9-2008

# The ROC Curves of Fused Independent Classification Systems

Michael B. Walsh

THE ROC CURVES OF FUSED INDEPENDENT CLASSIFICATION SYSTEMS

THESIS

Michael B. Walsh

AFIT/GAM/ENC/08/06

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

***AIR FORCE INSTITUTE OF TECHNOLOGY***

**Wright-Patterson Air Force Base, Ohio**

AFIT/GAM/ENC/08/06

# THE ROC CURVES OF FUSED INDEPENDENT CLASSIFICATION SYSTEMS

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Mathematics

Michael B. Walsh, B.A., B.S.

September 2008

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GAM/ENC/08/06

# THE ROC CURVES OF FUSED INDEPENDENT CLASSIFICATION SYSTEMS

Michael B. Walsh, B.A., B.S.

Approved:

| | |
|---|---|
| Dr. Mark E. Oxley<br>Thesis Advisor | date |
| Maj Steven N. Thorsen<br>Committee Member | date |
| Dr. Kenneth W. Bauer, Jr.<br>Committee Member | date |

*Acknowledgements*

I am so thankful for the great minds who have gone before me, shedding light in the darkness and bringing order out of chaos.

My wife and children deserve a lot of credit for the sacrifices they have made, and continue to make, so that I might pursue graduate studies in mathematics.

Michael B. Walsh

*Table of Contents*

*List of Figures*

AFIT/GAM/ENC/08/06


*Abstract*

The need for optimal target detection arises in many different fields. Due to the complexity of many targets, it is thought that the combination of multiple classification systems, which can be tuned to several individual target attributes or features, might lead to more optimal target detection performance. The ROC curves of fused independent two-label classification systems may be generated by the mathematical combination of their ROC curves to achieve optimal classifier performance without the need to test every Boolean combination. The monotonic combination of two-label independent classification systems which assign labels to the same target types results in a lattice of ROC curves which are epimorphic to the corresponding combinations of classification systems. Provided the ROC curves of individual systems are available, testing the lattice of ROC curves in software with existing individual ROC curves can represent a significant cost savings in the design of optimal classification systems.

# THE ROC CURVES OF FUSED INDEPENDENT CLASSIFICATION SYSTEMS

## I. Introduction

### 1.1 Motivation

The need to detect the presence of a target in temporal, spatial, or spectral settings arises in many fields of study; in medicine, the detection of a cancer; in marketing, the detection of the best customer base; in the national command structure, the detection of military targets in a theater of operation. The process of labeling or classifying a target typically begins with a sensor which detects certain attributes, generating raw data. The data might need further processing to allow for the extraction of desired features, which may not be directly measurable. Once criteria necessary for making a decision about the presence of a target is obtained, one can label or classify the targets and non-targets. Since most targets are composed of many parts, it may be necessary to detect multiple attributes prior to accurately assigning the target label. Hence it is often thought that the combination of multiple classification systems, which may use the same or diverse feature sets, gives more accurate and reliable information than the use of a single classification system.

### 1.2 Problem Statement

In two-class scenarios the combination of multiple classification systems may be done in many different ways. What is of interest are combinations yielding the best true positive rates while keeping the false alarm rate below acceptable thresholds. Since we are investigating two-label classification systems, it makes the most sense to use Boolean rules, thereby leveraging all that is known with regard to Boolean Algebras towards our field of target classification. In fact, if we consider the whole set, the empty set, the meet, join, and complement of every Boolean rule, we are indeed generating a Boolean Algebra of Classification System families. How well a particular Boolean combination performs can be quantified by using what is known as a Receiver Operating Characteristic (ROC) curve, which was originally developed to analyze radar signals and employed in signal detection theory [17]. For one to consider any fusion rule other than Boolean, one should be convinced that it performs better than or equal to any combination in our Boolean Algebra. When considering the Boolean combination of multiple classification systems, one would be most interested in finding a combination of Boolean operations on the classification systems which yields optimal performance without having the need to test each combination. This way a substantial cost savings could be realized.

### *1.3 Scope*

For the purpose of this thesis we will investigate the combination of multiple two-label independent classification systems. This is sometimes referred to as decision fusion or label fusion, not to be confused with data fusion, or feature fusion, both of which may occur earlier in the target detection process. We will restrict our attention to classifiers which assign the same target labels. Classifiers which assign labels at different levels in the same genera may be combined as well, but this is known as hierarchical fusion. We will not consider classifiers which have more than two labels (e.g. friend, foe, unknown) since the mathematics and transforms to handle these are beyond the capability of the ROC curve to represent. Independent classification systems will be considered, while correlated ones will be avoided to keep the manipulation of conditional probabilities manageable. Once a target has been classified, further refinements might be made, which can be grouped under what is known as target identification.

# II. Mathematical Background

This chapter will discuss the mathematical theory needed for the main results.

## 2.1 Philosophical Background

A classifier, which assigns a binary label (true/false or target/no target), does so based upon information provided to it, and in reference to a model. The classifier model contains a set of attributes or features germain to the target. The data which is fed to a classification system may be thought of as a combination of noise and signal. As the data is processed through algorithms and filters features are extracted from it. The features or attributes present in the data set are matched against a set of criteria from the model feature set. Each feature-based criterion may be a threshold value (real valued), a binary state (integer/discrete), an m-ary state (e.g. radio button), and the like. Complicated targets having multiple attributes may require more in-depth classifier models. Those attributes which are directly measurable are called data, while those which require processing to extract are called features. Ultimately all of the attributes necessary are assembled to compare with a classifier model to make the classification. For example, a fingerprint from the right index finger is sufficient to identify every living person, and that is based upon a set of attributes resident in that one fingerprint. In most situations, several attributes are compared with the model prior to a classification. In our research, we will restrict ourselves to single attribute based classifiers. This simplifies the math insofar as we can look at one attribute at a time, and if the attribute is real valued the threshold in the model may be varied to allow a look at true positive versus false positive rates as a function of the varied threshold, which we will call the classifier's parameter.

## 2.2 Types of Fusion

There are many ways of fusing outputs from multiple classifiers. Depending on user requirements, fusion may occur at the data, feature, or label phases of the classification process. Our focus will be the fusion of multiple labels to generate a total label.

In many applications, data from multiple sensors are pipelined to a common processor. These data may represent different aspects of an event. It is not as common to find systems which fuse identical data sets unless a further requirement of double, triple, or quadruple redundancy is imposed. Once the data is in a common bucket, whether that be spatial, temporal, or spectral, it may be processed through necessary algorithms to extract desired features. Often it is most economical to get disparate data from multiple sensors to a common CPU, where one can focus all of one's algorithmic development on one processor, and leave the individual sensors less sophisticated and less costly.

Extracting attibutes of interest often narrows the amount of data one has to carry around prior to target classification. If one can obtain the principal components of a matrix of data, in most cases, one can eliminate the components which have the least significance, reducing the size of the model [4]. Transmission of data also plays into this game, since the size of the data pipe might require an intelligent shrinking of the data set so as not to saturate the pipeline. For example, as satellite-borne sensors move towards hyperspectral data collections, with ever increasing data sets, the need to perform onboard feature extraction and feature fusion prior to transmission becomes paramount [14].

The definition of label fusion is the combination of classifier labels after the target/no target assignment has been made. The nice characteristic about this type of fusion is that the amount of data to be handled is quite small. In this research, which is restricted to signature classifiers with one parameter and a label space which is common among the classifiers, we have the ability of capturing the performance of each classifier as a function of its parameter with a ROC curve. If a classifier were to have multiple parameters, one could still generate (many) ROC curves by keeping all other parameters fixed while allowing one parameter to vary. From the collection of ROC curves one could chose the piecewise continuous frontier at each false positive threshold, saving the values of the parameters which yielded each point on the ROC curve frontier.

### *2.3 Example Fusion Scenarios*

The following diagrams show various ways classification systems may be formed and fused.

$$\begin{array}{ccccccccc}
 & & \mathcal{D}_1 & & \mathcal{F}_1 & & \mathcal{L}_1 & & \\
 & \mathbf{s}_1 & \text{Data} & \xrightarrow{\mathbf{p}_1} & \text{Feature} & \xrightarrow{\mathbf{a}_\theta} & \text{Label} & \mathbf{r} & \\
\mathcal{E} & \nearrow & \mathcal{D}_2 & & \mathcal{F}_2 & & \mathcal{L}_2 & \searrow & \mathcal{L} \\
\text{Event} & \mathbf{s}_2 & \text{Data} & \xrightarrow{\textbf{identity}} & \text{Feature} & \xrightarrow{\mathbf{b}_\phi} & \text{Label} & \longrightarrow & \text{Label} \\
 & \searrow & \mathcal{D}_3 & & \mathcal{F}_3 & & \mathcal{L}_3 & \nearrow & \\
 & \mathbf{s}_3 & \text{Data} & \xrightarrow{\mathbf{p}_3} & \text{Feature} & \xrightarrow{\mathbf{c}_\delta} & \text{Label} & &
\end{array}$$

Figure 2.1: Label Fusion of Multiple Classfication Systems.

$$\begin{array}{ccccccccc}
 & & \mathcal{D}_1 & \mathbf{r} & & & \mathcal{F}_1 & & \\
 & \mathbf{s}_1 & \text{Data} & \longrightarrow & \mathcal{D}_4 & \xrightarrow{\mathbf{p}_1} & \text{Feature} & \mathbf{r} & \\
\mathcal{E} & \nearrow & \mathcal{D}_2 & & \text{Data} & & & \searrow & \mathcal{L} \\
\text{Event} & \mathbf{s}_2 & \text{Data} & \nearrow & & & & & \text{Label} \\
 & \searrow & \mathcal{D}_3 & & & & \mathcal{F}_2 & \nearrow & \\
 & \mathbf{s}_3 & \text{Data} & \longrightarrow & \longrightarrow & \xrightarrow{\mathbf{p}_2} & \text{Feature} & &
\end{array}$$

Figure 2.2: Label Fusion After Data Fusion Has Occurred.

$$\begin{array}{ccccccccccc}
 & & \mathcal{D}_1 & & \mathcal{F}_1 & \mathbf{r} & & & & & \\
 & \mathbf{s}_1 & \text{Data} & \xrightarrow{\mathbf{p}_1} & \text{Feature} & \searrow & \mathcal{F}_4 & & \mathcal{L}_1 & \mathbf{r} & \\
\mathcal{E} & \nearrow & \mathcal{D}_2 & & \mathcal{F}_2 & & \text{Feature} & \xrightarrow{\mathbf{a}_\theta} & \text{Label} & \searrow & \mathcal{L} \\
\text{Event} & \mathbf{s}_2 & \text{Data} & \xrightarrow{\textbf{identity}} & \text{Feature} & \nearrow & & & & & \text{Label} \\
 & \searrow & \mathcal{D}_3 & & \mathcal{F}_3 & & & & \mathcal{L}_2 & \nearrow & \\
 & \mathbf{s}_3 & \text{Data} & \xrightarrow{\mathbf{p}_3} & \text{Feature} & \longrightarrow & \longrightarrow & \xrightarrow{\mathbf{b}_\phi} & \text{Label} & &
\end{array}$$

Figure 2.3: Label Fusion After Feature Fusion Has Occurred.

### 2.4 Classification System Theory

The following mathematical treatment is attributed to Schubert, Oxley, and Bauer [13].

Let $\mathcal{E}$ be a population set of outcomes. Let $\mathfrak{E}$ be a $\sigma$-algebra of subsets of $\mathcal{E}$, then $(\mathcal{E},\ \mathfrak{E})$ is a measurable space [10]. Let $P_{\mathcal{E}}$ be a probability measure defined on $\mathfrak{E}$, then $(\mathcal{E}, \mathfrak{E}, P_{\mathcal{E}})$ is a probability measure space[10]. Let $\mathbf{s}$ be a sensor that produces data as its output, i.e., $\mathbf{s}$ is a mapping of outcomes from the population set $\mathcal{E}$ to a datum. Let $\mathcal{D}$ denote the data set. Then we write $\mathbf{s} : \mathcal{E} \to \mathcal{D}$ or its diagram $\mathcal{E} \xrightarrow{\mathbf{s}} \mathcal{D}$. Let $\mathfrak{D}$ be a $\sigma$ -algebra of subsets of $\mathcal{D}$, then $(\mathcal{D}, \mathfrak{D})$ is a measurable space.[2]. A mapping, $\mathbf{p}$, defined on $\mathcal{D}$ is used to produce an element $\mathbf{x}$, called a feature. Let the mapping $\mathbf{p}$ represent a processor that takes a datum from $\mathcal{D}$ and produces a feature, i.e., $\mathbf{p} : \mathcal{D} \to \mathcal{F}$ or its diagram $\mathcal{D} \xrightarrow{\mathbf{p}} \mathcal{F}$ . Since $\mathbf{x}$ is typically a vector of real numbers, then, $\mathcal{F} \subset \mathbb{R}^N$ for some positive integer $N$. Let $\mathfrak{F}$ be a $\sigma$-algebra of subsets from $\mathcal{F}$, then ( $\mathcal{F}, \mathfrak{F}$) is a measurable space. Let $\Theta$ be a threshold set (or a set of parameters); typically, $\Theta = [0, 1]$ or $\Theta = \mathbb{R} = (-\infty, \infty)$. For each $\theta \in \Theta$ let $\mathbf{a}_\theta$ be a classifier mapping $\mathcal{F}$ into a label set $\mathcal{L}$. That is, $\mathbf{a}_\theta : \mathcal{F} \to \mathcal{L}$ or $\mathcal{F} \xrightarrow{\mathbf{a}_\theta} \mathcal{L}$ for each $\theta \in \Theta$. Thus, assume $(\mathcal{L}, \mathfrak{L})$ is a measurable space where $\mathfrak{L}$ is the power set of $\mathcal{L}$. For a two-class problem, examples of a label set could be $\mathcal{L} = \{$true, false$\}$, $\mathcal{L} = \{$T,F$\}, \mathcal{L} = \{0, 1\}$ or even $\mathcal{L} = \{$target,non-target$\}$. For some classifiers the label set is a continuum, e.g., $\mathcal{L} = \mathbb{R}$. In this thesis, $\mathcal{L} = \{t, n\}$ where $t =$ "target" and $n =$ "non-target". The simple graphical representation of these mappings is given in the following diagram.

$$\mathcal{E} \xrightarrow{\mathbf{s}} \mathcal{D} \xrightarrow{\mathbf{p}} \mathcal{F} \xrightarrow{\mathbf{a}_\theta} \mathcal{L} \,.$$

Define the system $\mathbf{A}_\theta$ to be the composition of these mappings for each $\theta \in \Theta$. That is, for each $\theta \in \Theta$, $\mathbf{A}_\theta = \mathbf{a}_\theta \circ \mathbf{p} \circ \mathbf{s}$. Graphically, the diagram for the system is written as

$$\mathcal{E} \xrightarrow{\mathbf{A}_\theta} \mathcal{L} \,.$$

The Receiver Operator Characteristic (ROC) Curve depicts the trade-off between true positives and false positives for every allowable threshold of the classifier. The performance functional (which will be defined later), when applied to the ROC curve, gives the classfication analyst something by which to measure the goodness of a classifier. The following mathematical development introduces the ROC curve.

Each mapping in the classification system, as well as the composition of mappings, has a *pre-image.* The general definition of a pre-image follows [9]:

**Definition 1.** (Pre-image) Let $\mathcal{X}$ and $\mathcal{Y}$ be a nonempty sets. Let the mapping $\mathbf{f}$ take an element from $\mathcal{X}$ and map it into $\mathcal{Y}$, that is, $\mathbf{f} : \mathcal{X} \rightarrow \mathcal{Y}$. Given a subset $Y \subset \mathcal{Y}$ we define the *pre-image* of $\mathbf{f}$ to be the subset in $\mathcal{X}$ by

$$\mathbf{f}^{\natural}(Y) = \{x \in \mathcal{X} : \mathbf{f}(x) \in Y\}.$$

Thus, the pre-image of a subset $Y$ in $\mathcal{Y}$ is all the elements in $\mathcal{X}$ that are mapped by $\mathbf{f}$ into $Y$.

The pre-image is sometimes called the *inverse image* [2], although the mapping $\mathbf{f}$ need not be invertible, yet the superscript $-1$ is used. Because this construction creates a *natural* mapping from subsets of $\mathcal{Y}$ into subsets of $\mathcal{X}$, the natural symbol $\natural$ will be used instead of $-1$. Therefore, we write $\mathbf{f}^{\natural}(Y) = X$. If we consider the entire classification system as a composition of mappings, then we can write the pre-image of a specific label, $\ell \in \mathcal{L} = \{\ell_1, \ldots \ell_n\}$, produced by the classification system $\mathbf{A}_\theta$ . Let $\mathcal{L}_{\ell_i} = \{\ell_i\}$ so that $\mathcal{L} = \mathcal{L}_{\ell_1} \cup \mathcal{L}_{\ell_2} \cup \ldots \mathcal{L}_{\ell_n}$ then $\mathbf{A}_\theta^{\natural}(\mathcal{L}_\ell) = \{e \in \mathcal{E} : \mathbf{A}_\theta(e) \in \mathcal{L}_\ell\}$. The use of pre-images allows us to take the resulting labels and express these in terms of the underlying probabilities [9].

Assume the label set is $\mathcal{L} = \{t, n\}$ where $t$ and $n$ may be real values or symbols and the label $t$ represents a "target" and the label $n$ represents a "non-target". Define $\mathcal{L}_t = \{t\}$ and $\mathcal{L}_n = \{n\}$. The event set $\mathcal{E}$ can be partitioned into a target event set containing all target outcomes and a non-target event set containing non-target outcomes. Denote the true target event set as $\mathcal{E}_t$ and the true non-target event set as $\mathcal{E}_n$. Thus, $\mathcal{E} = \mathcal{E}_t \cup \mathcal{E}_n$ and $\mathcal{E}_t \cap \mathcal{E}_n = \varnothing$.

In order to quantify how well the classification system $\mathbf{A}_\theta$ performs, we appeal to the probability measure space $(\mathcal{E}, \mathfrak{E}, P)$ to compute the following four performance quantifiers. Let $P_{TP}(\mathbf{A}_\theta)$ denote the probability of true positive classification of the classification system $\mathbf{A}_\theta$. Then $P_{TP}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an outcome, $e$, as a target label, $t$, given that the outcome really is a target outcome from the target event set, $\mathcal{E}_t$. Mathematically, $P_{TP}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{TP}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = t \mid e \in \mathcal{E}_t\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_t) \cap \mathcal{E}_t\right)}{P(\mathcal{E}_t)}.$$

Let $P_{FP}(\mathbf{A}_\theta)$ denote the probability of false positive classification of the system $\mathbf{A}_\theta$. Then $P_{FP}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an event outcome, $e$, as a target label, $t$, given that the outcome is really a non-target from the non-target set of the event set, $\mathcal{E}_n$. This is Type II error [7]. Mathematically, $P_{FP}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{FP}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = t \mid e \in \mathcal{E}_n\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_t) \cap \mathcal{E}_n\right)}{P(\mathcal{E}_n)}.$$

Let $P_{TN}(\mathbf{A}_\theta)$ denote the probability of true negative classification of the system $\mathbf{A}_\theta$. Then $P_{TN}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an event outcome, $e$, as a non-target label, $n$, given that the outcome really is a non-target outcome from the non-target event set, $\mathcal{E}_n$. Mathematically, $P_{TN}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{TN}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = n \mid e \in \mathcal{E}_n\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathcal{E}_n\right)}{P(\mathcal{E}_n)}.$$

Let $P_{FN}(\mathbf{A}_\theta)$ denote the probability of false negative classification by the system $\mathbf{A}_\theta$. Then $P_{FN}(\mathbf{A}_\theta)$ is the probability that the classification system $\mathbf{A}_\theta$ labels an event outcome, $e$, as a non-target label, $n$, given that the outcome is really a target outcome from the target event set, $\mathcal{E}_t$. This is known as Type I error [7]. Mathematically, $P_{FN}(\mathbf{A}_\theta)$ is defined by the conditional probability

$$P_{FN}(\mathbf{A}_\theta) = P\{\mathbf{A}_\theta(e) = n \mid e \in \mathcal{E}_t\} = \frac{P\left(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathcal{E}_t\right)}{P(\mathcal{E}_t)}.$$

Note that each of these four probabilities are dependent on the threshold value, $\theta$. A single value for each of these probabilities is computed for each value of $\theta$. As the value

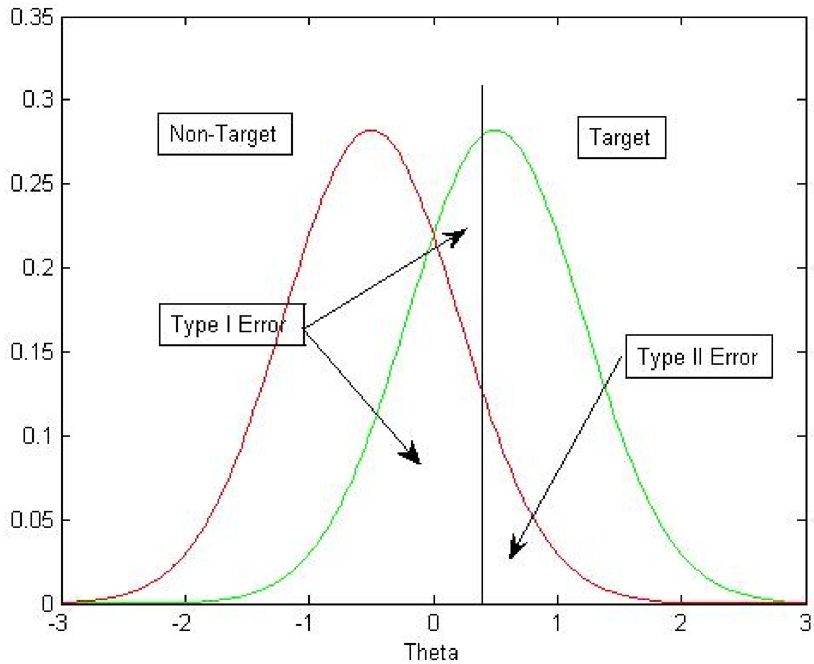


Figure 2.4: Typical Type I and Type II Errors.

of $\theta$ changes, so do the values of $P_{FP}(\mathbf{A}_\theta)$, $P_{TP}(\mathbf{A}_\theta)$, $P_{TN}(\mathbf{A}_\theta)$ and $P_{FN}(\mathbf{A}_\theta)$. A good illustration of these probabilities is found in Figure 2.4.

Define $\Theta$ as a set of possible thresholds and for each $\theta \in \Theta$, and the set of triples

$$\tau_{\mathbb{A}} = \{(\theta, P_{FP}(\mathbf{A}_\theta), P_{TP}(\mathbf{A}_\theta)) : \theta \in \Theta\}$$

to be the *trajectory* of $\mathbb{A}$ [9], [13]. We can project this trajectory onto the second and third component to yield the set

$$f_{\mathbb{A}} = \{(P_{FP}(\mathbf{A}_\theta), P_{TP}(\mathbf{A}_\theta)) : \theta \in \Theta\}.$$

If $\Theta$ is homeomorphic to the real numbers $\mathbb{R}$, then the trajectory $\tau_{\mathbb{A}}$ will be a curve in $\mathbb{R}^3$ and the projection $f_{\mathbb{A}}$ will be a curve in $\mathbb{R}^2$ (more specific, a curve in the unit square $[0,1] \times [0,1]$). Formally, this curve is called the ROC curve for the system family $\mathbb{A}$. For

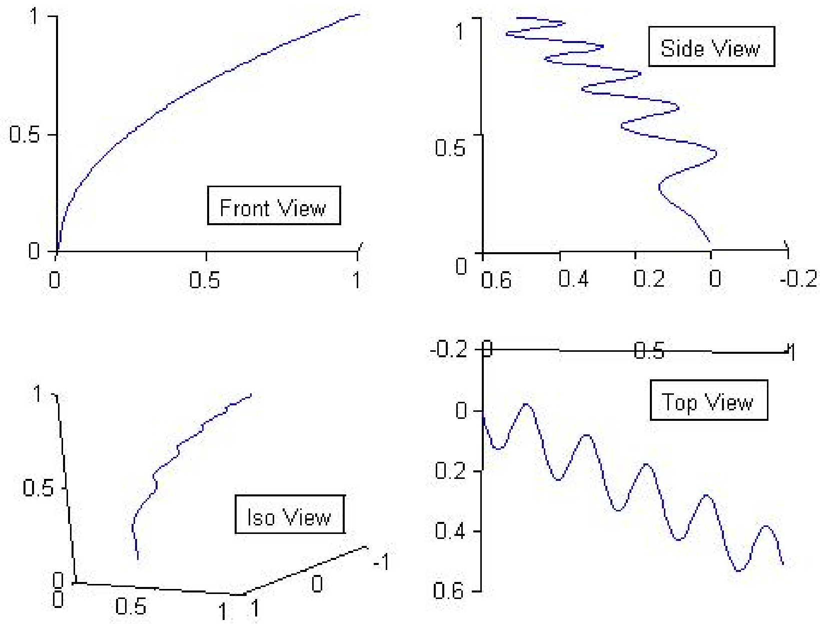


Figure 2.5: A ROC trajectory and its projection.

the case when $\Theta$ is discrete, the ROC "curve" is a set of discrete points. An example of this projection is given in Figure 2.5.

If $\Theta$ is a multi-dimensional set then this analysis will not yield a single curve in the $P_{FP}$-$P_{TP}$ plane. Instead, a collection of curves is created. Therefore, we choose the upper frontier to be the ROC curve as representative of the classifier performance.

**Definition 2.** (ROC function, ROC curve) Let $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ be a family of classification systems defined on the probability space $(\mathcal{E}, \mathfrak{E}, P)$ mapping to the label set $\mathcal{L} = \{t, n\}$ with parameter set $\Theta$. For each $p \in [0, 1]$ , define the set

$$\Theta_p \equiv \{\theta \in \Theta : P_{FP}(\mathbf{A}_\theta) \leq p\}.$$

For $p \in [0, 1]$, if $\Theta_p$ is nonempty then define

$$f_{\mathbb{A}}(p) = \max\{P_{TP}(\mathbf{A}_\theta) : \theta \in \Theta_p\}. \tag{2.1}$$

If $\Theta_p$ is empty then $f_{\mathbb{A}}(p)$ is not defined. The function $f_{\mathbb{A}}$ is called the *ROC function*. The graph of $f_{\mathbb{A}}$ is called the *ROC curve* [9].

In practice, the set $\Theta_p$ may be empty for certain values of $p$. We avoid the discussion of this case and assume that the ROC function is defined for all $p \in [0, 1]$. We make this clear by defining a *total ROC function*.

The set of total ROC functions may be defined as:

$$\mathscr{R} = \{f : [0, 1] \to [0, 1] \mid f \text{ is non-decreasing on } [0, 1]\}.$$

A property of a total ROC curve are given in the following theorem [9].

**Theorem 1.** *Let* $\mathbb{A} = \{\mathbf{A}_\theta : \theta \in \Theta\}$ *be a family of classification systems. Then* $f_{\mathbb{A}}$ *is a non-decreasing function. That is, for every* $p, q \in [0, 1]$ *with* $p \leq q$ *then* $f_{\mathbb{A}}(p) \leq f_{\mathbb{A}}(q)$.

*Proof.* Let $p, q \in [0, 1]$ with $p \leq q$ then $\Theta_p \subseteq \Theta_q$ therefore,

$$f_{\mathbb{A}}(p) = \max_{\theta \in \Theta_p} P_{TP}(\mathbf{A}_\theta) \leq \max_{\theta \in \Theta_q} P_{TP}(\mathbf{A}_\theta) = f_{\mathbb{A}}(q).$$

□

**Definition 3.** (Set of total ROC functions) Let the set of total ROC functions be denoted by

$$\mathscr{R} = \{f : [0, 1] \to [0, 1] \mid f \text{ is non-decreasing on } [0, 1]\}.$$

Notice that we do not require the functions to be continuous.

We write $f = g$ to mean the point-wise equality, that is, $f(p) = g(p)$ for all $p \in [0, 1]$.

### *2.5 Performance Measures*

The Receiver Operator Characteristic (ROC) Curve depicts the trade-off between true positives and false positives for every allowable threshold of the classifier. The performance functional (which will be defined later), when applied to the ROC curve, gives the classfication analyst something by which to measure the goodness of a classifier. The following mathematical development introduces the ROC curve.

The ROC Curve depicts the trade-off between true positives and false positives for every allowable threshold of the classifier. A performance functional, when applied to the ROC curve, gives the classfication analyst something by which to measure the goodness of a classifier. For example, if one pays attention to the upper left corner of the graph, ROC heaven as it were, one might be tempted to grade a ROC curve on how closely it approaches this corner [3]. For example, one might view the point of closest approach of the ROC curve to the corner by the euclidean norm or by the one norm as a measure of how well the system is performing. In the Neyman-Pearson method, a false alarm rate is specified, and the optimal performance is found by the fusion combination that maximizes the vertical distance above the chance line [8]. The chance line joins the vertices (0,0) and (1,1) in the ROC curve diagram. The Bayesian Risk of a ROC curve can be bounded by a line through the ROC heaven corner, whose slope depends upon the weighting between the cost of false negative versus the cost of false positive. If the costs are equal, then the line will be parallel to the chance line (45 deg), and the minimum cost is found by the tangent to the ROC curve that is parallel to the chance line. As the weights are adjusted, the tangent line tracks the angle of the Bayesian Risk bound [5]. Another measure of goodness is the area under the curve (AUC). While this does reward good performers, there are many cases where two ROC curves would achieve the same score, even though one would obviously be better by how much more to the left of the chart it lived than its "equal" counterpart. The Summary Receiver Operator Characteristic SROC curve represents another performance measure which has seen increased emphasis in the statistical community [11].

Consider the case when two sensors, $\mathbf{s}_1$ and $\mathbf{s}_2$, observe outcomes occurring in the same population set $\mathcal{E}$ . Assume they produce data in data sets $\mathcal{D}_1$ and $\mathcal{D}_2$. That is, $\mathbf{s}_1 : \mathcal{E} \to \mathcal{D}_1$ and $\mathbf{s}_2 : \mathcal{E} \to \mathcal{D}_2$. Further, assume sensors $\mathbf{s}_1$ and $\mathbf{s}_2$ each have a processor, $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively, which maps datum in the respective data sets, $\mathcal{D}_1$ and $\mathcal{D}_2$, to features in feature sets $\mathcal{F}_1$ and $\mathcal{F}_2$. In particular, assume $\mathbf{p}_1 : \mathcal{D}_1 \to \mathcal{F}_1$ and $\mathbf{p}_2 : \mathcal{D}_2 \to \mathcal{F}_2$. Suppose that the family of classifiers for $\mathbf{p}_1$ and $\mathbf{s}_1$ is given by $\{\mathbf{a}_\theta : \theta \in \Theta\}$ and that the family of classifiers for $\mathbf{p}_2$ and $\mathbf{s}_2$ is given by another family, $\{\mathbf{b}_\phi : \phi \in \Phi\}$. Let

$\mathbf{a}_\theta : \mathcal{F}_1 \rightarrow \mathcal{L}_1$ for each $\theta \in \Theta$ and $\mathbf{b}_\phi : \mathcal{F}_2 \rightarrow \mathcal{L}_2$ for each $\phi \in \Phi$. Then the labels that are produced from each of the classification systems are fused together to create an overall label for the outcome of interest. The composition of these mappings yield systems represented by the following diagram.

$$
\begin{array}{ccccccccc}
 & & \mathcal{D}_1 & \mathbf{p}_1 & \mathcal{F}_1 & \mathbf{a}_\theta & \mathcal{L}_1 & & \\
 & \mathbf{s}_1 & \text{Data} & \longrightarrow & \text{Feature} & \longrightarrow & \text{Label} & & \\
\mathcal{E} & \nearrow & & & & & & \searrow & \mathcal{L} \\
\text{Event} & & & & & & & & \text{Label} \\
 & \searrow & \mathcal{D}_2 & \mathbf{p}_2 & \mathcal{F}_2 & \mathbf{b}_\phi & \mathcal{L}_2 & \nearrow & \\
 & \mathbf{s}_2 & \text{Data} & \longrightarrow & \text{Feature} & \longrightarrow & \text{Label} & &
\end{array}
$$

We will suppress the text to simplify the diagram to the following

$$
\begin{tikzcd}
 & \mathcal{D}_1 \arrow[r, "p_1"] & \mathcal{F}_1 \arrow[r, "c_1"] & \mathcal{L}_1 \\
\mathcal{E} \arrow[ur, "s_1"] \arrow[dr, "s_2"'] & & & \\
 & \mathcal{D}_2 \arrow[r, "p_2"] & \mathcal{F}_2 \arrow[r, "c_2"] & \mathcal{L}_2
\end{tikzcd}
$$

For these two classification systems the compositions yield the systems $\mathbf{A}_\theta = \mathbf{a}_\theta \circ \mathbf{p}_1 \circ \mathbf{s}_1$ for each $\theta \in \Theta$ and $\mathbf{B}_\phi = \mathbf{b}_\phi \circ \mathbf{p}_2 \circ \mathbf{s}_2$ for each $\phi \in \Phi$. Thus, the individual diagrams are

$$\mathcal{E} \xrightarrow{\mathbf{A}_\theta} \mathcal{L}_1$$

$$\mathcal{E} \xrightarrow{\mathbf{B}_\phi} \mathcal{L}_2$$

and the two families of classification systems are given by $\mathbb{A} \equiv \{\mathbf{A}_\theta : \theta \in \Theta\}$ and $\mathbb{B} \equiv \{\mathbf{B}_\phi : \phi \in \Phi\}$ (See ref. [9]). The two classification systems developed above map outcomes from the population set into different data, feature, and label sets, which are then used to fuse the classification systems together.

There are, however, other ways to label outcomes from the event set. In this paper, classification systems can map outcomes into either the same or different data sets or the same or different feature sets. The sets which must remain the same for the mathematical development contained herein are the event set $\mathcal{E}$ and the two-class label set $\mathcal{L}$. Therefore, the classification systems must be acting from the same event set, map into either the same or different data and feature sets and eventually map into the same label set. These labels are combined together to generate one overall label for that outcome.

In this paper, we assume that the two classification systems are independent or uncorrelated. That is, the occurrence or non-occurrence of an event classified by one system will not affect the occurrence or non-occurrence of another event classified by the other system. This simplifies the expression of conditional probabilities.

We will also only consider two label classifiers.. That is, the label set for all systems considered, including each individual system and the fused classification system, contains two values or two classes. Examples of possible members of this label set were given previously, but the label set considered here is $\mathcal{L} = \{t, n\}$ where $t =$ "target" and $n =$ "non-target".

Using the premises of label fusion, a two-class label system, and classifier independence, representations for a two classification system are developed.

The OR rule is a binary operation defined on $\mathcal{L}$. Define the OR operation by $\vee$. Its definition is given in the table:

| $\vee$ | $t$ | $n$ |
|---|---|---|
| $t$ | $t$ | $t$ |
| $n$ | $t$ | $n$ |

Then the new classification system $\mathbf{C}^{\text{OR}}_{\theta,\phi}$ is defined by the point-wise OR operation

$$\mathbf{C}^{\text{OR}}_{\theta,\phi}(e) = \mathbf{A}_\theta(e) \vee \mathbf{B}_\phi(e) \quad \text{for all } e \in \mathcal{E} \tag{2.2}$$

and yields a new classification system family $\mathbb{C}^{\text{OR}} = \{\mathbf{C}^{\text{OR}}_{\theta,\phi} : \theta \in \Theta, \phi \in \Phi\}$. Thus, to be labeled as "target", either the label from $\mathbf{A}_\theta$ or $\mathbf{B}_\phi$ must be the "target" $t$ label [9].

The AND rule is a binary operation defined on $\mathcal{L}$. We denote this operation by $\wedge$. Its definition is given in the table:

| $\wedge$ | $t$ | $n$ |
|---|---|---|
| $t$ | $t$ | $n$ |
| $n$ | $n$ | $n$ |

The new classification system $\mathbf{C}_{\theta,\phi}^{\text{AND}}$ is defined by the point-wise AND operation on its output, that is,

$$\mathbf{C}_{\theta,\phi}^{\text{AND}}(e) = \mathbf{A}_{\theta}(e) \wedge \mathbf{B}_{\phi}(e) \quad \text{for all } e \in \mathcal{E}. \tag{2.3}$$

This produces a new classification system family $\mathbb{C}^{\text{AND}} = \{\mathbf{C}_{\theta,\phi}^{\text{AND}} : \theta \in \Theta, \phi \in \Phi\}$. Thus, to be labeled as "target", both the label from $\mathbf{A}_{\theta}$ and $\mathbf{B}_{\phi}$ must be the target "$t$ " label [9].

The NOT rule is a unary operation defined on $\mathcal{L}$. We denote the NOT operation by $\neg$. Its definition is given in the table:

| $\neg$ | $t$ | $n$ |
|---|---|---|
| | $n$ | $t$ |

Then the new classification system $\neg\mathbf{A}_{\theta}$ is defined by the point-wise NOT operation

$$[\neg\mathbf{A}_{\theta}]\,(e) \equiv \neg[\mathbf{A}_{\theta}(e)] \quad \text{for all } e \in \mathcal{E} \tag{2.4}$$

and yields a new classification system family $\mathbb{C}^{\text{NOT}} = \{\neg\mathbf{A}_{\theta} : \theta \in \Theta\}$. Thus, to be labeled as "target", the label from system $\mathbf{A}_{\theta}$ must be the "non-target" $n$ label. For brevity we write $\neg\mathbb{A} = \mathbb{C}^{\text{NOT}}$ . Clearly, the NOT rule is not a fusion rule, but it will be used in certain situations [9].

A fusion rule is a method of combining multiple classifiers presumably with the intent of achieving better performance. Since the outcome of our classifiers is binary, either target/no target, then it is reasonable that some Boolean rule might express the optimal combination of classifiers. It is a well known fact that the total number of possible binary outputs of $k$ combinations of two label classifiers is $2^{2^k}$ [16]. Listed below is all possible binary outcomes with just $k = 2$ binary classifiers.

| $u_1$ | $u_2$ | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | $f_{12}$ | $f_{13}$ | $f_{14}$ | $f_{15}$ | $f_{16}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |

Figure 2.6: Possible Fusion Rules for Two Binary Decisions [16].

One can demonstrate that, using ANDs, ORs, and NOTs one can populate or cover this entire set. But that is not the object here. Many of the rules listed in Figure 2.6 are poor performers. Identifying the best performers without having to evaluate every possible combination will be desirable. Varshney shows that including monotonic fusion rules eliminates the majority of the poor performers. A detailed definition of monotonic fusion rules is found on pages 63-64 of Varshney [16]. Before proceeding further, a formal definition of fusion is: Let $\mathfrak{R}$ be a set of rules:

$$\mathfrak{R} = \{\mathbf{r} : \mathcal{L} \times \mathcal{L} \longrightarrow \mathcal{L}\}$$

Label fusion with respect to $\rho$ is:

$$\rho(\mathbf{r}^*(\mathbb{A}, \mathbb{B})) = \max_{\mathbf{r} \in \mathfrak{R}} \rho(\mathbf{r}(\mathbb{A}, \mathbb{B})) > \max\{\rho(\mathbb{A}), \rho(\mathbb{B})\} \tag{2.5}$$

Monotonicity of fusion rules is analogous to the monotonicity of switching functions or finding the simplest *disjunctive normal form* for a given truth function [6]. Examples of poor performers which never escape the chance line are the constant fusion rules which either always assign the target label, or always the non-target label. When fusing multiple classification systems, exclusionary rules are not desirable. Suppose we had classification systems A, B, and C. A fusion rule which says always believe A and disregard systems B and C is not a fusion rule, since it does not deliver results strictly better than any individual classification system in Equation 2.5. If each classifier is doing better than chance, it will become evident that Boolean meet(AND) and join(OR) are all that is needed to optimize the label fusion.

The focus of our research will be restricted to label fusion. Further we require that each classifier being combined has identical labels. Since the classifiers being combined are ordered towards the same target/no target label, and due the desire for monotonic optimal fusion rules, we only need consider the Boolean"OR" and the Boolean"AND". Use of the"NOT" unary operator is not necessary for optimal fusion. There are no exceptions. In the case of the classifier that operates under the chance line, a application of the"NOT" to that classifier would precede any attempt to fuse with other classifiers.

### *2.6 Boolean Algebra*

The definition of a Boolean Algebra is given below [9] [1].

**Definition 4.** A Boolean Algebra is an algebraic structure, denoted by $(\mathscr{A}, =, \wedge, \vee, \neg)$ where

$\mathscr{A}$ is a nonempty set of elements;

$=$ denotes element equality;

$\wedge$ is a binary operation on elements in $\mathscr{A}$ called AND, conjunction, or meet;

$\vee$ is a binary operation on elements in $\mathscr{A}$ called OR, disjunction, or join;

$\neg$ is a unary operation on elements in $\mathscr{A}$ called NOT or negation (or complementation).

And the following axioms hold true:

1. $\mathscr{A}$ is closed w.r.t. $\wedge, \vee$ and $\neg$. For every $a, b \in \mathscr{A}$

$$a \wedge b \in \mathscr{A} \qquad a \vee b \in \mathscr{A} \qquad \overline{a} \in \mathscr{A}$$

2. $\mathscr{A}$ is associative w.r.t. $\wedge$ and $\vee$. For every $a, b, c \in \mathscr{A}$

$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \qquad (a \vee b) \vee c = a \vee (b \vee c)$$

3. $\mathscr{A}$ is commutative w.r.t. $\wedge$ and $\vee$. For every $a, b \in \mathscr{A}$

$$a \wedge b = b \wedge a \qquad a \vee b = b \vee a$$

4. $\mathscr{A}$ has unique identities w.r.t. $\wedge$ and $\vee$. There exists unique elements $l, u \in \mathscr{A}$ such that for every $a \in \mathscr{A}$

$$a \wedge u = a \qquad a \vee l = a$$

5. $\mathscr{A}$ is absorptive w.r.t. $\wedge$ and $\vee$. For every $a, b \in \mathscr{A}$

$$a \wedge (a \vee b) = a \qquad a \vee (a \wedge b) = a$$

6. $\mathscr{A}$ is distributive w.r.t. $\wedge$ and $\vee$. For every $a, b, c \in \mathscr{A}$

$$\begin{aligned} a \wedge (b \vee c) &= (a \wedge b) \vee (a \wedge c) \\ a \vee (b \wedge c) &= (a \vee b) \wedge (a \vee c) \end{aligned}$$

7. $\mathscr{A}$ contain complements w.r.t. $\wedge$ and $\vee$. For every $a \in \mathscr{A}$

$$a \wedge \overline{a} = l \qquad a \vee \overline{a} = u$$

There are several other properties that follow from these axioms, see [1] for a larger list.

## *2.7 Lattice*

**Definition 5.** A Lattice is an algebraic structure, denoted by $(\mathscr{L}, =, \wedge, \vee)$ where

$\mathscr{L}$ is a nonempty set of elements;

$=$ denotes element equality;

$\wedge$ is a binary operation on elements in $\mathscr{L}$ called AND, conjunction, or meet;

$\vee$ is a binary operation on elements in $\mathscr{L}$ called OR, disjunction, or join.

And the following axioms hold true:

1. $\mathscr{L}$ is closed w.r.t. $\wedge and \vee$. For every $a, b \in \mathscr{L}$

$$a \wedge b \in \mathscr{L} \qquad\qquad a \vee b \in \mathscr{L}$$

2. $\mathscr{L}$ is associative w.r.t. $\wedge$ and $\vee$. For every $a, b, c \in \mathscr{L}$

$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \qquad (a \vee b) \vee c = a \vee (b \vee c)$$

3. $\mathscr{L}$ is commutative w.r.t. $\wedge$ and $\vee$. For every $a, b \in \mathscr{L}$

$$a \wedge b = b \wedge a \qquad\qquad a \vee b = b \vee a$$

4. $\mathscr{L}$ has unique identities w.r.t. $\wedge$ and $\vee$. There exists unique elements $l, u \in \mathscr{L}$ such that for every $a \in \mathscr{L}$

$$a \wedge u = a \qquad\qquad a \vee l = a$$

5. $\mathscr{L}$ is absorptive w.r.t. $\wedge$ and $\vee$. For every $a, b \in \mathscr{L}$

$$a \wedge (a \vee b) = a \qquad\qquad a \vee (a \wedge b) = a$$

The lattice contains only combinations of meets and joins, which are the monotonic subset of all possible combinations of meet, joins, and complements in the Boolean algebra (See page 41 of [15]). When we investigate the optimum combination of classifiers, we are only interested in unique monotonic rules [16].

## *III. Main Results*

### *3.1 Introduction*

If it can be established that there exists an epimorphism between a Boolean algebra of ROC curves and a Boolean algebra of classification system families, then finding the best combination of classification system families can be done by combining their respective ROC curves, obviating the need for additional testing of each and every combination. Given how expensive it might be to generate each datum on a single ROC curve, imagine how expensive it would become to generate each datum on every ROC curve arising out of a Boolean algebra. To put this in perspective, for any k two-label classifiers, there are $2^{2^k}$ possible Boolean combinations [16]:

| Number of Systems | Number of Fusion Rules |
|---|---|
| 2 | 16 |
| 3 | 256 |
| 4 | 65,536 |

For the purposes of label fusion of identical labels, utilizing ROC curves which all fall above the chance line, there is no need to include the NOT as it would be counterproductive towards improving overall classification performance. Utilizing only the Boolean join and meet, our Boolean algebra reduces to a lattice[16]. If we can show that a join between two ROC curves is epimorphicly equivalent (onto) to an OR (join) between their respective classification systems, and if we can show the same for the AND (meet), then we can show it for any finite combination using meets and joins. Since we are interested in optimizing the best combination of a finite number of classification systems, it will be good to know that what we learn from optimizing ROC performance will be equivalent to optimization of the classification systems.

The first step will be to show that the meet and join of ROC curves is equivalent to the AND and OR of classification families.

*3.2 OR Formula*

We will capitalize on the development of Schubert [12] which proves the formula for the OR of ROC curves. We start with the development of $P_{TP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi)$ and $P_{FP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi)$ [12]. Recall that $\mathcal{L}_n = \{n\}$.

$$\begin{aligned}
P_{TP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi) &= 1 - P_{FN}(\mathbf{A}_\theta \vee \mathbf{B}_\phi) \\
&= 1 - P([\mathbf{A}_\theta \vee \mathbf{B}_\phi]^\natural(\mathcal{L}_n)|\mathcal{E}_t) \\
&= 1 - P(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathbf{B}_\phi^\natural(\mathcal{L}_n)|\mathcal{E}_t) \\
&= 1 - P(\mathbf{A}_\theta^\natural(\mathcal{L}_n)|\mathcal{E}_t) P(\mathbf{B}_\phi^\natural(\mathcal{L}_n)|\mathcal{E}_t) \quad \textit{by independence} \\
&= 1 - P_{FN}(\mathbf{A}_\theta) P_{FN}(\mathbf{B}_\phi) \\
&= 1 - [1 - P_{TP}(\mathbf{A}_\theta)]\,[1 - P_{TP}(\mathbf{B}_\phi)] \\
&= P_{TP}(\mathbf{A}_\theta) + P_{TP}(\mathbf{B}_\phi) - P_{TP}(\mathbf{A}_\theta) P_{TP}(\mathbf{B}_\phi). \qquad (3.1)
\end{aligned}$$

$$\begin{aligned}
P_{FP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi) &= 1 - P_{TN}(\mathbf{A}_\theta \vee \mathbf{B}_\phi) \\
&= 1 - P([\mathbf{A}_\theta \vee \mathbf{B}_\phi]^\natural(\mathcal{L}_n)|\mathcal{E}_n) \\
&= 1 - P(\mathbf{A}_\theta^\natural(\mathcal{L}_n) \cap \mathbf{B}_\phi^\natural(\mathcal{L}_n)|\mathcal{E}_n) \\
&= 1 - P(\mathbf{A}_\theta^\natural(\mathcal{L}_n)|\mathcal{E}_n) P(\mathbf{B}_\phi^\natural(\mathcal{L}_n)|\mathcal{E}_n) \quad \textit{by independence} \\
&= 1 - P_{TN}(\mathbf{A}_\theta) P_{TN}(\mathbf{B}_\phi) \\
&= 1 - [1 - P_{FP}(\mathbf{A}_\theta)]\,[1 - P_{FP}(\mathbf{B}_\phi)] \\
&= P_{FP}(\mathbf{A}_\theta) + P_{FP}(\mathbf{B}_\phi) - P_{FP}(\mathbf{A}_\theta) P_{FP}(\mathbf{B}_\phi). \qquad (3.2)
\end{aligned}$$

Let $r \in [0,1]$ be a value for the probability of false positive for the fused classifier $\mathbf{C}_{\theta,\phi}^{\mathrm{OR}} = \mathbf{A}_\theta \vee \mathbf{B}_\phi$, then $f_{\mathbf{A}_\theta \vee \mathbf{B}_\phi}(r)$ is the value of the probability of true positive for classifier $\mathbf{C}_{\theta,\phi}^{\mathrm{OR}}$. For $p, q, r \in [0,1]$ define

$$\begin{aligned}
\Theta_p &= \{\theta \in \Theta : P_{FP}(\mathbf{A}_\theta) = p\} \\
\Phi_q &= \{\phi \in \Phi : P_{FP}(\mathbf{B}_\phi) = q\} \\
\Psi_r^{\mathrm{OR}} &= \{(\theta, \phi) \in \Theta \times \Phi : P_{FP}(\mathbf{C}_{\theta,\phi}^{\mathrm{OR}}) = r\}.
\end{aligned}$$

**Theorem 2.** *For every $r \in [0,1]$ then*

$$\Psi_r^{OR} = \bigcup_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \Theta_p \times \Phi_q.$$

*Proof.* Choose $r \in [0,1]$ and let it be fixed. Let

$$(\theta', \phi') \in \bigcup_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \Theta_p \times \Phi_q$$

then there exists some $p'$,$q' \in [0,1]$ such that $\theta' \in \Theta_{p'}$ , $\phi' \in \Phi_{q'}$ , and $p' + q' - p'q' = r$. From the definitions of $\Theta_p$, $\Phi_q$, and $\Psi_r^{\mathrm{OR}}$ we have

$$\begin{aligned} p' &= P_{FP}(\mathbf{A}_{\theta'}) \\ q' &= P_{FP}(\mathbf{B}_{\phi'}) \\ r &= P_{FP}(\mathbf{C}^{\mathrm{OR}}_{\theta,\phi}) \end{aligned}$$

which implies

$$P_{FP}(\mathbf{A}_{\theta'}) + P_{FP}(\mathbf{B}_{\phi'}) - P_{FP}(\mathbf{A}_{\theta'})P_{FP}(\mathbf{B}_{\phi'}) = r = P_{FP}(\mathbf{C}^{\mathrm{OR}}_{\theta',\phi'})$$

and so

$$(\theta', \phi') \in \Psi_r^{\mathrm{OR}}.$$

thus,

$$\bigcup_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \Theta_p \times \Phi_q \subseteq \Psi_r^{\mathrm{OR}}.$$

On the other hand, let $(\theta', \phi') \in \Psi_r^{\mathrm{OR}}$ then $P_{FP}(\mathbf{C}^{\mathrm{OR}}_{\theta',\phi'}) = r$. Observe that $\theta' \in \Theta_{p'}$ for some $p' \in [0,1]$ and $\phi' \in \Phi_{q'}$ for some $q' \in [0,1]$ we have that $1-p' \in [0,1]$ and $1-q' \in [0,1]$ so that $(1-p')(1-q') \in [0,1]$ and therefore,

$$1 - (1-p')(1-q') \in [0,1].$$

Since

$$1-(1-p')(1-q') = p' + q' - p'q'$$

then

$$p' + q' - p'q' \in [0,1].$$

Thus, there exists real numbers $p'$ and $q'$ such that

$$p' + q' - p'q' = r$$

which implies that

$$(\theta', \phi') \in \bigcup_{\substack{p,q\in[0,1] \\ p+q-pq=r}} \Theta_p \times \Phi_q.$$

Since $(\theta', \phi')$ were chosen arbitrary then,

$$\Psi_r^{\mathrm{OR}} \subseteq \bigcup_{\substack{p,q\in[0,1] \\ p+q-pq=r}} \Theta_p \times \Phi_q.$$

Combining results we have set equality

$$\Psi_r^{\mathrm{OR}} = \bigcup_{\substack{p,q\in[0,1] \\ p+q-pq=r}} \Theta_p \times \Phi_q.$$

Since $r \in [0,1]$ was chosen arbitrarily these sets are equal for every $r \in [0,1]$. $\square$

To form the ROC curve for the fused classification system, we want to maximize $P_{TP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi)$ and minimize $P_{FP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi)$. Consider the constrained optimization problem for every $r \in [0,1]$:

$$
\begin{aligned}
f_{\mathbb{A}\vee\mathbb{B}}(r) &= \max_{(\theta,\phi)\in\Psi_r^{\mathrm{OR}}} P_{TP}(\mathbf{A}_\theta \vee \mathbf{B}_\phi) \\
&= \max_{(\theta,\phi)\in\Psi_r^{\mathrm{OR}}} P_{TP}(\mathbf{A}_\theta) + P_{TP}(\mathbf{B}_\phi) - P_{TP}(\mathbf{A}_\theta)P_{TP}(\mathbf{B}_\phi) \\
&= \max_{(\theta,\phi)\in \bigcup\limits_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \Theta_p\times\Phi_q} P_{TP}(\mathbf{A}_\theta) + P_{TP}(\mathbf{B}_\phi) - P_{TP}(\mathbf{A}_\theta)P_{TP}(\mathbf{B}_\phi) \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \max_{(\theta,\phi)\in\Theta_p\times\Phi_q} P_{TP}(\mathbf{A}_\theta) + P_{TP}(\mathbf{B}_\phi) - P_{TP}(\mathbf{A}_\theta)P_{TP}(\mathbf{B}_\phi) \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \max_{(\theta,\phi)\in\Theta_p\times\Phi_q} [1-[1-P_{TP}(\mathbf{A}_\theta)]\,[1-P_{TP}(\mathbf{B}_\phi)]] \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \left[1 - \min_{\Theta_p\times\Phi_q} [1-P_{TP}(\mathbf{A}_\theta)]\,[1-P_{TP}(\mathbf{B}_\phi)]\right] \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \left[1 - \min_{\theta\in\Theta_p} [1-P_{TP}(\mathbf{A}_\theta)] \min_{\phi\in\Phi_q} [1-P_{TP}(\mathbf{B}_\phi)]\right] \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} \left[1 - [1-\max_{\theta\in\Theta_p} P_{TP}(\mathbf{A}_\theta)][1-\max_{\phi\in\Phi_q} P_{TP}(\mathbf{B}_\phi)]\right] \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} [1-[1-f_{\mathbb{A}}(p)]\,[1-f_{\mathbb{B}}(q)]] \\
&= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} [f_{\mathbb{A}}(p)+f_{\mathbb{B}}(q)-f_{\mathbb{A}}(p)f_{\mathbb{B}}(q)]
\end{aligned}
$$

Note that $q$ is a function of $p$ such that $p+q-pq=r$. Solving for $q$ in terms of $p$ for $r$ fixed yields

$$q = Q(p) = \frac{r-p}{1-p}$$

for $0 \le p \le r$. Therefore, an equivalent formula is

$$
\begin{aligned}
f_{\mathbb{A}\vee\mathbb{B}}(r) &= \max_{\substack{p,q\in[0,1]\\ p+q-pq=r}} [f_{\mathbb{A}}(p)+f_{\mathbb{B}}(q)-f_{\mathbb{A}}(p)f_{\mathbb{B}}(q)] \\
&= \max_{p\in[0,r]} [f_{\mathbb{A}}(p)+f_{\mathbb{B}}(Q(p))-f_{\mathbb{A}}(p)f_{\mathbb{B}}(Q(p))]\,. \qquad (3.3)
\end{aligned}
$$

Now that we have a formula for the join [12], this motivates the creation of a new symbol that represents this operation. Given $f, g \in \mathscr{R}$ we will write

$$[f \sqcup g](r) \equiv \max_{\substack{p,q \in [0,1] \\ p+q-pq=r}} [f(p) + g(q) - f(p)g(q)] \tag{3.4}$$

We read $f \sqcup g$ as "$f$ or $g$". We use the symbol $\sqcup$ rather than $\vee$ in order to distinguish it from dealing with classification systems [9].

Next we will test to see if this operation satisfies the properties of a lattice.

1. (idempotent) Since $\mathbb{A} \vee \mathbb{A} = \mathbb{A}$ it follows that

$$f_{\mathbb{A}} \sqcup f_{\mathbb{A}} = f_{\mathbb{A} \vee \mathbb{A}} = f_{\mathbb{A}}$$

2. (commutativity) Testing for commutativity of the join,

$$f_{\mathbb{A}} \sqcup f_{\mathbb{B}} = f_{\mathbb{A} \vee \mathbb{B}} = f_{\mathbb{B} \vee \mathbb{A}} = f_{\mathbb{B}} \sqcup f_{\mathbb{A}}$$

   By commutativity of the families with regard to the join $\vee$, then the commutativity of the join $\sqcup$ is satisfied.

3. (associativity) Observe that

$$(f_{\mathbb{A}} \sqcup f_{\mathbb{B}}) \sqcup f_{\mathbb{C}} = f_{(\mathbb{A} \vee \mathbb{B}) \vee \mathbb{C}} = f_{\mathbb{A} \vee (\mathbb{B} \vee \mathbb{C})} = f_{\mathbb{A}} \sqcup (f_{\mathbb{B}} \sqcup f_{\mathbb{C}})$$

4. (identity) Define $f_{\mathbb{N}}(p) \equiv 0$ for every $p \in [0, 1]$. Then

$$f_{\mathbb{A}} \sqcup f_{\mathbb{N}} = f_{\mathbb{A}}$$

5. (maximal element) Define $f_{\mathbb{T}}(p) \equiv 1$ for every $p \in [0, 1]$. Then

$$f_{\mathbb{A}} \sqcup f_{\mathbb{T}} = f_{\mathbb{T}}$$

### 3.3 AND Formula

We also will employ the development of Schubert [12] which develops the formula for the AND of ROC curves. Consider the development of the probabilities of true and false positive ($P_{TP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi)$ and $P_{FP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi)$, respectively) for the AND label-fusion rule under the assumption of independence. Recall $\mathcal{L}_t = \{t\}$

$$\begin{aligned} P_{TP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi) &= P([\mathbf{A}_\theta \wedge \mathbf{B}_\phi]^\natural(\mathcal{L}_t)|\mathcal{E}_t) \\ &= P(\mathbf{A}_\theta^\natural(\mathcal{L}_t) \cap \mathbf{B}_\phi^\natural(\mathcal{L}_t)|\mathcal{E}_t) \\ &= P(\mathbf{A}_\theta^\natural(\mathcal{L}_t)|\mathcal{E}_t)P(\mathbf{B}_\phi^\natural(\mathcal{L}_t)|\mathcal{E}_t) \\ &= P_{TP}(\mathbf{A}_\theta)P_{TP}(\mathbf{B}_\phi). \end{aligned} \tag{3.5}$$

$$\begin{aligned} P_{FP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi) &= P([\mathbf{A}_\theta \wedge \mathbf{B}_\phi]^\natural(\mathcal{L}_t)|\mathcal{E}_n) \\ &= P(\mathbf{A}_\theta^\natural(\mathcal{L}_t) \cap \mathbf{B}_\phi^\natural(\mathcal{L}_t)|\mathcal{E}_n) \\ &= P(\mathbf{A}_\theta^\natural(\mathcal{L}_t)|\mathcal{E}_n)P(\mathbf{B}_\phi^\natural(\mathcal{L}_t)|\mathcal{E}_n) \\ &= P_{FP}(\mathbf{A}_\theta)P_{FP}(\mathbf{B}_\phi). \end{aligned} \tag{3.6}$$

Let $p$ be a value for the probability of false positive for classifier $\mathbf{A}_\theta$ then $f_{\mathbb{A}}(p)$ is the value of the probability of true positive for classifier $\mathbf{A}_\theta$. Similarly, let $q$ be a value for the probability of false positive for classifier $\mathbf{B}_\phi$ then $f_{\mathbb{B}}(q)$ is the value of the probability of true positive for classifier $\mathbf{B}_\phi$ . Let $r$ be a value for the probability of false positive for the fused classifier $\mathbf{A}_\theta \wedge \mathbf{B}_\phi$, then $f_{\mathbf{A}_\theta \wedge \mathbf{B}_\phi}(r)$ is the value of the probability of true positive for classifier $\mathbf{A}_\theta \wedge \mathbf{B}_\phi$. Define the sets

$$\begin{aligned} \Theta_p &= \{\theta \in \Theta : P_{FP}(\mathbf{A}_\theta) = p\} \\ \Phi_q &= \{\phi \in \Phi : P_{FP}(\mathbf{B}_\phi) = q\} \\ \Psi_r^{\text{AND}} &= \{(\theta, \phi) \in \Theta \times \Phi : P_{FP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi) = r\}. \end{aligned}$$

**Theorem 3.** *For every $r \in [0,1]$ then*

$$\Psi_r^{AND} = \bigcup_{\substack{p,q \in [0,1] \\ pq = r}} \Theta_p \times \Phi_q.$$

*Proof.* Choose $r \in [0,1]$ and let it be fixed. Let

$$(\theta', \phi') \in \bigcup_{\substack{p,q \in [0,1] \\ pq = r}} \Theta_p \times \Phi_q$$

Then $\theta' \in \Theta_{p'}$ and $\phi' \in \Phi_{q'}$ for some $p', q' \in [0,1]$ such that $p'q' = r$. From the definitions of $\Theta_p$, $\Phi_q$, and $\Psi_r^{\text{AND}}$ we see that

$$\begin{aligned} p' &= P_{FP}(\mathbf{A}_{\theta'}) \\ q' &= P_{FP}(\mathbf{B}_{\phi'}) \\ r &= P_{FP}(\mathbf{A}_{\theta} \wedge \mathbf{B}_{\phi}) \end{aligned}$$

which implies

$$P_{FP}(\mathbf{A}_{\theta'}) P_{FP}(\mathbf{B}_{\phi'}) = P_{FP}(\mathbf{A}_{\theta} \wedge \mathbf{B}_{\phi})$$

and therefore, $(\theta', \phi') \in \Psi_r^{\text{AND}}$. Since $(\theta', \phi')$ were arbitrary and hence, $p', q' \in [0,1]$ then

$$\Psi_r^{\text{AND}} \supseteq \bigcup_{\substack{p,q \in [0,1] \\ pq = r}} \Theta_p \times \Phi_q.$$

Let $(\theta', \phi') \in \Psi_r^{\text{AND}}$ then $P_{FP}(\mathbf{A}_{\theta} \wedge \mathbf{B}_{\phi}) = r$. For some $p', q' \in [0,1]$ such that $\theta' \in \Theta_{p'}$ and $\phi' \in \Phi_{q'}$, we observe that

$$p'q' = P_{FP}(\mathbf{A}_{\theta'}) P_{FP}(\mathbf{B}_{\phi'}) = P_{FP}(\mathbf{A}_{\theta} \wedge \mathbf{B}_{\phi}) = r$$

which implies

$$(\theta', \phi') \in \Theta_{p'} \times \Phi_{q'} \subseteq \bigcup_{\substack{p,q \in [0,1] \\ pq = r}} \Theta_p \times \Phi_q.$$

Hence

$$\Psi_r^{\text{AND}} \subseteq \bigcup_{\substack{p,q\in[0,1] \\ p+q-pq=r}} \Theta_p \times \Phi_q$$

Combining set containments yields set equality. Since $r \in [0,1]$ was chosen arbitrary, then

$$\Psi_r^{\text{AND}} = \bigcup_{\substack{p,q\in[0,1] \\ pq=r}} \Theta_p \times \Phi_q$$

for all $r \in [0,1]$. □

To form the ROC curve for the fused classification system, we want to maximize $P_{TP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi)$ and minimize $P_{FP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi)$. Consider the constrained optimization problem for every $r \in [0,1]$:

$$\begin{aligned}
f_{\mathbb{A}\wedge\mathbb{B}}(r) &= \max_{(\theta,\phi)\in\Psi_r^{\text{AND}}} P_{TP}(\mathbf{A}_\theta \wedge \mathbf{B}_\phi) \\
&= \max_{(\theta,\phi)\in\Psi_r^{\text{AND}}} P_{TP}(\mathbf{A}_\theta) P_{TP}(\mathbf{B}_\phi) \\
&= \max_{(\theta,\phi)\in \bigcup_{\substack{p,q\in[0,1] \\ pq=r}} \Theta_p\times\Phi_q} P_{TP}(\mathbf{A}_\theta) P_{TP}(\mathbf{B}_\phi) \\
&= \max_{\substack{p,q\in[0,1] \\ pq=r}} \left[ \max_{(\theta,\phi)\in\Theta_p\times\Phi_q} P_{TP}(\mathbf{A}_\theta) P_{TP}(\mathbf{B}_\phi) \right] \\
&= \max_{\substack{p,q\in[0,1] \\ pq=r}} \left[ [\max_{\theta\in\Theta_p} P_{TP}(\mathbf{A}_\theta)][\max_{\phi\in\Phi_q} P_{TP}(\mathbf{B}_\phi)] \right] \\
&= \max_{\substack{p,q\in[0,1] \\ pq=r}} f_{\mathbb{A}}(p) f_{\mathbb{B}}(q).
\end{aligned}$$

Next we will test to see if this operation satisfies the properties of a lattice.

1. (idempotent)Since $\mathbb{A} \wedge \mathbb{A} = \mathbb{A}$ it follows that

$$f_{\mathbb{A}} \sqcap f_{\mathbb{A}} = f_{\mathbb{A}\wedge\mathbb{A}} = f_{\mathbb{A}}$$

2. (commutativity) Testing for commutativity of the join,

$$f_{\mathbb{A}} \sqcap f_{\mathbb{B}} = f_{\mathbb{A}\wedge\mathbb{B}} = f_{\mathbb{B}\wedge\mathbb{A}} = f_{\mathbb{B}} \sqcap f_{\mathbb{A}}$$

By commutativity of the reals with regard to multiplication, the commutivity of the meet is satisfied.

3. (associativity) Observe that

$$(f_{\mathbb{A}} \sqcap f_{\mathbb{B}}) \sqcap f_{\mathbb{C}} = f_{(\mathbb{A} \wedge \mathbb{B}) \wedge \mathbb{C}} = f_{\mathbb{A} \wedge (\mathbb{B} \wedge \mathbb{C})} = f_{\mathbb{A}} \sqcap (f_{\mathbb{B}} \sqcap f_{\mathbb{C}})$$

4. (identity)

$$f_{\mathbb{A}} \sqcap f_{\mathbb{T}} = f_{\mathbb{A}}$$

5. (minimal element)

$$f_{\mathbb{A}} \sqcap f_{\mathbb{N}} = f_{\mathbb{N}}$$

### *3.4 Epimorphism*

Our main result is the following theorem [9].

**Theorem 4.** *Let $\mathscr{G} = \{\mathbb{A}^{(1)}, \mathbb{A}^{(2)}, \ldots, \mathbb{A}^{(K)}\}$ be a collection of $K$ families of total classification systems that are mutually independent. Let $(CSF(\mathscr{G}), =, \wedge, \vee)$ denote the Lattice of total, independent classification system families generated by $\mathscr{G}$. Let $\mathscr{F} = \{f_{\mathbb{A}^{(1)}}, f_{\mathbb{A}^{(2)}}, \ldots, f_{\mathbb{A}^{(K)}}\}$ be the collection of $K$ ROC curves corresponding to $\mathscr{G}$. Then $(ROC(\mathscr{F}), =, \sqcap, \sqcup)$ is a Lattice of ROC curves that is epimorphic to $(CSF(\mathscr{G}), =, \wedge, \vee)$.*

*Proof.* Define the mapping

$$F : (CSF(\mathscr{G}), =, \wedge, \vee) \to (ROC(\mathscr{F}), =, \sqcap, \sqcup)$$

to be

$$F(\mathbb{A}) \equiv f_{\mathbb{A}}.$$

Then it is clear that

$$Dom(F) = CSF(\mathscr{G})$$

and,

$$F(\mathbb{A} \vee \mathbb{B}) = F(\mathbb{A}) \sqcup F(\mathbb{B})$$

while,

$$F(\mathbb{A} \wedge \mathbb{B}) = F(\mathbb{A}) \sqcap F(\mathbb{B})$$

then it is clear that

$$Ran(F) = ROC(\mathscr{F}).$$

If $\mathbb{A} \neq \mathbb{B}$ but $f_{\mathbb{A}} = f_{\mathbb{B}}$ then $F$ is not one-one. Thus, the lattices $CSF(\mathscr{G})$ and $ROC(\mathscr{F})$ are epimorphic. $\square$

## IV. Examples

The best way to show the power of label fusion is to incorporate formulas 3.7 and 3.3 for AND and OR into some Matlab code and visually evaluate the results.

Suppose we have three ROC curves from classification systems $\mathbb{A}, \mathbb{B}, \mathbb{C}$; with the curves defined as follows:

$$f_{\mathbb{A}}(p) = \begin{cases} 2p, & 0 \le p \le 0.4 \\ \\ p/3 + 2/3, & 0.4 \le p \le 1.0 \end{cases}$$

$$f_{\mathbb{B}}(p) = \tanh(4p)$$

$$f_{\mathbb{C}}(p) = p^{1/3}$$

The following plots will show how these ROC curves combine, and how the optimal fusion rules result in better performance overall.

### 4.1 Two Classifiers

With only two systems, fusion can either be the AND, or the OR. Which one is better is dependent on the shape of the original ROC curves.

### 4.2 Three Classifiers

Using only AND's and OR's from the Lattice, that is the monotonic combinations, we can see that combining classifiers has benefits which range across the total ROC curve. It should be noted in Figure 4.5 that the majority vote, delivers remarkably good results.

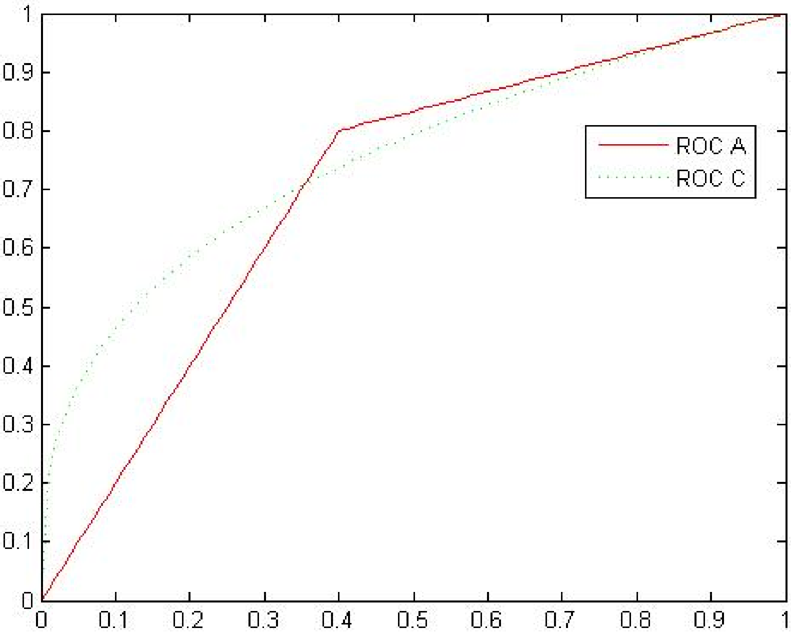


Figure 4.1: ROC Curves of Classification System Families $\mathbb{A}$ and $\mathbb{C}$.

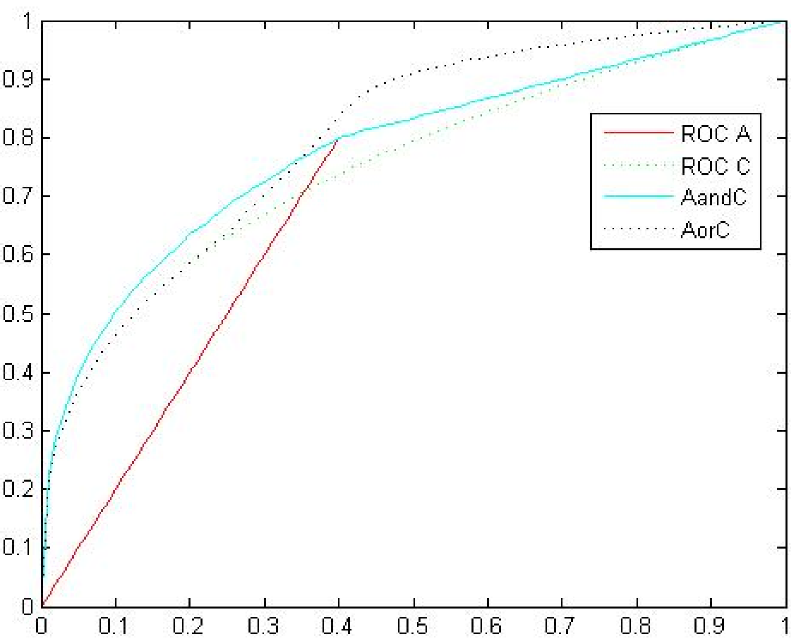


Figure 4.2: ROC Curves of Classification Systems $\mathbb{A}$ and $\mathbb{C}$, Label Fused.

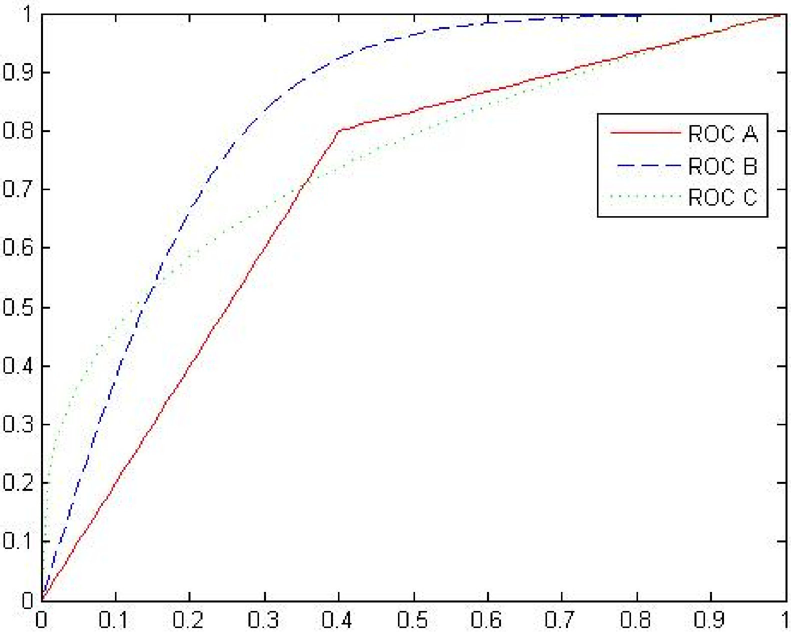


Figure 4.3: ROC Curves of Classification System Families $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$.

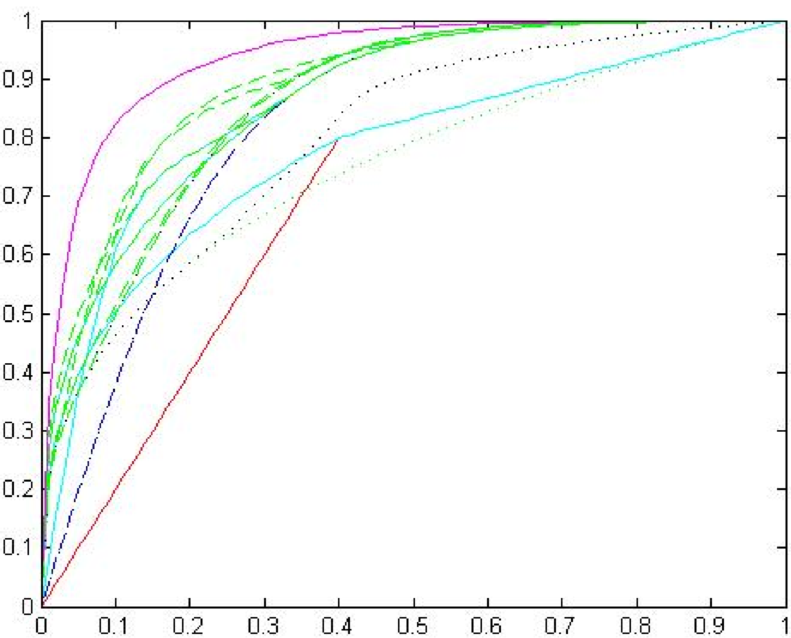


Figure 4.4: ROC Curves of Classification Systems $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$, Label Fused.

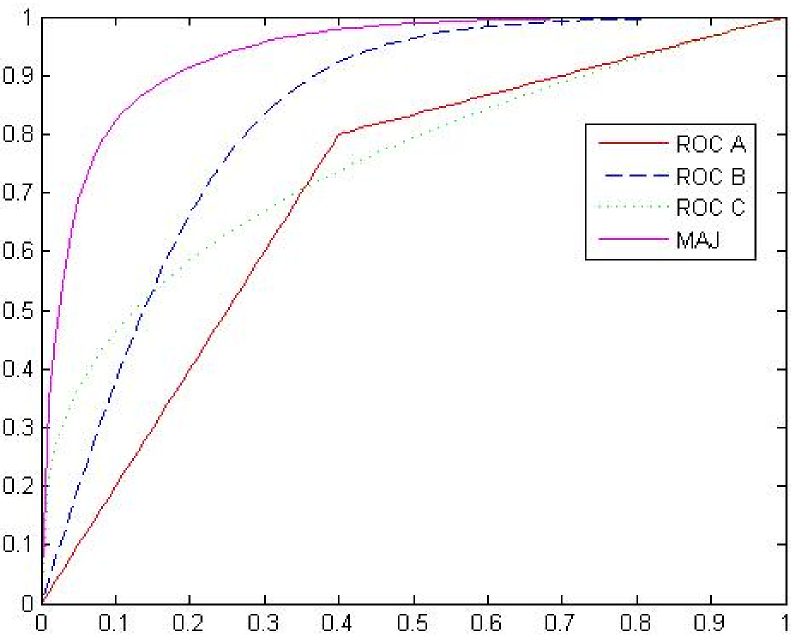


Figure 4.5: ROC Curves of Classification Systems $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$ with Majority Rule.

# *V. Conclusions*

## *5.1 Summary*

We began this thesis discussing the need for optimal target detection arising in many different fields. It was proposed that the combination of multiple classification systems might lead to more optimal target detection performance. We restricted our study to the label fusion of multiple independent two-label classification systems. By doing so we were able to quantify the performance of their combination in the same way as is done for the individual systems; by means of the ROC curve. Provided our assumptions were maintained, we found that the ROC curve of two classification systems which were AND'ed was onto the meet of the individual ROC curves. We found the same result for the ROC of the OR'ed classification systems and the ROC of their join. These results were checked against a list of properties of binary operators to convince us that we were working with lattices, and that the Lattice of Classification System Families is epimorphic to the Lattice of ROC curves. We applied the label fusion techniques to some examples to show visually how increased ROC performance may be achieved by the optimal combination of classification systems. The majority vote was a stellar performer for the systems we chose to test. What we did not show, and possibly cannot show, is that a given ROC curve is one-one with the classification system which produced it. Since a ROC curve is a measure of how well a classification system is performing, it may not be unique, as multiple classification systems might enjoy the same measure of performance. So we held short from claiming that the lattice of ROC curves is isomorphic to the lattice of classification systems. Even though our developments only represented a surjective relationship between classification system families and their ROC curves, it still was and is a noteworthy accomplishment. Given the enormous cost of building and testing combinations of classification systems, and generating their ROC performance, testing the lattice of ROC curves in software with existing individual ROC curves can represent a significant cost savings in the design of optimal classification systems.

### *5.2 Future Work*

It may be possible to show, and it may be worth proving, that the ROC curves of optimal combinations of classification systems are unique, allowing us to assert that there exists an isomorphism between the lattice of ROC curves and the lattice of classification systems. This might prove useful when attempting to reverse engineer what went into each classification system given its ROC performance.

It might also be worthwhile to prove that the lattice of ROC curves is indeed a distributive lattice. This would require that it satisfy the distributive property. We were not able to verify this.

We briefly discussed some different ways of evaluating the ROC performance of a given combination of classification systems. Area Under the Curve (AUC), Neyman-Pearson, and Bayesian Risk were examples. If a functional could be defined that could predict optimal ROC performance without having to form the set of ROC curves generated by the lattice of classification system families, a further cost savings might be achieved by reducing the amount of computation to deliver optimal performance.

*Bibliography*


[1] J. C. Abbott. *Sets, lattices and boolean algebras.* Allyn and Bacon, 1969.

[2] T. Apostol. *Mathematical analysis.* Addison-Wesley, MA, 1974.

[3] J. P. Egan. *Signal detection theory and ROC analysis.* Academic Press, New York NY, 1975.

[4] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent component analysis.* Wiley Interscience, NY, 2001.

[5] N. A. Macmillan and C. D. Creelman. *Detection theory.* Lawrence Erlbaum, 2004.

[6] E. Mendelson. *Boolean algebra and switching circuits.* McGraw-Hill, 1970.

[7] D. C. Montgomery. *Design and analysis of experiments.* John Wiley and Sons, Inc., NJ, Sixth Edition, 2005.

[8] J. Neyman and E. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, 1933.

[9] M. E. Oxley, S. N. Thorsen, and C. M. Schubert. A boolean algebra of receiver operating characteristic curves. In 10th international conference on information fusion, pages paper 1412. 2007.

[10] S. I. Resnick. *A probability path.* Birkhauser, Boston MA, 2005.

[11] A. S. Rosman and M. A. Korsten. Application of summary receiver operating characteristics (SROC) analysis to diagnostic clinical testing. *Advances in Medical Sciences* 52, 2007.

[12] C. M. Schubert, *Quantifying correlation and its effects on system performance in classifier fusion.* Ph.D. Thesis, Wright-Patterson AFB OH, 2005.

[13] C. M. Schubert, M. E. Oxley, and K. W. Bauer, Jr. A comparison of ROC curves for label-fused within and across classifier systems. In 8th international conference on information fusion. 2005.

[14] J. H. Schurmer. Hyperspectral imaging from space. *Air Force Research Laboratories Technology Horizons*, 2003, December.

[15] F. Smarandache and J. Dezert. *Advances and applications of dsmt for information fusion.* American Research Press, Rehoboth NM, 2004.

[16] P. K. Varshney. *Distributed detection and data fusion.* Springer-Verlag, NY, 1997.

[17] T. D. Wickens. *Elementary signal detection theory.* Oxford University Press, NY, 2004.

| **REPORT DOCUMENTATION PAGE** | *Form Approved* *OMB No. 0704-0188* |
|---|---|



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| 08-09-2008 | Master's Thesis | Sept 2006 - Sept 2008 |

**4. TITLE AND SUBTITLE**
The ROC Curves of Fused Independent Classification Systems

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Walsh, Michael B.

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

**8. PERFORMING ORGANIZATION REPORT NUMBER**
AFIT/GAM/ENC/08-06

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Intentionally left blank

**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**

**14. ABSTRACT**
The need for optimal target detection arises in many different fields. Due to the complexity of many targets, it is thought that the combination of multiple classification systems, which can be tuned to several individual target attributes or features, might lead to more optimal target detection performance. The ROC curves of fused independent two-label classification systems may be generated by the mathematical combination of their ROC curves to achieve optimal classifier performance without the need to test every Boolean combination. The monotonic combination of two-label independent classification systems which assign labels to the same target types results in a lattice of ROC curves which are epimorphic to the corresponding combinations of classification systems. Provided the ROC curves of individual systems are available, testing the lattice of ROC curves in software with existing individual ROC curves can represent a significant cost savings in the design of optimal classification systems.

**15. SUBJECT TERMS**
Receiver Operating Characteristic (ROC) Curves, Target Classification, Decision Fusion, Lattice of ROC Curves

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | Mark E. Oxley |
| U | U | U | UU | 48 | **19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636 x4515 |