (PNN) acting on the Moving and Stationary Target Acquisition and Recognition (MSTAR) mixed targets data set. The spread parameter of the PNNs serves as one threshold distinguishing the PNN classification systems from one another, and a second parameter is a cropping proportion used in processing the image data. Using computer simulation, a warfighter can choose a threshold that minimizes risk under the assumption of temporarily fixed costs.					
15. SUBJECT TERMS Classification, Risk, Costs, Risk Analysis, ROC Analysis, Classification Systems, Performance, MSTAR, Empirical Statistics, Monte Carlo Simulation, PNN					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	9	10	