

Receiver Operating Characteristic Analysis for Detecting Explosives-related Threats[‡]

Alexander Venzin[‡]

Department of Mathematics and Statistics
Graduate School of Engineering and Management
Air Force Institute of Technology
2950 Hobson Way
Wright-Patterson Air Force Base, Ohio 45433

November 6, 2012

Abstract

The Department of Homeland Security (DHS) and the Transportation Security Administration (TSA) are interested in developing a standardized testing procedure for determining the performance of candidate detection systems. This document outlines a potential method for judging detection system performance as well as determining if combining the information from a legacy system with a new system can significantly improve performance. In this document, performance corresponds to the Neyman-Pearson criterion applied to the Receiver Operating Characteristic (ROC) curves of the detection systems in question. A simulation was developed to investigate how the amount of data provided by the vendor in the form of the ROC curve effects the performance of the combined detection system. Furthermore, the simulation also takes into account the potential effects of correlation and how this information can also impact the performance of the combined system.

1 Introduction

The Department of Homeland Security (DHS) is currently searching for methods that are capable of determining the quality and efficiency of potential explosives detection systems. Suppose that system \mathbb{A} represents a legacy x-ray machine currently employed by Transportation Security Administration (TSA) in airports. The x-ray machine is designed to classify the atomic number/density of the material being examined. Suppose further that

*Funding Document: DOE (VBIED Phase II) IA# DE-SC0006992

[†]AFIT\EN Technical Report 11-2012-01

[‡]The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the US Government.

company ‘B’ develops a separate x-ray detection system. Denote this system by \mathbb{B} . Company B is claiming that their system considerably outperforms the legacy system \mathbb{A} , however, the proposed cost of development is a concern. Furthermore, Company B is protected by law from disclosing their proprietary information concerning how results were obtained and the data that was used in their algorithms. Given these constraints, the question becomes whether or not Company B’s detection system is some pre-determined threshold better than the legacy system \mathbb{A} , or, if using the information generated by both systems will produce results worthy of operating both systems simultaneously.

1.1 Research Goals

When detecting events of interest, a combination of detection systems is believed to perform better than individual detection systems. Intuitively, the performance of the combined detection system is reliant on several factors including the nature of the individual systems, the method in which the systems are fused together, and how the systems relate, or correlate, with each other. Because of these factors, the combined system may or may not perform better than an individual system operating at its optimal parameter values. The research goals are:

1. Develop a method for evaluating potential fusion of detection systems.
2. Develop an optimization problem that will generate viable solutions.
3. Examine how correlation effects the fused system performance.

1.2 Research Scope

We consider a set of scenarios that could help to determine the potential effectiveness of a proposed detection system. The first scenario involves Company B sharing their Receiver Operating Characteristic (ROC) curve along with an explanation of how and what the system is designed to detect. In the second scenario, Company B provides three points on their ROC curve. We examine the data when the company does and does not provide details regarding the detection system. In the final scenario, Company B gives a single point on their ROC curve, without any explanation of how their sensor/detection system determines targeted explosive material. Using these scenarios, we attempt to determine bounds for the performance of the combined systems.

1.3 Research Report

The remainder of this report is divided into the following sections: Methods, Results, Simulation, and Conclusions. The methods section presents mathematical background. The results section contains the new mathematical methods derived for use in the simulation section. The simulation section will walk the reader through the three scenarios. The conclusion section discusses the results and highlights future research that have potential to aid in the detection of explosives-related threats.

2 Methods

The mathematical background needed is presented in this section. Detection systems will be discussed in generic terms to illustrate how they can be modeled as a composition of measurable mappings.

2.1 Detection System Theory

Let Γ to be a population set of outcomes, i.e., the underlying event set. Within Γ are outcomes that can be labeled as “explosive material present” and “no explosive material present”. For brevity, let t denote “explosive material present”, i.e., t is short for “target”, and let n denote “no explosive material present”, i.e., n is short for “non-target”. Define the label set $\Lambda = \{t, n\}$. Suppose there exists a truth mapping $\mathbf{T} : \Gamma \rightarrow \Lambda$ such that \mathbf{T} partitions the population set with $\{\Gamma_t, \Gamma_n\}$ where

$$\Gamma_t = \{x \in \Gamma : \mathbf{T}(x) = t\} \quad \text{and} \quad \Gamma_n = \{x \in \Gamma : \mathbf{T}(x) = n\}$$

That is,

$$\Gamma_t \cup \Gamma_n = \Gamma \quad \text{and} \quad \Gamma_t \cap \Gamma_n = \emptyset$$

In this context, a detection system is designed to be the truth mapping \mathbf{T} , or approximate it as close as possible. To do this one constructs a sensor, denoted by s , that produces (raw) datum. Let Δ denote this data set. We assume the sensor is a function, producing a single datum for a given outcome x , that is, $s : \Gamma \rightarrow \Delta$. The raw data is refined, or processed, with a processor p that extracts features $f \in \Phi$. In most circumstances the feature f is a vector of real numbers. The processor is assumed to be a function so $p : \Delta \rightarrow \Phi$. Next one designs a classifier a that inputs a feature vector and outputs a label, t or n , thus, $a : \Phi \rightarrow \Lambda$ is a function. Typically, a classifier has parameters that can be varied. Let Θ be a set of parameters for the classifier, therefore, for each $\theta \in \Theta$ let a_θ denote a classifier, and the collection $\{a_\theta : \theta \in \Theta\}$ is called a family of classifiers. The graphical representation of these mappings is given in the following *diagram*

$$\Gamma \xrightarrow{s} \Delta \xrightarrow{p} \Phi \xrightarrow{a_\theta} \Lambda$$

The composition of these mappings creates a detection system, that is, for every $\theta \in \Theta$, define the composition of the mappings to be

$$A_\theta = a_\theta \circ p \circ s$$

The collection of detection systems $\mathbb{A} = \{A_\theta : \theta \in \Theta\}$ is called a family of detection systems, or simply a detection system family (DSF). Suppose \mathbb{A} represents a legacy x-ray detection system family currently employed by TSA. The x-ray system is designed to classify the atomic number/density of the material being examined. Suppose further that Company B develops a separate x-ray detection system. This system may also be modeled as discussed above.

2.2 Performance of Detection Systems

The Receiver Operating Characteristic (ROC) curve quantitatively assesses detection system performance. The ROC curve is useful as it graphically depicts the difference between false positive rate and true positive rate for each parameter threshold of the detection system. The detection system registers a true positive (TP) when it labels an element from the explosive population event set to which it is tuned as an “explosive.” The detection system registers a false positive (FP) when it labels an element from the non-explosive population event set as an “explosive.” A ROC curve is created by graphing FP rates along the horizontal (X) axis and TP rates along the vertical (Y) axis. The detection system approximates the true ROC curve using empirical data produced during the composition of mappings.

Consider a detection system family \mathbb{A} . Denote the event set composed of explosives that system A_θ is designed to classify as Γ_t . Denote the event set composed of non-explosives system B_π is tuned to classify as Γ_n . Let $P_{TP}(B_\pi)$ denote the probability that B_π correctly labels an outcome $t \in \Gamma_t$ as an “explosive.” This is the definition of a true positive detection by system B_π as the system correctly labeled an explosive with the correct label. Mathematically, this can be modeled using conditional probability

$$P_{TP}(B_\pi) = P\{B_\pi(x) = t : x \in \Gamma_t\} = \frac{P(B_\pi^{-1}(\{t\}) \cap \Gamma_t)}{P(\Gamma_t)}$$

Let $P_{FP}(B_\pi)$ denote the probability that B_π incorrectly labels an outcome $x \in \Gamma_n$ as an “explosive.” This is an instance of a false positive classification by system B_π as it falsely labeled the non-explosive event with an explosive label. This, too, may be modeled using conditional probability

$$P_{FP}(B_\pi) = P\{B_\pi(x) = t : x \in \Gamma_n\} = \frac{P(B_\pi^{-1}(\{t\}) \cap \Gamma_n)}{P(\Gamma_n)}$$

Note that these probabilities are dependent upon the parameters of the detection system. A single probability value is associated with each specific parameter. These probabilities change as the parameter values change. With these two value defined, it is possible to define the *receiver operating characteristic* for system B_π . Define π to be the parameter set for system B_π . Formally, the ROC function and ROC curve are defined as

Definition 1 (*ROC function, ROC curve*) Let $\mathbb{A} = \{A_\theta : \theta \in \Theta\}$ be a detection system family defined on Γ . The ROC function for \mathbb{A} is denoted by $f_{\mathbb{A}}$ and is defined as follows: for every false positive value $p \in [0, 1]$

$$f_{\mathbb{A}}(p) = \max\{P_{TP}(A_\theta) : \theta \in \Theta \text{ and } P_{FP}(A_\theta) \leq p\}.$$

The ROC curve is the graph of the ROC function [5].

2.3 Combining Detection Systems

Let \mathbb{B} denote a detection system family with a different sensor s_2 , processor p_2 , and classifier b_π , for each $\pi \in \Pi$. The diagram for this detection system is

$$\Gamma \xrightarrow{s_2} \Delta_2 \xrightarrow{p_2} \Phi_2 \xrightarrow{b_\pi} \Lambda_2$$

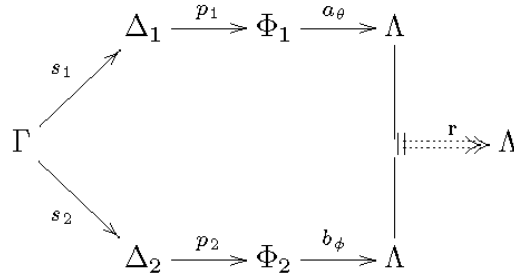
and the composition of the mappings is

$$B_\pi = b_\pi \circ p_2 \circ s_2$$

with $\mathbb{B} = \{B_\pi : \pi \in \Pi\}$ denoting the DSF. Note that the sensor, s_2 , processor, p_2 , and classifier b_π are different from those defined for DSF \mathbb{A} above.

3 Label Rules

Now that a method is in place to create the ROC curve for individual detection systems, it is important to consider how to combine these systems in an attempt to increase predictive power. In order to use the performance of the individual detection systems to quantify the performance of the combined system, there must be a method that accounts for the different explosive and non-explosive partitions that the detection systems are tuned to label. These partitions are addressed using the concepts of within and across label rules.



This diagram depicts the label rule \mathbf{r} that combines the two systems into a new single system.

3.1 Within Label Rule

The within label rule applies to situations where two detection systems are designed to classify the same target and non-target partitions. The rigorous definition follows.

Definition 2 (*Within Label Rule*) Suppose $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$ is a finite label set and $\mathbf{\Lambda}$ is its power set such that $(\Lambda, \mathbf{\Lambda})$ is a measurable space. Let $G_\Lambda = \{\Gamma_{\lambda_1}, \Gamma_{\lambda_2}, \dots, \Gamma_{\lambda_M}\}$ be the truth partition of Γ with respect to Λ . Let G be a σ -algebra on Γ that contains G_Λ then (Γ, G) is a measurable space. If the detection systems $B_1, B_2, \dots, B_N : \Gamma \rightarrow \Lambda$ are measurable mappings designed to map Γ_{λ_m} to the label λ_m for each $m = 1, \dots, M$, then the label rule \mathbf{r} that combines the collection of detection system produces the new detection system

$$B_0 = \mathbf{r}(B_1, B_2, \dots, B_N)$$

is said to be a within-label rule [5].

3.2 Across Label Rules

There are three pre-defined scenarios that may be defined as across type fusion

- 1) Case I – Each detection system labels mutually disjoint target types.
- 2) Case II – One detection system labels a subset of targets of the other detection system.
- 3) Case III – The targets of two detection systems overlap, creating a subset of targets labeled by both systems.

The rigorous definition follows.

Definition 3 (*Across Label Rule*) Suppose $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$ is a finite label set and $\mathbf{\Lambda}$ is its power set such that $(\Lambda, \mathbf{\Lambda})$ is a measurable space. Let $G_\Lambda = \{\Gamma_{\lambda_1}, \Gamma_{\lambda_2}, \dots, \Gamma_{\lambda_M}\}$ be the truth partition of Γ with respect to Λ . Let G be a σ -algebra on Γ that contains G_Λ then (Γ, G) is a measurable space. The partitions $\Lambda^{(0)}, \Lambda^{(1)}, \dots, \Lambda^{(N)} \subset \Lambda$. For each $n = 0, \dots, N$, the integer $M^{(n)} = \text{card}(\Lambda^{(n)}) \leq M$, and the partition $\Lambda^{(n)}$ is congruent to the label set $\Lambda^{(n)} = \{w_1^{(n)}, \dots, w_{M^{(n)}}^{(n)}\}$. For each $n = 0, \dots, N$, the partition $G^{(n)} \subset G$ is the true partition of Γ with respect to $\Lambda^{(n)}$. If the detection systems $B_1 : \Gamma \rightarrow \Lambda^{(1)}, B_2 : \Gamma \rightarrow \Lambda^{(2)}, \dots, B_N : \Gamma \rightarrow \Lambda^{(N)}$ are designed to map each event set $\Gamma_w \in G^{(n)}$ to the corresponding $w \in \Lambda^{(n)}$, then for every $n = 1, 2, \dots, N$, the fusion rule \mathbf{r} that combines the collection of detection systems yielding the new detection system system

$$B_0 = \mathbf{r}(B_1, B_2, \dots, B_N)$$

is said to be an across-label rule [5].

3.3 Boolean AND Label Rule

The AND rule is a binary operation that is defined on the label set, $*$. This operator will be defined using the logical AND symbol, \wedge . It is defined in the following table.

\wedge	t	n
t	t	n
n	n	n

Now, consider detection systems \mathbb{A}_θ and \mathbb{B}_π . The Boolean AND detection system $C_{(\theta, \pi)}^{AND}$ may be defined as the following.

$$C_{(\theta, \pi)}^{AND}(x) = A_\theta(x) \wedge B_\pi(x) \quad \text{for all } x \in \Gamma$$

That is, the combined detection system returns a "t" label (meaning "explosive") only when both detection systems \mathbb{A} and \mathbb{B} label the same outcome as being an explosive.

3.4 Boolean OR Label Rule

The Boolean OR rule is also a binary operation defined on $*$. The OR operator will be defined using the logical OR symbol, \vee . It is defined in the following truth table:

\vee	t	n
t	t	t
n	t	n

Consider systems \mathbb{A} and \mathbb{B} . The Boolean OR label-fused detection system is defined as the following

$$C_{(\theta,\pi)}^{OR}(x) = A_{\theta}(x) \vee B_{\pi}(x) \text{ for all } x \in \Gamma.$$

Using the truth table as a reference, notice that the combined OR detection system labels an element $x \in \Gamma$ as an explosive if either one or both detection systems labels the element as an explosive.

3.5 Within and Across Label Rules

Let "t" denote the pre-determined explosives outcome and let "n" denote the pre-determined non-explosive outcome that \mathbb{A} and \mathbb{B} are tuned to classify. The following table outlines the combined system response depending on the label designation of detection systems \mathbb{A} and \mathbb{B} for two Boolean rules of interest.

Fused Labels	Decision if $x \in \Gamma_t$	Decision if $x \in \Gamma_n$
$A_{\theta}^{-1}(\{t\}) \wedge B_{\pi}^{-1}(\{t\})$	TP	FP
$A_{\theta}^{-1}(\{t\}) \wedge B_{\pi}^{-1}(\{n\})$	FN	TN
$A_{\theta}^{-1}(\{n\}) \wedge B_{\pi}^{-1}(\{t\})$	FN	TN
$A_{\theta}^{-1}(\{n\}) \wedge B_{\pi}^{-1}(\{n\})$	FN	TN
$A_{\theta}^{-1}(\{t\}) \vee B_{\pi}^{-1}(\{t\})$	TP	FP
$A_{\theta}^{-1}(\{t\}) \vee B_{\pi}^{-1}(\{n\})$	TP	FP
$A_{\theta}^{-1}(\{n\}) \vee B_{\pi}^{-1}(\{t\})$	TP	FP
$A_{\theta}^{-1}(\{n\}) \vee B_{\pi}^{-1}(\{n\})$	FP	TP

Regardless of fusion scheme (within/across), these outcomes remain true.

3.6 Correlation

In [5], the functions for the ROC curves associated with the different types of label rules were derived under the assumption that detection systems had some fixed level of correlation. Using the results of [5,7], correlation between two detection systems takes the functional form below.

$$\rho(A_{\theta}, B_{\pi}) = \frac{\mathcal{C}_{A_{\theta} \wedge B_{\pi}} - \mathcal{C}_{A_{\theta}} \mathcal{C}_{B_{\pi}}}{\sqrt{\mathcal{C}_{A_{\theta}}(1 - \mathcal{C}_{A_{\theta}})} \sqrt{\mathcal{C}_{B_{\pi}}(1 - \mathcal{C}_{B_{\pi}})}}$$

where

$$\begin{aligned}
\mathcal{C}_{A_\theta} &= P(\Gamma_t)P_{TP}(A_\theta) + P(\Gamma_n)P_{FP}(A_\theta) \\
\mathcal{C}_{B_\pi} &= P(\Gamma_t)P_{TP}(B_\pi) + P(\Gamma_n)P_{FP}(B_\pi) \\
\mathcal{C}_{A_\theta \wedge B_\pi} &= P(\Gamma_t)P_{TP}(A_\theta \wedge B_\pi) + P(\Gamma_n)P_{FP}(A_\theta \wedge B_\pi)
\end{aligned}$$

The formulas \mathcal{C}_{A_θ} , \mathcal{C}_{B_π} , $\mathcal{C}_{A_\theta \wedge B_\pi}$ are weighted functions involving the false positive and true positive rates corresponding to detection systems \mathbb{A} and \mathbb{B} . They arise naturally from the expressions for the expected value of detection systems \mathbb{A} and \mathbb{B} . When the two detection systems are uncorrelated, the correlation term goes to zero, i.e., when $\mathcal{C}_{A_\theta \wedge B_\pi} = \mathcal{C}_{A_\theta}\mathcal{C}_{B_\pi}$. When these functions are maximized over their respective parameter sets, the values from the ROC curves for detection systems \mathbb{A} and \mathbb{B} may then be used. These maximized functions will be referred to as $T_{\mathbb{A}}(p)$ and $T_{\mathbb{B}}(q)$ for detection systems \mathbb{A} and \mathbb{B} , respectively. These functions take the same general form regardless of choice of within or across label rule, however the prior probability weighting changes with choice of label rule. It is further derived in [5,7] that the ROC function for any AND label fused system where the correlation remains constant for any combination of parameters θ and π takes the following form for every false positive value $r \in [0, 1]$.

$$f_C^{AND}(r) = \frac{1}{P(\Gamma_t)} \max_{pq=r} [g_\rho(T_{\mathbb{A}}(p), T_{\mathbb{B}}(q)) - P(\Gamma_n)r]$$

The function g_ρ is derived from the correlation expression and takes the following form

$$g_\rho(\xi, \eta) = \rho\sqrt{\xi(1-\xi)}\sqrt{\eta(1-\eta)} + \xi\eta.$$

In a similar fashion, the ROC function for the OR label-rule system under the assumption of constant correlation between the detection systems may be defined for every false positive value $r \in [0, 1]$

$$f_C^{OR}(r) = \frac{1}{P(\Gamma_t)} \max_{p+q-pq=r} [h_\rho(T_{\mathbb{A}}(p), T_{\mathbb{B}}(q)) - P(\Gamma_n)r].$$

The function h_ρ is also derived from the correlation expression and is defined as the following.

$$h_\rho(\xi, \eta) = \xi + \eta - \rho\sqrt{\xi(1-\xi)}\sqrt{\eta(1-\eta)} - \xi\eta.$$

For a more in depth discussion of the derivation of these terms, the reader is directed to [3, 5, 6, 7, 9].

4 Results

The goal of this research is to develop a framework for testing how information from detection systems should be combined and the necessary constraints that make the combined system the optimal choice. Assume that detection system family \mathbb{A} corresponds to the legacy system currently employed by the TSA. The competing detection system family and fusion candidate has been modeled by system family \mathbb{B} . Given the ROC functions associated with

these two families, consider the simple Neyman-Pearson optimization criterion for the ROC function of the combined system $f_{\mathbb{C}}$

$$\phi_{NP}(f_{\mathbb{C}}) = \max_{0 \leq p \leq a} f_{\mathbb{C}}(p) = f_{\mathbb{C}}(a).$$

The false positive value of a is fixed and usually chosen between 0.01 and 0.20. The constraint associated with the optimization problem is the following

$$\phi_{NP}(f_{\mathbb{C}}) > \max \{ \phi_{NP}(f_{\mathbb{A}}), \phi_{NP}(f_{\mathbb{B}}) \}.$$

This defines label fusion with respect to the quantifier ϕ_{NP} . The label rule r that combines the two detection system families \mathbb{A} and \mathbb{B} to produce the family $\mathbb{C} = r(\mathbb{A}, \mathbb{B})$ is called “label fusion” with respect to the quantifier ϕ_{NP} if

$$\phi_{NP}(f_{r(\mathbb{A}, \mathbb{B})}) > \max \{ \phi_{NP}(f_{\mathbb{A}}), \phi_{NP}(f_{\mathbb{B}}) \}.$$

4.1 Simulations

In this section, three scenarios are considered to demonstrate the theory. Consider first a scenario where company B supplies a description of the targets that system \mathbb{B} is built to detect and a complete ROC curve. In this instance, company B has developed a detection system that estimates the density and atomic number of the object in question. The legacy detection system currently in use performs the same operation and is therefore designed to find the same type of target. Therefore the within label rule will be used to combine the data. Suppose that the ROC function associated with system \mathbb{A}

$$f_{\mathbb{A}}(p) = p^{1/7}$$

where $p \in S = \{0, 0.01, 0.02, \dots, 1\}$ and the ROC function associated with system \mathbb{B} is

$$f_{\mathbb{B}}(q) = \left(\frac{2}{\pi} \arcsin(q) \right)^{\frac{1}{10}}$$

where $q \in S$. These ROC functions were used so that the two individual ROC curves would cross. Company B is given the benefit of the doubt of being the superior detection system when the false positive rate is low. The variable $\alpha = P(\Gamma_t)$ corresponds to the prior probability of observing a target event. The fusion of these two ROC curves is outlined in the following algorithm.

Algorithm 4 Fusion of two ROC Curves

Let $\mathbf{p} = \mathbf{q} = [0 \ 0.01 \ 0.02 \ \dots \ 1]$;
 $\alpha = \frac{1}{50}$;
 $T_A(\mathbf{p}) = \alpha f(\mathbf{p}) + (1 - \alpha) \mathbf{p}$;
 $T_B(\mathbf{q}) = \alpha f(\mathbf{q}) + (1 - \alpha) \mathbf{q}$;
function [AND/OR] = AND/OR($\rho, r, T_A(\mathbf{p}), T_B(\mathbf{q}), \alpha$)
 $n = 1$;

```

while  $n \leq \text{length}(\boldsymbol{\rho})$ ;
for  $i = 1 : \text{length}(T_A(\mathbf{p}))$ 
    for  $j = 1 : \text{length}(T_B(\mathbf{q}))$ 
        for  $k = 1 : \text{length}(\boldsymbol{\rho})$ 
             $AND(i, j, k) = g_\rho(\boldsymbol{\rho}(k), T_A(\mathbf{p}(i)), T_B(\mathbf{q}(j)), r(i, j), \alpha)$ 
             $OR(i, j, k) = h_\rho(\boldsymbol{\rho}(k), T_A(\mathbf{p}(i)), T_B(\mathbf{q}(j)), r(i, j), \alpha)$ 
        end
    end
end
pg =  $AND/OR(:, :, n)$ ; % i x j x 1 array containing true positive and false positive estimates evaluated
with constant correlation  $\rho(n)$ 
roc =  $\text{sortrows}(\text{cat}(2, r(:, :), pg(:, :)), 1)$ ; %creates (i x j) x 2 matrix of false positive and true positive
pairs and sorts by false positive value.
c2 =  $\text{roc}(:, 2)$ ; %resets true positive rate estimates  $< 0 = 0$  (This can occur depending on correlation
value)
btp =  $\text{find}(c2 < 0)$ ;
if  $\text{isempty}(btp) \sim 1$ ;
c2(btp) = 0;
end
ind =  $\text{find}(c2 > 1)$ ; %resets values  $> 1 = 1$ 
if  $\text{isempty}(ind) \sim 1$ ;
c2(ind) = 1;
end
for  $i = 2 : \text{length}(roc)$ 
c2(i) =  $\max(c2(i - 1), c2(i))$ ;
end
 $AND/OR(:, :, n) = \text{cat}(2, \text{roc}(:, 1), c2)$ ; %Frontier/ROC curve corresponding to  $\rho(n)$  where  $\text{cat}$ 
is the function that concatenates vectors  $\text{roc}(:, 1)$  and  $c2$  along the column
n =  $n + 1$ ;
end

```

Suppose in particular that $a = 0.10$. If it is assumed that the two detection systems are uncorrelated, then the following values of performance may be computed.

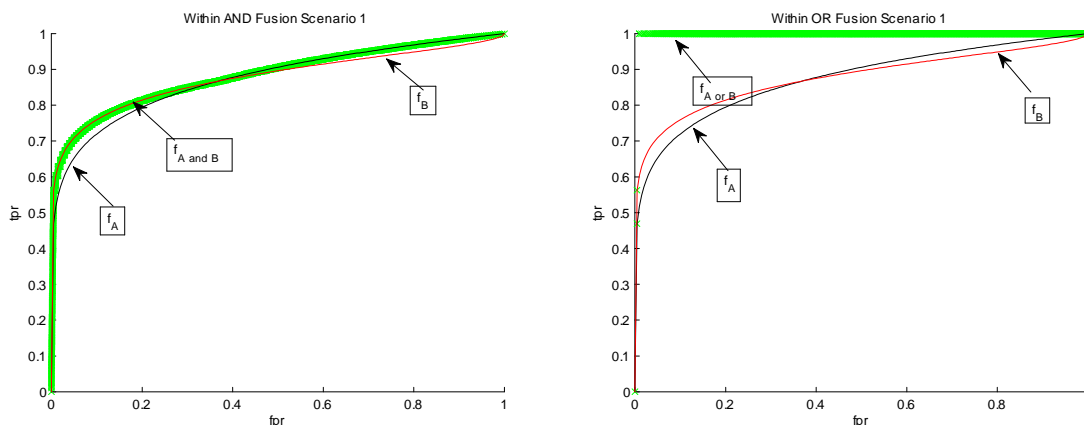


Figure 1: ROC curves for individual and combined system.

<i>System</i>	<i>Performance (a = 0.10)</i>
\mathbb{A}	0.7197
\mathbb{B}	0.7594
$\mathbb{A} \wedge \mathbb{B}$	0.7594
$\mathbb{A} \vee \mathbb{B}$	1

Under the assumption systems are uncorrelated.

Combining the two systems under the Boolean AND rule produces a performance that is only as good as detection system \mathbb{B} and therefore does not meet the constraint. Fusing under the OR rule produces an estimate of perfect classification. Using the results of [5,6], this same estimate can accommodate for varying levels of correlation between individual detection systems \mathbb{A} and \mathbb{B} .

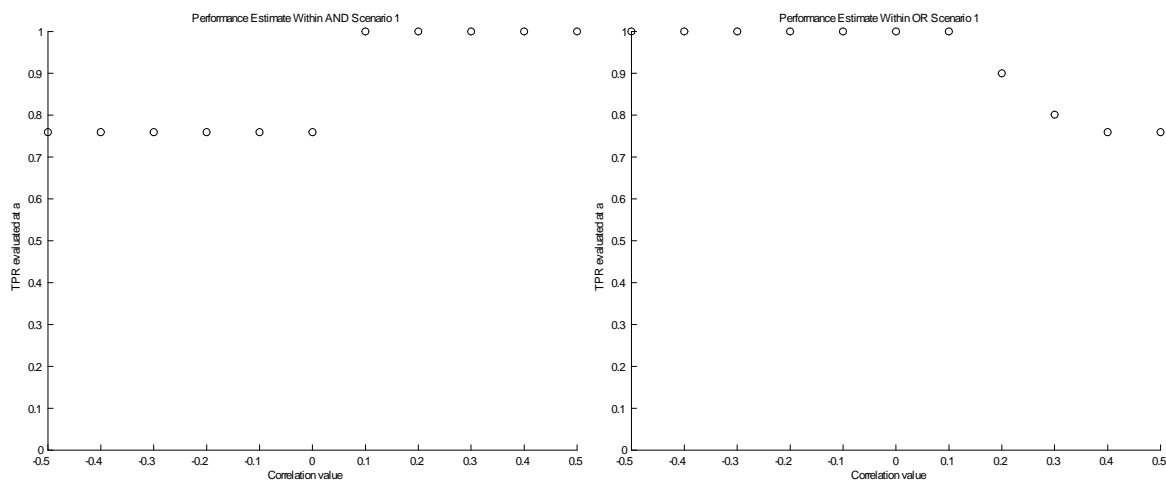


Figure 2: Performance estimates at different levels of constant correlation.

<i>Fusion Rule</i>	<i>min</i>	<i>max</i>
<i>AND</i>	0.7594	1
<i>OR</i>	0.7594	1

min and max over $\rho \in [-\frac{1}{2}, \frac{1}{2}]$

Negatively correlated detection systems combined under the AND rule have decreased performance while the opposite is true of negatively correlated systems combined under the OR rule. Similarly, positively correlated detection systems combined under the AND rule received a boost to performance while the same systems combined under the OR rule demonstrate decreased performance.

Now consider a situation where company B begins to restrict the amount of data they are willing to submit for testing. Company B will no longer provide a complete ROC curve, but rather three points surrounding the optimal point. It is assumed that the points (0,0) and (1,1) constitute two points on the ROC curve. In particular, consider the "ROC curve"

fpr	tpr
0	0
0.09	0.7514
0.10	0.7594
0.11	0.7667
1	1

$f_{\mathbb{B}}$

Given that the performance provided by company B is comprised of only three points, then technically any non-decreasing curve that passes through these points is a valid candidate for the ROC curve corresponding to detection system \mathbb{B} . If company B is further willing to disclose that their detection system is designed to detect a set of explosives and legacy system \mathbb{A} is designed to detect a specific subset of those explosives, then there is enough information to suggest that across type II fusion is most appropriate.

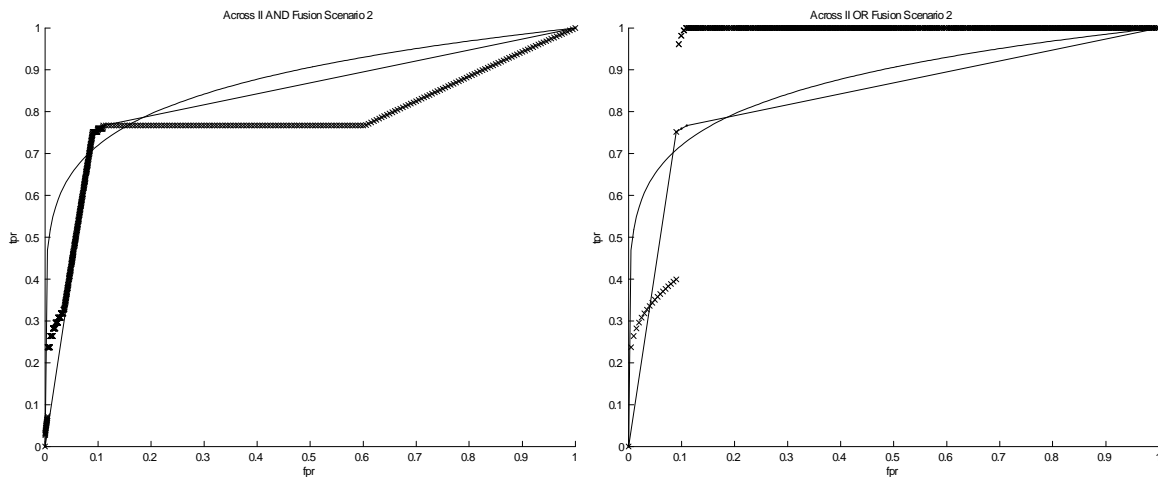


Figure 3: Resultant curves when company B provides only three data points.

$System$	$Performance (a = 0.10)$
\mathbb{A}	0.7197
\mathbb{B}	0.7594
$\mathbb{A} \wedge \mathbb{B}$	0.7594
$\mathbb{A} \vee \mathbb{B}$	0.9816

Under the assumption systems are uncorrelated.

As a reference, consider the resultant ROC curves if more data were present.

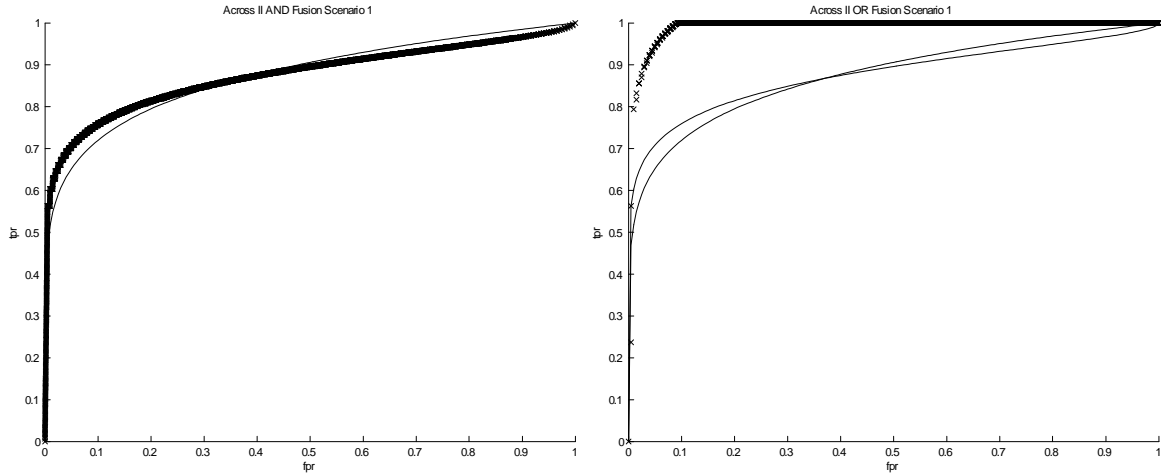


Figure 4: The across II fused ROC curves when company B provides a full ROC curve.

<i>System</i>	<i>Performance</i> ($a = 0.10$)
\mathbb{A}	0.7197
\mathbb{B}	0.7594
$\mathbb{A} \wedge \mathbb{B}$	0.7594
$\mathbb{A} \vee \mathbb{B}$	1

Under the assumption systems are uncorrelated.

Figure 4 represents the "complete" ROC curves corresponding to combination under the across II rule. Notice that the estimate for the performance under the AND rule remains the same regardless of the whether a complete curve or only three data points are present. Interestingly the estimate of performance under the OR decreases marginally when the amount of information becomes restricted.

Suppose now that information regarding the type of explosive outcome detection system \mathbb{B} has been designed to detect is no longer available. Seeing as the partitioning of target events could be different for legacy system \mathbb{A} as compared to system \mathbb{B} , all of the within and across fusion scenarios must be tested exhaustively.

<i>Fusion Rule</i>	<i>Boolean Rule</i>	<i>Performance</i>	<i>min</i>	<i>max</i>
<i>Within</i>	<i>AND</i>	0.7594	0.7594	1
<i>Across I</i>	<i>AND</i>	0.4297	0.4297	1
<i>Across II</i>	<i>AND</i>	0.7594	0.7594	1
<i>Across III</i>	<i>AND</i>	0.5131	0.5131	1
<i>Within</i>	<i>OR</i>	1	0.7594	1
<i>Across I</i>	<i>OR</i>	0.6559	0.4297	1
<i>Across II</i>	<i>OR</i>	0.9816	0.7594	1
<i>Across III</i>	<i>OR</i>	0.6240	0.5131	1

Performance calculated under the assumption systems are uncorrelated. Min and max over $\rho \in [-\frac{1}{2}, \frac{1}{2}]$.

Knowing nothing about the type of fusion that should be applied, it is unclear which output should be regarded as the best estimate of fused performance. Across I and across III label rules can be ignored in this instance as both produce performances which are less than either individual system. If performance was the only consideration, then it is clear that within OR fusion would be the rule of choice. However, It could be that detection system \mathbb{B} has been designed to detect the chemical signature of RDX while detection system \mathbb{A} has been specifically designed to detect nitro-based explosives. This would be categorized as an across I scenario and hence the within label rule would not be applicable. Incorrectly applying the wrong label rule in this instance would lead to a gross overestimation of combined detection system performance.

Finally, consider when the company developing detection system \mathbb{B} provides only one value for the performance of their system as well as no details regarding the targets that are classified by the detection system. It is assumed that the points $(0, 0)$ and $(1, 1)$ constitute two points on the ROC curve. The curve provided by company B is now

fpr	tpr
0	0
0.10	0.7594
1	1

$f_{\mathbb{B}}$

Suppose that the testing agency is considering using the within label rule to combine \mathbb{A} and \mathbb{B} . Given only a single point, the combined ROC curves under the within AND and within OR rules are the following

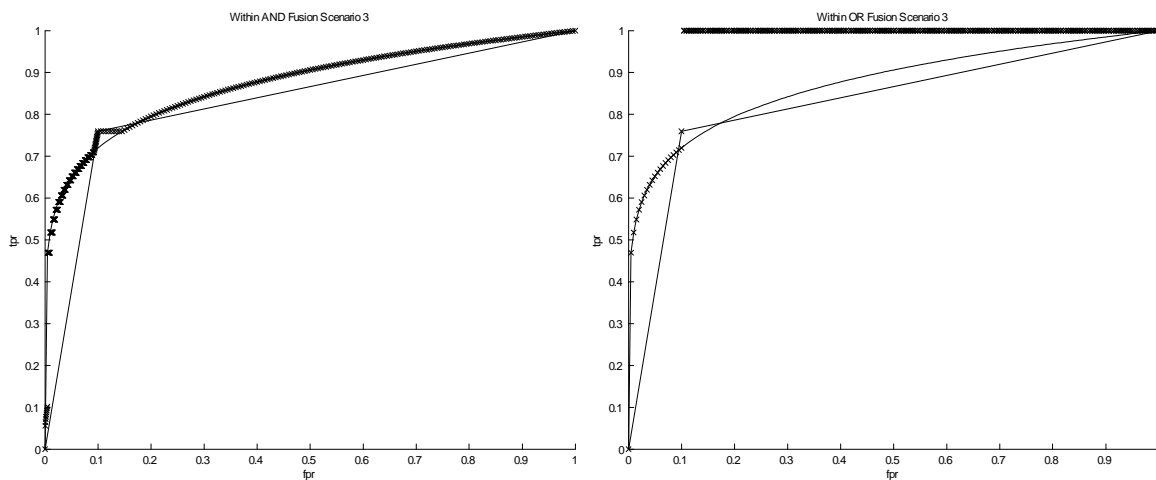


Figure 5: Individual and combined ROC curves when $f_{\mathbb{B}}$ is a single point.

<i>System</i>	<i>Performance ($a = 0.10$)</i>
\mathbb{A}	0.7197
\mathbb{B}	0.7594
$\mathbb{A} \wedge \mathbb{B}$	0.7594
$\mathbb{A} \vee \mathbb{B}$	0.7594

Under the assumption systems are uncorrelated.

This data implies that the within label rule is not applicable. This is surprising given that with as few as three points, fusing under the within OR rule could produce considerably better performance. Given that the most appropriate label rule is unknown, then exhaustively testing all other types may provide at least one potential fusion candidate.

<i>Fusion Rule</i>	<i>Boolean Rule</i>	<i>Performance</i>	<i>min</i>	<i>max</i>
<i>Within</i>	<i>AND</i>	0.7594	0.7594	1
<i>Across I</i>	<i>AND</i>	0.4297	0.4297	1
<i>Across II</i>	<i>AND</i>	0.7594	0.7594	1
<i>Across III</i>	<i>AND</i>	0.5131	0.5131	1
<i>Within</i>	<i>OR</i>	0.7594	0.7594	0.7594
<i>Across I</i>	<i>OR</i>	0.4297	0.4297	0.4297
<i>Across II</i>	<i>OR</i>	0.7594	0.7594	0.7594
<i>Across III</i>	<i>OR</i>	0.5131	0.5131	0.5131

Performance calculated under the assumption systems are uncorrelated. Min and max over $\rho \in [-\frac{1}{2}, \frac{1}{2}]$.

Though the estimate of performance under the AND rule remains unchanged, the estimate under the OR rule has changed for all label rules. Given only this information, the conclusion would be that combination under any rule would not be advised. The lack of information is not only detrimental to TSA in this instance, but also falsely suggests that detection system \mathbb{B} may only have limited usefulness. Recall that it is assumed in this instance that detection system \mathbb{B} has superior performance over the false positive interval of interest. Though every outcome would point to potential funding for company B, it would make more sense to fund a system if it was also a viable fusion candidate in the future.

5 Conclusions

The evidence provided by this research suggests that it is essential that the agency testing new candidate systems collect as much information as possible from the developers. This information is not only useful to the agency, but may also demonstrate the superiority of a particular agencies' detection system. Knowing the type of target and non-target outcome that the candidate system is designed to classify is arguably more important than the volume of data concerning the performance of the system. Having as little as three ROC curve points is potentially enough information to preserve the optimality of the combined performance of two ROC curves.

5.1 Future Research

Potential future research could involve the development of an optimization framework for assessing potential detection systems that meet the standards of DHS and the TSA. Extending the capability of fusion of detection systems beyond Boolean rules as well as more than two detection systems could also be advantageous. Finally, developing a distribution that governs the prior probability of threats could prove useful in applications.

6 Acknowledgements

We acknowledge the financial support for this research from the Department of Energy and Pacific Northwest National Laboratory.

7 References

- [1] J. M. Hill, M. E. Oxley, K. W. Bauer. Receiver operating characteristic curves and fusion of multiple classifiers. Presented at Information Fusion 2003, Proceedings of the Sixth International Conference of 2003.
- [2] M. E. Liggins II, Chee-Yee Chong, I. Kadar, M. G. Alford, V. Vannicola and S. Thomopoulos. Distributed fusion architectures and algorithms for target tracking. Proceedings of the IEEE 85(1), pp. 95-107, 1997.
- [3] M. E. Oxley, S. N. Thorsen, C. M. Schubert, "A Boolean Algebra of Receiver Operating Characteristic Curves," 10th International Conference on Information Fusion, 2007.
- [4] M. Petrakos, I. Kannelopoulos, J. A. Benediktsson and M. Pesaresi. The effect of correlation on the accuracy of the combined classifier in decision level fusion. Presented at Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International. 2000.
- [5] C. M. Schubert, "Quantifying correlation and its effects on system performance in classifier fusion," Ph. d. Dissertation, AFIT/DS/ENC/05-01 Air Force Institute of Technology, Wright-Patterson AFB OH, 2005.
- [6] C. M. Schubert, M. E. Oxley, K. W. Bauer. The inclusion of correlation effects in the performance of multiple sensor and classifier systems. Presented at Aerospace Conference, 2005 IEEE. 2005 .
- [7] C. M. Schubert, M. E. Oxley, K. W. Bauer Jr. A comparison of ROC curves for label-fused within and across classifier systems. Presented at Information Fusion, 2005 8th International Conference on, 2005.
- [8] S. N. Thorsen, M. E. Oxley. A description of competing fusion systems. Information Fusion 7(4), pp. 346-360. 2006.
- [9] A. M. Venzin, "Quantifying Performance Bias in Label Fusion," M. S. Thesis, AFIT/GAM/ENC/12-04 Air Force Institute of Technology, Wright-Patterson AFB, OH 2012.