3-2003

An Investigation of the Effects of Correlation in Sensor Fusion

Susan A. Storm

AN INVESTIGATION OF THE EFFECTS OF CORRELATION IN SENSOR FUSION

THESIS

Susan A. Storm, Captain, USAF

AFIT/GOR/ENS/03-22

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GOR/ENS/03-22

AN INVESTIGATION OF THE EFFECTS OF CORRELATION IN SENSOR FUSION

THESIS

Presented to the Faculty

Department of Operational Sciences

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 Operations Research

Susan A. Storm, BS

Captain, USAF

March 2003

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GOR/ENS/03-22

AN INVESTIGATION OF THE EFFECTS OF CORRELATION IN SENSOR FUSION

Susan A. Storm, BS
Captain, USAF

Approved:

//signed//

Kenneth W. Bauer (Chairman) date

//signed//

Mark E. Oxley (Member) date

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Dr. Kenneth Bauer, for his guidance and support throughout my thesis effort. His insight and experience was appreciated. His ability to think through complex problems and data hang-ups, and his willingness to teach me these methods will carry throughout my Air Force career. I would also like to thank my reader, Dr. Mark Oxley for his insight and support and for always ensuring proper mathematics notation and methods are utilized. I would like to thank Capt Trevor Laine for his support in data generation and Matlab coding.

I am indebted to my classmates for their support during not only the thesis process, but also my whole time at AFIT. Thank you for all of your help. I would like to express my special thanks to my family for their love and support during this process. And finally special thanks goes to my fiancé for his constant love, understanding and support during the thesis process and during our time at AFIT.

Susan A. Storm

Table of Contents

Page

List of Figures

Page

List of Tables

Abstract

This thesis takes the first step towards the creation of a synthetic classifier fusion-testing environment. The effects of data correlation on three classifier fusion techniques were examined. The three fusion methods tested were the ISOC fusion method (Haspert, 2000), the ROC "Within" Fusion method (Oxley and Bauer, 2002) and the simple use of a Probabilistic Neural Network (PNN) as a fusion tool. Test situations were developed to allow the examination of various levels of correlation both between and within feature streams. The effects of training a fusion ensemble on a common dataset versus an independent data set were also contrasted. Some incremental improvements to the ISOC procedure were discovered in this process.

AN INVESTIGATION OF THE EFFECTS OF CORRELATION IN SENSOR FUSION

I. Introduction

1.1 General Issue

During combat, weapons systems operators are tasked by the Air Tasking Order (ATO) to correctly identify hostile forces. After determining that a target is hostile, they are required to debilitate or eliminate this hostile force. During this process these operators rely on sensors in their system to correctly identify these targets. The level of targeting accuracy is dependent on the information gained by the sensors. Combining data with another sensor that is focusing on the same target can enhance the targeting information gained by a specific sensor. The combination of this information is called sensor fusion. Current fusion techniques typically assume that the data received by the targeting sensors is independent. This independence assumption is not always valid. An assumption of independence that is not valid, or correlation that is present, can lead to miscalculations in the sensor fusion procedure and, possibly, the misidentification of targets. The most costly outcome of these miscalculations is fratricide, the killing of friendly forces by friendly fire. Another potential error is the misclassification of hostile targets as friendly, therefore eliminating them as viable targets. The adverse effects of these costly misidentifications suggests that a study of the effects of the independence assumption with regards to the accuracy of fused targeting information would be of great interest.

1.2 Background

Air Force Doctrine specifically sets standards for the "accuracy" of sensor information required for correct target identification. The definition of target identification depends on the designation given by the system user. These designations range from a simple friendly or hostile determination, to a specific determination of a particular target from a particular enemy. This research focuses on the classification of a target as friendly or hostile.

Several sensors are present in each weapons system used for target identification. The readouts from these sensors are fused to make a final identification of a specific target as friendly or hostile. Recent research in target classification and the accuracy of this classification has led to several sensor fusion models. The purpose of these models is to determine if the combinatorial mechanics of fusion need to be updated in the weapons systems. The current fusion model employed by Air Combat Command (ACC) is the Identification System Operating Curve (ISOC) (Haspert, 2000). Another fusion model that is relevant to this research is the Receiver Operating Curve (ROC) fusion model (Oxley and Bauer, 2002). New methods in neural networks suggest that a probabilistic neural net could also be used in data fusion. All of these fusion models assume that the data from each sensor is independent. The data from real world sensors are typically correlated to different degrees, and this correlation leads to problems in identification accuracy.

1.3 Problem Statement

In this thesis the effects of sensor data correlation on fusion models were investigated. This research explores how the degree of correlation in classification data affects the degree of accuracy in a fusion context. In this thesis we consider a two-class problem in which we simplify the sensor target determination to friendly or hostile. The research examines correlation effects across three different fusion techniques. The last step in this research is to present the research findings to ACC/DRSA and the sensor fusion community.

1.4 Research Objectives

The goal of this thesis is to exercise several fusion models, on several techniques, across interesting data sets to assess the outcomes. The fusion models explored are the ISOC fusion model, the ROC "Within" fusion model and a probabilistic neural net (PNN) used as a fusion tool. Due to unavailability of real-world data and for control purposes, we generated artificial data for this study.

1.5 Research Methodology

This thesis employs methodology involving the use of three different fusion models. The ISOC and ROC models both use logical rules to combine given sensor outputs. These rules involve complicated logical "and" and logical "or" rules that determine the best classification accuracy. In the ISOC method these rule combinations are used to form the Identification System Operating Curve. The "within" receiver operating curve (ROC) curve method determines the optimal operating thresholds for each sensor. When applied to the data these thresholds are designed to yield the highest possible true positive rate for a given false positive rate. The probabilistic neural net uses a probability-based Bayes classifier to classify data. All three of these methods are applied to a friend/foe identification problem using toy data.

1.6 Scope of Research

This research is limited by the identification of friendly and hostile forces. The determination of a specific target is not discussed here, but the methods highlighted in this thesis can be used on all types of target identifications. The analysis will be used by ACC/DRSA, AFOSR, and AFRL/SNA on existing data fusion programs and to further their research. This research is a basis for further study into the fusion of correlated data.

1.7 Relevance

The fusion of information for target determination is an area specified in Air Force targeting doctrine. Air Force Doctrine and targeting guidance requires a certain level of information accumulation before engaging a target (AFPAM 14-210, 1998). One way this information accumulation is achieved is through sensor fusion.

A primary mission of the US Air Force is air superiority. Intelligence and targeting information are two tools that the Air Force employs to achieve the air superiority goal. Through this research, the warfighter will gain a better understanding of the ideal way to identify a target as friendly or hostile. Through this effort, the sensor fusion community will expand their knowledge and have a better understanding of how to give the warfighter the best information available on a specific target.

1.8 Outline of Thesis

This thesis is divided into the following five chapters: Introduction, Literature Review, Methodology, Findings and Analysis, and Conclusions. A brief description of each follows.

Chapter 1: Introduction - This chapter discusses the background, focus of research, research objectives, and relevance of this thesis document.

Chapter 2: Literature Review – This chapter begins with the Air Force doctrine and targeting guidance that designates the need for such research. Following this doctrine is a discussion of research that has been accomplished concerning the independence of data during fusion. Finally this chapter discusses the fusion models and tools that are used in this research.

Chapter 3: Methodology – This chapter begins by discussing the two major cases of data generation. These cases include a data containing a single feature stream and data containing multiple feature streams. The correlation introduced into multiple feature stream data is also discussed. Finally this chapter shows the experimental design employed in this thesis research.

Chapter 4: Findings and Analysis – This chapter presents the results of the data fusion when the two major cases of data are modeled. This chapter shows the results of the fusion tools when novel methodology is introduced into the fusion models and a comparison of the results from these models.

Chapter 5: Conclusions and Recommendations – In this final chapter, the research results are reviewed. The relevance of the research effort is shown and recommendations for further research are provided.

II. Literature Review

2.1 Introduction

The purpose of this chapter is to provide a thorough review of literature relevant to this research effort. First, this chapter provides a description of Air Force Doctrine and documentation specific to targeting. Second, this chapter presents an in-depth discussion of the assumption independence in three areas: within the data, within the sensors, and within the fusion model. Additionally, this chapter reviews current multivariate fusion techniques that will be used in the analysis of the data.

2.2 Air Force Guidance

"Every joint air operations plan (JAOP) should include a desired outcome, target set, and a mechanism for achieving the desired outcome." (AFDD2-1, 2000). Proper target identification is one mechanism for achieving a desired outcome. Correctly identifying a target ensures that a weapons system operator has all the necessary information to make an informed decision about the target set. In order to assure that this desired outcome is reached operators utilize precision employment. "Precision employment is the direct application of force that is used to degrade an adversary's capability or will, or the employment of forces to affect an event." (AFDD2-1, 2000). Precision employment includes the application of force and supplies to achieve the desired result along with the required information to make that employment truly precise (AFDD 2-1, 2000). Given a desired outcome, or goal, and precision employment, which include a required information level, an operator has all the tools necessary to effectively engage a target.

"When identifying a target the Air Force uses physical characteristics that are the visually discernable features." (AFPAM 14-210, 1998). "The target shape, size, composition, reflectivity and radiation propagation, determine to a large extent the type and number of weapons, weapon systems, or sensors needed to accomplish the attack or intelligence objective." (AFPAM 14-210, 1998). To properly apply sensor information, the operators need to insure that the information thresholds have been met. This threshold is the point in time when one has accumulated enough information to make a valid decision (AFPAM 14-210, 1998). As Figure 1 suggests, independent information sources, taken by themselves, do not provide enough information to reach this threshold, but when the sources are combined, the threshold is reached (AFPAM 14-210, 1998). It is also important to note that the point of adequacy for information is adjustable depending on the fidelity of information both collected and needed for targeting (AFPAM 14-210, 1998).

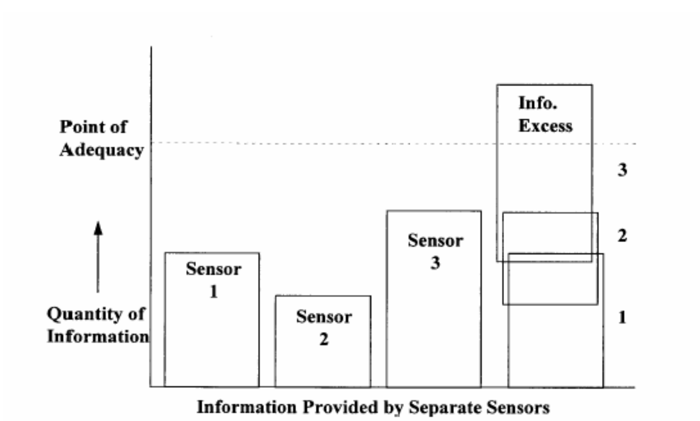

Figure 1: Information Accumulation

Combat identification can be considered the weakest part of the military's kill chain. Links in the chain include searching, detecting, tracking, classifying, identifying, assigning, solution of fire control calculations, weapons launch, mid-course guidance, weapon acquisition of the target, terminal homing, fusing, target damage, and kill assessment (Haspert, 2000). This thesis focuses on the classifying and identification links in this chain through sensor fusion.

The definition of sensor fusion, for the purposes of this thesis, is the combination of the outputs of several disparate ID sensors in a weapons system (Haspert, 2000). In a strict sense, this thesis actually addresses classifier fusion. We assume the sensors have fed their data to classifiers, and it is the classifier outputs that are fused. We use the words "sensors" and "classifiers" synonymously. Traditional sensor fusion uses fixed rules that are easy for operators to implement; however, these rules do not always lead to the optimum target ID (Haspert, 2000). The desired overall effect of this fusion is an improvement in classification accuracy (Shipp and Kuncheva, 2002).

Most fusion techniques assume data and sensor independence. This assumption of independence stems from the conditional probabilities required by most sensor fusion methods (Willett, et al., 2000). The use of conditional probabilities with the assumption of independence leads to more simplified equations and proofs and also leads to fewer calculations required by the user. In terms of a weapons system operator, this means quicker real-time targeting results, which are typically preferred.

2.3 The "good", "bad", and "ugly"

In the sensor fusion process, the goal is to find an optimal set of rules that will give the operator all the information needed for precision employment. The assumption of most fusion models is that the targeting information from the sensors is conditionally independent. This assumption allows the modeler to find a set of logical rules that can be applied to the sensor outputs. These rules will combine the information resulting in the most accurate targeting information available.

In the paper "The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise" Willett, Swazek, and Blum try to find a set of "rules" similar to those of the conditionally independent case and evolve those for the dependent case (Willett, et al., 2000). The focus of this paper was on optimum fusion rules because these are more well understood than the design of optimum sensor rules (Willett, et al., 2000). This thesis research also focuses on the optimum fusion rules used in the ISOC fusion model and the ROC fusion model. When the logical "and", "or", and "xor" fusion rules are divided into three cases of dependent fusion, it was determined that different numerical methods are needed for each problem (Willett, et al., 2000). It was shown that the logical "and" and the logical "or" rules can be analyzed in the same manner (Willett, et al., 2000). Thus, only the logical "and" rule needs to be considered for characterization during sensor fusion (Willett, et al., 2000).

In order to further characterize these rules, the set of all possible Gaussian mean-shift problems was divided into three regions called "good", "bad", and "ugly" (Willett, et al., 2000). Mathematically it can be proven that any problem in the good region must use optimum sensor rules like those used under the assumption of conditional independence (Willett, et al., 2000). For any problem in the bad region, the optimum decision rule could not use single interval decision regions at both sensors (Willett, et al., 2000). The ugly region was dominated by the logical "xor" rule.

In systems using the logical "xor" rule, it can be shown that the usual single-threshold sensor quantization rules can never be optimal; either one sensor must be ignored or several intervals must be considered (Willett, et al., 2000). These regions are complicated due to the fact that these are unconnected (Willett, et al., 2000). These decision regions would require a large number of thresholds for this rule to be useful in the traditional manner and in the dependent case (Willett, et al., 2000).

For the purposes of this document, the rules that are considered here are the logical "and" and the logical "or". As evident from this research, given that a combat identification system's sensors are operating in the "good" region of the threshold spectrum, the same rules that apply under conditional independence can be applied to a dependent case. The question remains, "Will these fusion rules perform adequately given correlated data?"

2.4 Relationships Between Combination and Diversity

In the paper "Relationships between combination methods and measures of diversity in combining classifiers" by Shipp and Kuncheva, the authors discuss the difference between methods of classifier information combination and measures of diversity. Classifier combination is defined as the fusion "rule" that is used to unite data from several sensors. Measures of diversity can be defined as the differences in the resulting data from a sensor. For example, it would not be beneficial to combine two identical data sets because the user would not gain any useful improvement or more information from the combination (Shipp and Kuncheva, 2002). It was found that relationships between different methods of classifier combination and measures of diversity are primarily dependent on this diversity of the data (Shipp and Kuncheva, 2002). If a classification method (or sensor) is not very diverse, the combination methods typically employed, i.e. majority vote, maximum, minimum, average, etc., do not improve notably over a single best classifier (Shipp and Kuncheva, 2002).

The authors also found that there is an interesting correlation between combination methods and diversity measurements (Shipp and Kuncheva, 2002). A diversity measurement such as negative dependence, independence or orthogonality can be overcome depending on the combination method that is employed (Shipp and Kuncheva, 2002). This also means that a diversity measurement can have a completely negative effect on the sensor fusion and cause a loss of information instead of a gain. It was also found that the correlation between combination methods and diversity methods is not consistent. The authors show that each set of diversity measurements can have an optimal classifier combination, but this problem remains open for further research (Shipp and Kuncheva, 2002).

It is typically assumed that the more diverse a set of data is, or the more different types of sensors trained on a particular target, the better the information from the combination of those sensors will be. This is not always the case and is dependent on the types of data analyzed by the sensors and the methods used in the combination of the identification from those sensors. This suggests that for any given set of sensors an optimum fusion rule can be found; yet there is not one optimum fusion rule for any set of sensors (Haspert, 2000).

2.5 Fusion Methods

Three methods of sensor fusion are compared in this thesis. These are the ISOC fusion model, the ROC "Within" fusion model, and a probabilistic neural net. These models are developed very differently, but have the same goal. Specifically, these methods seek to produce a fused classifier that produces the highest true identification of a hostile force, while realizing the smallest possible rate of identifying a friend as hostile.

2.5.1 ISOC Model

The Identification System Operating Curve or ISOC method is a novel algorithm that, given a set of sensors from a weapons system, provides the best fusion rule to determine optimum targeting (Haspert, 2000). The mathematical reasoning in this model is nontrivial, but the resulting technique involves trivial calculations to determine if a set of ID sensor reports will result in a hostile declaration (Haspert, 2000). This methodology requires the user to shift from the current fixed-ID rules of engagement, to adaptive rules. An adaptive rule takes data specific to a particular target and finds the optimum rule for that particular data set. These adaptive rules would require sensor ID probability values as part of the sensor classification report through a sensor performance matrix (Haspert, 2000).

2.5.1.1 Sensor Performance Matrices

Combat Identification Systems (CIS) process data through several sensors and combine the results from theses sensors to form a series of friend/foe identifications (Haspert, 2000). In order to use the ISOC system, the sensor must produce a sensor performance matrix as an output. This performance can be represented in a table format, as seen in Table 1.

Table 1: Sensor Performance Matrix

	a	*b*
T_1	$P(a\|T_1)$	$P(b\|T_1)$
T_2	$P(a\|T_2)$	$P(b\|T_2)$

In this matrix, T_1 and T_2 are the two types of targets, friends or hostiles and *a* and *b* represent two possible ID sensor outputs. The conditional probability represented by $P(a|T_1)$ is the probability of the sensor designating the target as an *a* given that the true target type is T_1. From these sensor performance matrices, the ISOC algorithm consists of a nontrivial algorithm comprised of several trivial calculations.

2.5.1.2 Combat ID System States

Let N_S denote the number of sensors on target. Let *i* be the index for those sensors where $1 \leq i \leq N_S$. Let n_i be the number of indicator states for sensor *i*. Let k_i be the index of states for sensor *i* where $1 \leq k_i \leq n_i$. There will be a total of N distinct configurations of the total system given by

$$N = \prod_{i=1}^{N_s} n_i \ .$$

Let S_j be the j^{th} configuration of a combat identification system (CIS); this is a vector of dimension N_S. Thus $S_j = (s_1^j, s_2^j, \ldots, s_n^j)$ where, for instance, $s_1^j \equiv$ state of the 1^{st} sensor in the j^{th} configuration. Table 2 shows these combinations.

Table 2: Sensor Output State Combinations

j	S_j
1	$(s_1^1, s_2^1, s_3^1, \dots, s_n^1)$
2	$(s_1^2, s_2^2, s_3^2, \dots, s_n^2)$
3	$(s_1^3, s_2^3, s_3^3, \dots, s_n^3)$
:	:
:	:
N	$(s_1^N, s_2^N, s_3^N, \dots, s_n^N)$

Under the assumption that the sensor indications are independent, the probability of a sensor configuration given truth is calculated by multiplying the probabilities of the individual sensors, in a given output state combination, given the same truth. This is shown in the following equation

$$P(S_j \mid T) = \prod_{i=1}^{N_S} P(s_i^j \mid T).$$

In this equation T is defined as the true target type where $T \in \{F, H\}$ and $F \equiv$ target is a friend and $H \equiv$ target is hostile. After all the probabilities have been calculated using all possible output state combinations, the fusion rules must be defined (Ralston, 1998).

2.5.1.3 Fusion Rules

The identification fusion rule must resolve all possible conflicting indications from two or more of the individual sensors, specifically whether or not to declare a target "hostile" and hence engageable for each of the N states of the system (Ralston, 1998). In a case where only two ID designations are used, friend and hostile, a complete ID fusion rule can be expressed as a vector, $R = (r_1, r_2, \ldots, r_N)$ of dimension N where $r_i \in \{0,1\}$, i = 1, 2, …N.

The probability of a specific fusion rule is given by the summation of all the output state combinations multiplied by the given rule set vector. This probability is defined below.

$$P(R \mid T) = \sum_{j=1}^{N} P(S_j \mid T) \cdot r_j$$

so

$$P(R \mid T) = \sum_{j=1}^{N} \left(\prod_{i=1}^{N_S} P(s_i^j \mid T) \right) \cdot r_j$$

The crux of this model is to choose the fusion rule $R(j) = r_j$ that maximizes the probability of declaring a hostile target hostile, while minimizing the probability of declaring a friendly target hostile (Haspert, 2000). The total number of distinct possible rules is 2^N. It is virtually impossible to test all these rules for a large N. A subset of all possible fusion rules that will represent the best performance, for a given sensor suite, can be defined and selected. It is possible to determine how closely an optimum fusion rule may be approached with a given set of sensors using their performance matrices.

At the beginning two fusion rules are immediately obvious, "never declare hostile" and "always declare hostile". Let $R(j) = r_j$, that is R(j) is the j^{th} component of R. The "never declare hostile" rule means that R(j) = 0 for all j and is the most conservative rule (Ralston, 1998). The next most conservative rule is to engage in the single state j for which the likelihood ratio P(j|H)/P(j|F) is largest (Ralston, 1998). The likelihood ratios should always be ordered if there are multiple rules to be considered (Egan, 1975). This gives the maximum true hostile identification rate with the minimum number of friends identified as hostile. The next most conservative rule allows engagement on both this state and also on the system state with the next highest likelihood ratio. By repeating this process, we create successively less conservative rules of engagements until the “always engage” rule, R(j) = 1 for all j, is reached (Ralston, 1998).

From this logic an algorithm to create the ISOC boundary can be implemented. As before let T $\in$ {F, H} where F means the target is a friend and H means the target is Hostile.

1. Compute $P(S_j|T)$ for all j in T. These come from the sensor performance matrices.

2. Compute $P(S_j|H)/P(S_j|F) \equiv LR^j$ where LR^j is the likelihood ratio for a given sensor output state combination.

3. Order the set $\{LR^j \mid j = 1, \ldots, N\}$ from highest to lowest as

$$LR_{[1]}^{j_1} > LR_{[2]}^{j_2} > \ldots > LR_{[N]}^{j_N}$$

4. Pick S_j associated with $LR_{[N]}^{j_N}$ to add to the rule (i.e. make it $r_{j_N} = 1$ in R).

5. Go to 3 unless $r_j = 1$ for all j.

The key to this method is that the N distinct configurations of the CIS are mathematically tested using objective sensor performance data and then “turned on” in decreasing order of likelihood ratio (Ralston, 1998). If a system has N states there will be N+1 points plotted that connect the two obvious rules (Ralston, 1998). Each point provides an alternative trade-off between effectiveness and fratricide. Any point in the set is a rational and objective rule of engagement. No alternative rule can provide higher hostile target identification at the same or lower fratricide rate. No alternative rule can provide lower fratricide at the same or higher defense effectiveness. Although the best trade-off among these alternatives will depend on combat requirements, the collective set of objective fusion rules completely and objectively characterizes the performance of the specific suite of identification sensors being analyzed (Haspert, 2000).

This trajectory of objective fusion rules summarizes the performance of an ID system in the same way that a ROC curve summarizes the detection/false alarm performance of a detection system producing the ID system operating characteristic curve or ISOC (Ralston, 1998). The most optimum of these rules is then determined by cost or other user-defined characteristics (Haspert, 2000).

2.5.1.4 Implementation of the ISOC Method

To implement the ISOC method the classifier outputs are fused. For a two-class problem with three classifiers, there are eight possible output states. If a "Hostile" decision of a particular classifier is denoted with an "H" and the "Friendly" decision with an "F", the eight possible states are listed in Table 3 below.

Table 3: Output States for a Two Class, Three Classifier System

State (C_1, C_2, C_3)	
1. (H, H, H)	5. (H, F, F)
2. (H, H, F)	6. (F, H, H)
3. (H, F, H)	7. (F, F, H)
4. (F, H, F)	8. (F, F, F)

Next (following Section 2.5.1.3) the following likelihood ratios are calculated for all states j:

$$LR = \frac{P\{S_j \mid H\}}{P\{S_j \mid F\}},$$

where $P\{S_j|H\}$ is the likelihood of state j given a hostile target, and $P\{S_j|F\}$ is the likelihood of state j given a friendly target. Then the ratios are ordered from least to greatest such that

$$LR_{[1]} \leq LR_{[2]} \leq \ldots \leq LR_{[8]}.$$

Once the likelihood ratios are ordered, the most likely output state probability is chosen to be a part of a rule set. The second most likely output state probability is then added to this $LR_{[1]}$ to form a second rule. Each rule set consists of three logical "and"s for the output state and up to seven logical "or"s for the fullest rule set. This continues until there are eight possible rules, based on the ordered probabilities. These rules then form the Identification System Operating Characteristic (ISOC). An example of all the possible rule sets and the dominating ISOC curve can be seen in Figure 2. Out of the dominating ISOC curve, an optimum fusion rule can be determined.

The optimal rule is determined by a cost function. The cost is calculated by the following equation

$$C_T = (C_{FN} \times P_H \times p_{FN}) + (C_{FP} \times P_F \times p_{FP})$$

where $C_T \equiv$ total cost of misclassification, $C_{FN} \equiv$ cost of not classifying a hostile as hostile, $P_H \equiv$ *a priori* probability target is hostile, $p_{FN} \equiv$ probability hostile is not declared hostile, $C_{FP} \equiv$ cost of declaring a friend as hostile, $P_F \equiv$ *a priori* probability target is friendly, and $p_{FP} \equiv$ probability friend is declared hostile. The user of the combat identification system sets the costs CFN and CFP. For the purpose of this research these costs are both set to 1. The arrow in Figure 2 points out the optimal ISOC rule for this example.

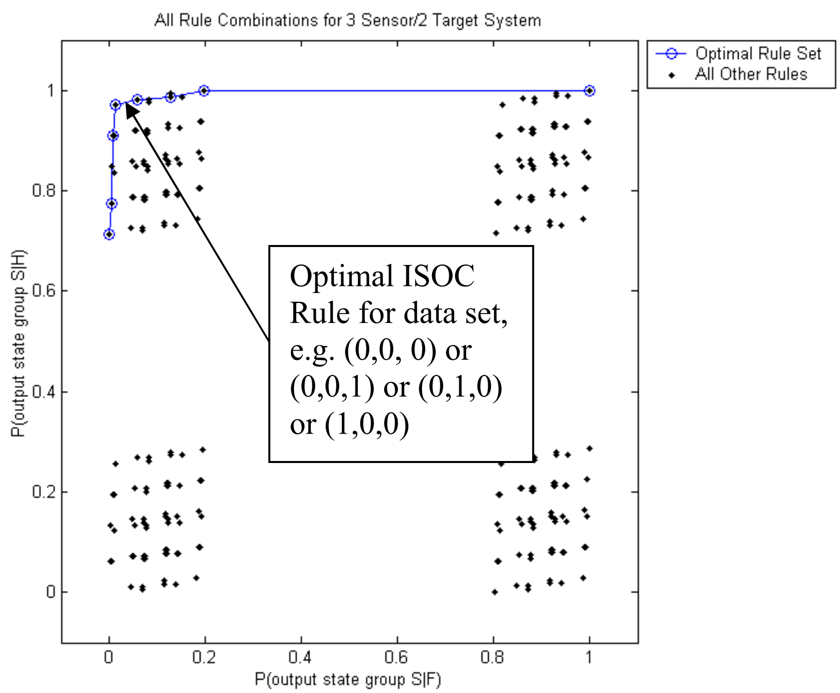

Figure 2: Example of ISOC Rule Sets and Dominating ISOC Curve

2.5.2 ROC Fusion Model

The Receiver Operating Curve (ROC) fusion model combines the results of two or more classifiers into an overall target classification for the combat identification system. Two types of ROC fusion are discussed in the section: across fusion and within fusion. The basic concept of the across ROC fusion model is that two classifiers (sensors) are defined on two different feature sets (X, Y). These feature sets map into two different label sets. These label sets are then combined or fused into a single system label set. In within ROC fusion different sensors are applied to the same feature set (Oxley and Bauer, 2002).

2.5.2.1 Single Classifier

The simplest case of a classification system is a single sensor/classifier. When a single classifier is present, a threshold set $\Theta = [0,1]$ is defined. For each element of that set ($\theta \in \Theta$) there is a classifier A_θ defined to classify the feature set X into a label set L. For a two-class problem that label set could be $L = \{0,1\}$ or any continuum $L = \Re$ (Oxley and Bauer, 2002). This methodology can be seen in Figure 3 below.

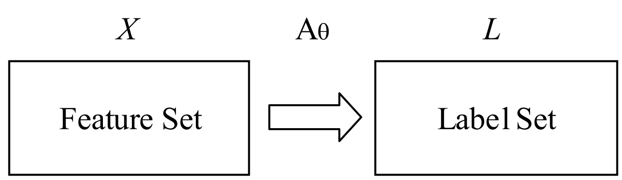

Figure 3: Label Set Methodology – Single Classifier

2.5.2.2 Across ROC Fusion

In a system of two classifiers or sensors, X and Y relate to events occurring in the same event set (Oxley and Bauer, 2002). These produce feature vectors in different feature sets X and Y. These feature sets are mapped into label sets L and M through classifiers A_θ and B_ϕ. For each element of a threshold set, there is a combination of the two classifiers for a concatenated feature set or $C_{\theta,\phi}(x, y) = ((A_\theta(x),B_\phi(y))$. The question of interest is "How can one combine two different classifiers acting on different feature sets to produce results better than the individual classifiers separately?" The answer to this question lies in the probabilities of true positive and the probability of false positives.

These probabilities can be written as sets of conditional probabilities, where each classifier maintains it own label set, i.e., classifier A has a label set L while classifier B has the label set M (Oxley and Bauer, 2002). Figure 4 below shows this process

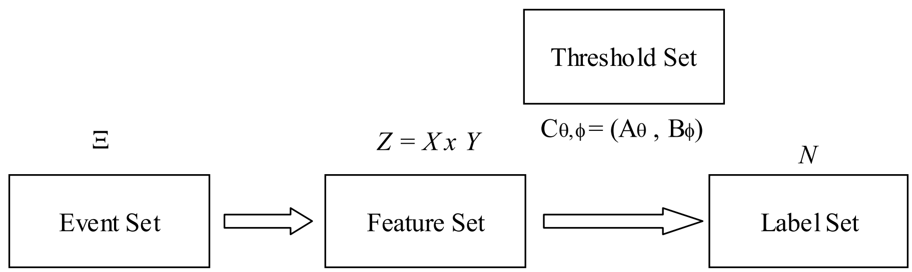

Figure 4: Label Set Methodology – Two Classifiers (Oxley and Bauer, 2002)

2.5.2.3 Fusion Rules

The fusion rule used to combine these two classifiers is the logical "or" rule. These are combined using the following Theorem (Oxley and Bauer, 2002).

$$\begin{gathered}\text{Assuming that the classifiers } A_\theta \text{ and } B_\phi \text{ are independent,}\\ \text{then for every set } Z \in \mathbf{Z} \text{ such that } Z = X \times Y \text{ where } \mathrm{X} \subset X\\ \text{and } \mathrm{Y} \subset Y \text{ then } \Pr(\mathrm{X} \times \mathrm{Y}) = \Pr(\mathrm{X}) \cdot \Pr(\mathrm{Y}).\end{gathered}$$

Using this theorem we can then find the following probabilities for false positives and true positives.

$$P_{FP}(C_{\theta,\phi}) = P_{FP}(A_\theta) + P_{FP}(B_\phi) - P_{FP}(A_\theta) \cdot P_{FP}(B_\phi)$$

Using the *a priori* probabilities of the corresponding classifiers and letting α =

$\Pr(X_{tar})$ and $\beta = \Pr(Y_{tar})$ and adding this to the similar definition of the probability of a true positive we can see the following result. Let $P^A{}_{TP} = P_{TP}(A_\theta)$, $P^A{}_{FP} = P_{FP}(A_\theta)$, $P^B{}_{TP} = P_{TP}(B_\phi)$, and $P^B{}_{FP} = P_{FP}(B_\phi)$ and let $\gamma = \alpha + \beta - \alpha\beta$ to simplify the equation. Then

$$\begin{aligned} P_{TP}(C_{\theta,\phi}) = &\frac{(1-\alpha)\beta}{\gamma} P_{FP}^{A} + \frac{\alpha}{\gamma} P_{TP}^{A} + \frac{\alpha(1-\beta)}{\gamma} P_{FP}^{B} + \frac{\beta}{\gamma} P_{FP}^{B} \\ &- \frac{(1-\alpha)\beta}{\gamma} P_{FP}^{A} P_{TP}^{B} - \frac{\alpha(1-\beta)}{\gamma} P_{TP}^{A} P_{FP}^{B} - \frac{\alpha\beta}{\gamma} P_{TP}^{A} P_{TP}^{B} \end{aligned}$$

These results may be verified from the tables discussed in Section 2.5.2.4.

2.5.2.4 Joint Probabilities

Based on the previous definitions and the statistical definition of conditional probability, assuming independence, the following conditional probability table is produced. Let L_{tar} be defined as the event that classifier A declares a target, L_{non} be defined as the event that classifier A declares a non-target and similarly for classifier B we have the labels M_{tar} and M_{non} (Oxley and Bauer, 2002).

Table 4: ROC Curve Conditional Probability Table for Two Systems and Two Classifiers

		TRUTH			
		$X_{tar} \times Y_{tar}$	$X_{tar} \times Y_{non}$	$X_{non} \times Y_{tar}$	$X_{non} \times Y_{non}$
LABEL	$L_{tar} \times M_{tar}$	$P^A{}_{TP}P^B{}_{TP}$	$P^A{}_{TP}P^B{}_{FP}$	$P^A{}_{FP}P^B{}_{TP}$	$P^A{}_{FP}P^B{}_{FP}$
	$L_{tar} \times M_{non}$	$P^A{}_{TP}P^B{}_{FN}$	$P^A{}_{TP}P^B{}_{TN}$	$P^A{}_{FP}P^B{}_{FN}$	$P^A{}_{FP}P^B{}_{TN}$
	$L_{non} \times M_{tar}$	$P^A{}_{FN}P^B{}_{TP}$	$P^A{}_{FN}P^B{}_{FP}$	$P^A{}_{TN}P^B{}_{TP}$	$P^A{}_{TN}P^B{}_{FP}$
	$L_{non} \times M_{non}$	$P^A{}_{FN}P^B{}_{FN}$	$P^A{}_{FN}P^B{}_{TN}$	$P^A{}_{TN}P^B{}_{FN}$	$P^A{}_{TN}P^B{}_{TN}$

In this table TRUTH is the true target class and LABEL is the classifier label from the feature vector. From this table and adding in the *a priori* probabilities, we finally get the joint probability table seen below.

Table 5: ROC Curve Joint Probability Table for Two Systems and Two Classifiers

		TRUTH			
		$X_{tar} \times Y_{tar}$	$X_{tar} \times Y_{non}$	$X_{non} \times Y_{tar}$	$X_{non} \times Y_{non}$
LABEL	$L_{tar} \times M_{tar}$	$P^A_{TP}P^B_{TP}\alpha\beta$	$P^A_{TP}P^B_{FP}\alpha(1\text{-}\beta)$	$P^A_{FP}P^B_{TP}(1\text{-}\alpha)\beta$	$P^A_{FP}P^B_{FP}(1\text{-}\alpha)(1\text{-}\beta)$
	$L_{tar} \times M_{non}$	$P^A_{TP}P^B_{FN}\alpha\beta$	$P^A_{TP}P^B_{TN}\alpha(1\text{-}\beta)$	$P^A_{FP}P^B_{FN}(1\text{-}\alpha)\beta$	$P^A_{FP}P^B_{TN}(1\text{-}\alpha)(1\text{-}\beta)$
	$L_{non} \times M_{tar}$	$P^A_{FN}P^B_{TP}\alpha\beta$	$P^A_{FN}P^B_{FP}\alpha(1\text{-}\beta)$	$P^A_{TN}P^B_{TP}(1\text{-}\alpha)\beta$	$P^A_{TN}P^B_{FP}(1\text{-}\alpha)(1\text{-}\beta)$
	$L_{non} \times M_{non}$	$P^A_{FN}P^B_{FN}\alpha\beta$	$P^A_{FN}P^B_{TN}\alpha(1\text{-}\beta)$	$P^A_{TN}P^B_{FN}(1\text{-}\alpha)\beta$	$P^A_{TN}P^B_{TN}(1\text{-}\alpha)(1\text{-}\beta)$

An example of how to interpret this table is to consider

$$\Pr(C_{\theta,\phi}{}^{-1}[L_{tar} \times M_{non}] \cap X_{tar} \times Y_{non}) = P^A_{TP}P^B_{TN}\alpha(1-\beta),$$

which is the 2,2 entry in the table. This entry can be read as the probability of classifier A_θ indicating “target” and classifier B_ϕ indication a “non-target”. In the logical “or” case this is where classifier A is looking at features that are due to the established “target” vector, while classifier B responds to the features in the established “non-target” vector.

Again, the above method of ROC curve fusion is called “across” fusion. “Across” fusion combines the results of two classifiers that are looking at two different feature sets in the same event set. This type of ROC fusion can be used for different sensor types that are acting on two targets of different types. It can also be applied to a single target type. For instance, when the sensors in a weapons system combine radar data with thermal data to determine a target the radar sensor and the thermal sensor are looking at different properties of the same target. A different set of conditional and joint probabilities are produced when two sensors are looking at the diverse properties of the same target. These probabilities form the basis of "within" fusion.

2.5.2.5 Within Fusion (adapted to friend/hostile problem)

The method of ROC curve fusion where different sensors use the same feature set is called "within" fusion. An Air Force example would be two different radars tracking a single target. The mathematics behind this method is slightly different, due to the use of only one label set, as such nothing precludes the use of two feature sets with this method.

Let Ξ be an event set. Let X be the set of data vectors whose image is contained in X, the set of feature vectors (Clutz, 2002). Let X_h be the set of system feature vectors indicating a hostile target. Let $p_h = Pr(x \in X_h)$ be the prior probability that a hostile will be indicated. Likewise the definitions associated with friendly targeting are X_f and $p_f = (1 - p_h) = Pr\ (x \in X_f)$. There are only two states of the target (friendly or hostile) in the label set. Two sensors **A** and **B** have associated classifiers A_θ and B_ϕ, where $\theta \in \Theta$ and $\phi \in \Phi$ and Θ and Φ are admissible sets of parameters associated with tuning each classifier. These classifiers assume the data is independent (Clutz, 2002).

$C_{\theta,\phi}$ is used to denote the concatenated classifier of the A_θ and B_ϕ. This classifier returns two labels l_1 and l_2. A rule, R, transforms these two labels into a single label. Such as $R\ (l_1, l_2) = l_1 \vee l_2$ where the $\vee$ operator is defined as the "logical or" rule and L is the label set (Clutz, 2002). $D_{\theta,\phi}$ is used to denote the fused classifier, $D_{\theta,\phi} = A_\theta(x) \vee B_\phi(x)$ (Clutz, 2002). As in all the other methods of fusion and the definition of true positive and false positive are the same. This method uses the same notation as before,

$$P^A_{TP} = \Pr(A_\theta(x) \in L_h \mid x \in X_h)$$
$$P^A_{FP} = \Pr(A_\theta(x) \in L_h \mid x \in X_f)$$
$$P^A_{TN} = \Pr(A_\theta(x) \in L_f \mid x \in X_f)$$
$$P^A_{FN} = \Pr(A_\theta(x) \in L_f \mid x \in X_h).$$

The definitions for B are similar (Clutz, 2002). A conditional probability table similar to the "across" fusion table is shown below.

Table 6: ROC Curve Conditional Probability Table for One System and Two Classifiers

		Classifier Report $C_{\theta,\phi} = (A_\theta, B_\phi)$			
		H, H	**H, F**	**F, H**	**F, F**
True State	**Friend**	$P^A_{FP}P^B_{FP}$	$P^A_{FP}P^B_{TN}$	$P^A_{TN}P^B_{FP}$	$P^A_{TN}P^B_{TN}$
	Hostile	$P^A_{TP}P^B_{TP}$	$P^A_{TP}P^B_{FN}$	$P^A_{FN}P^B_{TP}$	$P^A_{FN}P^B_{FN}$

From this conditional table, it can be seen how the following joint probability table lists the possible outcomes as disjoint events. The general formulation is

$\Pr(C_{\theta,\phi}(x) \in (L_i \times L_j) \cap (x \in X_k))$

$= \Pr((A_\theta(x), B_\phi(x)) \in (L_i \times L_j) \mid (x \in X_k))\Pr(x \in X_k)$

$= \Pr(A_\theta(x) \in L_i \mid (x \in X_k) \Pr (B_\phi(x)) \in L_j \mid (x \in X_k) \Pr(x \in X_k)$

where $i, j, k \in \{h,f\}$.

Table 7: ROC Curve Joint Probability Table for One System and Two Classifiers

		Classifier Report $C_{\theta,\phi} = (A_\theta, B_\phi)$			
		H, H	**H, F**	**F, H**	**F, F**
True State	**Friend**	$P^A_{FP}P^B_{FP}\, p_f$	$P^A_{FP}P^B_{TN}\, p_f$	$P^A_{TN}P^B_{FP}\, p_f$	$P^A_{TN}P^B_{TN}\, p_f$
	Hostile	$P^A_{TP}P^B_{TP}\, p_h$	$P^A_{TP}P^B_{FN}\, p_h$	$P^A_{FN}P^B_{TP}\, p_h$	$P^A_{FN}P^B_{FN}\, p_h$

The above table shows the probability of occurrence for each possible event as a product of the individual probabilities. These are mutually exclusive and collectively exhaustive. The designation "H" means the classifier has reported a hostile, "F" means a friend. The designation "H,H" means that 'classifier A_θ reports hostile, classifier B_ϕ reports hostile'. ROC curves for each classifier consist of a set where a probability of true positive value (ordinate) is specified for each probability of false positive value (abscissa) (Clutz, 2002). The "within" fusion method uses these coordinate pairs, at common set points along the abscissa, to create the new ROC curve (Clutz, 2002).

This methodology was developed using $(P^A{}_{FP}, P^A{}_{TP})$ and $(P^B{}_{FP}, P^B{}_{TP})$ as data pairs. The result point will be $(P^C{}_{FP}, P^C{}_{TP})$ (Clutz, 2002). The probability of false positive for $C_{\theta,\phi}$ is the probability that $C_{\theta,\phi}$ declares a hostile; given the target is a friend. The classifier will declare a hostile in three cases, using the "logical or" rule. Note that

$$P^C{}_{FP} = 1 - P^C{}_{TN}.$$

Using Bayes rule we can see that

$$P^D{}_{TN} = \Pr(D_{\theta,\phi}(x) \in L_f \mid x \in X_f)$$

$$= \Pr((A_\theta(x) \vee B_\phi(x)) \in L_f \mid x \in X_f)$$

$$P^C{}_{TN} = \Pr((A_\theta(x) \in L_f) \cap (B_\phi(x) \in L_f) \mid x \in X_f).$$

Using the law of conditional probability

$$= [P^A{}_{TN}][P^B{}_{TN}].$$

Thus, the point on the fused ROC curve is given by (Clutz, 2002)

$$(P^C{}_{FP}, P^C{}_{TP}) = (P^A{}_{FP} + P^B{}_{FP} - P^A{}_{FP}P^B{}_{FP}, P^A{}_{TP} + P^B{}_{TP} - P^A{}_{TP}P^B{}_{TP}).$$

As in the "across" fusion method, the "within" method assumes independence. A different rule could be developed without independence that assumes operating points are set *a priori*. The within fusion rule provides an upper bound for the fused ROC curve C. This rule allows for the combination of any number of classifiers. This is accomplished by fusing 2 classifiers, then fusing the resulting curve with another. This becomes an iterative process and continues until all classifiers are fused (Clutz, 2002).

2.5.3 Probabilistic Neural Network (PNN) Fusion Model

The probabilistic neural network fusion method entails simply training a PNN to learn the simultaneous outputs of two classifiers and thereby fuse these two classifiers. The PNN has been used successfully to solve many diverse classification problems (Wasserman and Nostrand, 1993.) Compared with a standard back-propagation algorithm, the PNN offers the following advantages: rapid training; convergence to a Bayes Optimal Classifier; addition or deletion of data from the training set without retraining; and confidence assessment for its outputs (Wasserman and Nostrand, 1993).

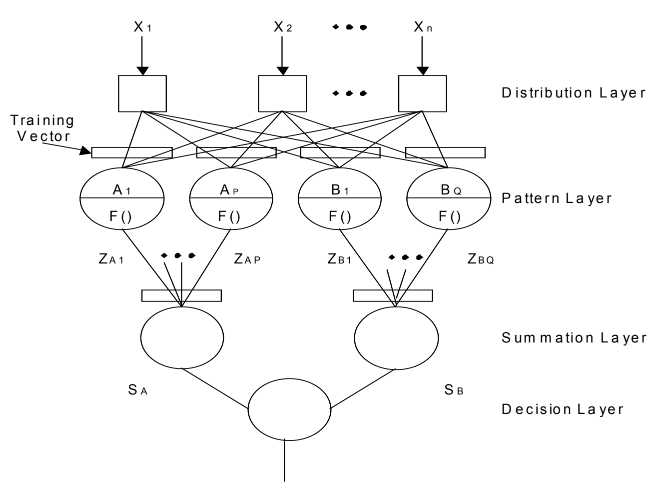

Figure 5: A Probabilistic Neural Network (Wasserman and Nostrand, 1993)

A two-class PNN network is shown Figure 5. An input vector $\mathbf{X} = (x_1\ x_2\ \ldots\ x_n)$ is applied to the neurons of a distribution layer. This vector is to be classified by the neural network. The distribution layer of this network serves as a connection point and the neurons do not perform any computations (Wasserman and Nostrand, 1993). A specific training vector is using to calculate a set of weights, where each weight has the value of a component of that vector. The pattern layer neurons are grouped by the known classification of the associated training vector and each of these neurons sums the weighted inputs from the distribution layer neurons. After summations, the pattern layer neuron applies the non-linear function $f(\cdot)$ to that sum producing output Z_{ci}. In this output c indicates the class of the associated training vector while i indicates the pattern layer neuron computing that class (Wasserman and Nostrand, 1993). The exponential function for Z_{ci} is

$$Z_{ci} = \exp\left[\left(\frac{\mathbf{X}_{Ri}^T \mathbf{X}_i - 1}{\sigma^2}\right)\right]$$

where the input vector $\mathbf{X} = (x_1, x_2, \ldots x_n)$ and the set of weights associated with a given pattern neuron represent a training vector $\mathbf{X}_{Ri} = (x_{R1}, x_{R2} \ldots, x_{Rn})$.

Each neuron in the summation layer receives all patter layer outputs for a given class. The equation for the summation of a specific class, S_c is

$$S_c = \sum_{i=1} \exp\left[\frac{\left(\mathbf{X}^T \mathbf{X}_{Ri} - 1\right)}{\sigma^2}\right]$$

In the decision layer, each neuron forms a comparison based on the decision rule

$$D(\mathbf{X}) = \theta_R \text{ if}$$
$$\sum_{i=1}^{n_R} \exp\left[\frac{\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}_{Ri} - 1\right)}{\sigma^2}\right] \geq \sum_{i=1}^{n_S} \exp\left[\frac{\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}_{Ri} - 1\right)}{\sigma^2}\right].$$

In this comparison the neuron outputs a one if S_a is greater than S_b and zero otherwise (Wasserman and Nostrand, 1993). This output indicates the class of the current input vector. A probabilistic neural net can be easily extended to an arbitrary number of classes by adding pattern layer neurons and summation layer neurons for each class (Wasserman and Nostrand, 1993).

2.6 Chapter Summary

Several topics were discussed in this chapter. It was shown that Air Force Doctrine and targeting guidance requires a specified level of information accumulation. This level of accumulation can be achieved through fusing the information from several sensors. The level of information accumulation that is required is dependent on the specific target, but it can be assumed that this information cannot be collected safely through one information source alone. Several sources are required, which leads to sensor fusion. Due to the complexity of sensor fusion, several models have been developed and assumptions made in those models must be closely inspected.

First the latest research in the independence of fusion rules and their dependence on data diversity was discussed. Next the different fusion models we chose were reviewed. The three models that we chose were the ISOC fusion model, the ROC "Within" fusion model, and a probabilistic neural net as a fusion tool.

III. Methodology

3.1 Introduction

This chapter discusses the methodology employed in this research. First, this chapter shows the data generation process. Data was generated for this research due to lack of real-world data and for correlation control purposes. Two major cases of data were applied to the fusion tools discussed in Section 2.5. Finally, the different types of correlation are discussed. The results of these methods can be seen in Chapter IV.

3.2 Data Generation – 2 Major Cases

Two cases were considered in this research: a single feature set and multiple feature sets. In all cases the outputs of two or more classifiers were fused using the three separate fusion techniques discussed in Section 2.5.

3.2.1 Single Feature Set

In the first case of generation, data is developed with a single feature set. This feature set has four features and was generated using a U(0,1) distribution. The following non-linear mapping function *f(x)* was developed to incorporate the four features.

$$f(x) = x_1 + \pi x_1 x_2 - \frac{1+\sqrt{5}}{2x_2} + \sqrt{x_3} + \frac{x_1}{x_4}$$

where each x_i is uniformly distributed [0,1] with an expected value of 0.5. When the expected value of the U(0,1) is input into this function, *f(x) = -0.2436*. If the result of this function is greater than the mean value, then feature vector **X** is labeled class 0.

Otherwise the target is said to be in class 1. Six independent data sets of 100 exemplars were generated using this method. One of these sets was used as a validation set. The other four data sets were used in different realizations as explained below.

3.2.1.1 One Realization

In this case the classifiers were trained with one realization of a single feature set F_1. This can be seen in Figure 6 below. The three classifiers that were used were C_1- linear discriminants, C_2 – quadratic discriminants, and C_3 – a probabilistic neural net (PNN). Once these classifiers were trained, a separate validation set was applied to the results, and the posterior probabilities from this validation set were fused using the ISOC fusion method, the ROC "Within" Fusion, and the PNN fusion method.

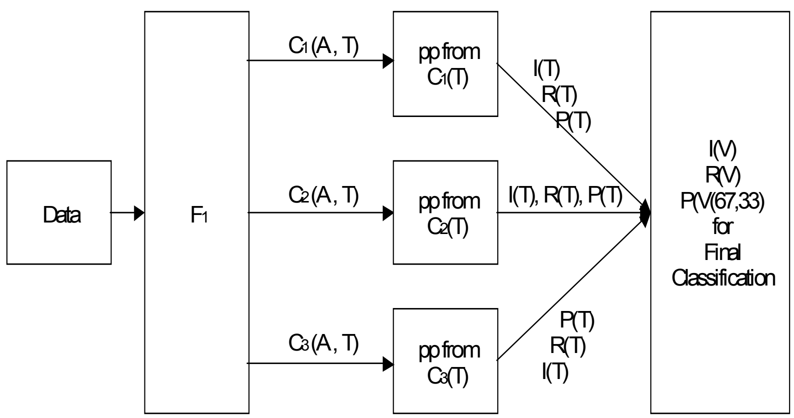

Figure 6: One Realization Flowchart

In Figure 6 the following variables were used: $F_1 \equiv$ Feature Set 1 with four features, $A \equiv 100$ training exemplars with four features, $T \equiv 100$ test exemplars with four features, $V \equiv 100$ validation exemplars with four features, $I \equiv$ ISOC Fusion Application, R ≡ ROC Fusion Application, and P ≡ PNN Fusion Application. The symbol $C_1(A,T)$ signifies that classifier 1, linear discriminant analysis, was trained on data set A and tested on data set T. The symbol I(V) shows that the posterior probabilities from the validation set V were fused using the optimal ISOC rule. The symbol P(V(67,33)) defines that in the PNN fusion, 67 posterior probability exemplars from the validation set were used for training the neural net, while 33 exemplars were used for application of the PNN. A single realization is one way to utilize this data set. Another utilization technique involves multiple realizations of one feature set.

3.2.1.2 Multiple Realizations

When using multiple realization of a feature set, the classifiers were trained using three independent realizations of the data set. This method can be seen in Figure 7 below. The three classifiers that were used were C_1- linear discriminant analysis, C_2 – quadratic discriminant analysis, and C_3 – a probabilistic neural net (PNN). Once these classifiers were trained, the validation set was applied to the results, and the posterior probabilities from this validation set were fused using the ISOC fusion method, the ROC "Within" Fusion, and the PNN fusion method.

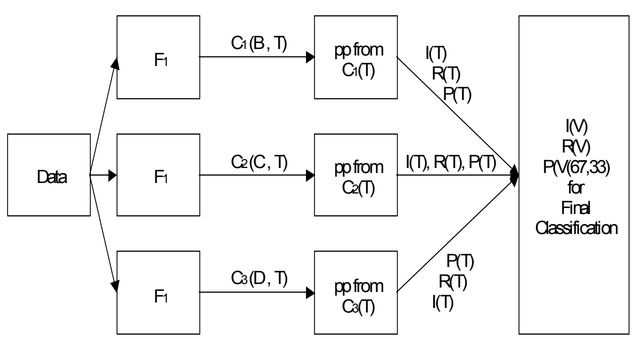

Figure 7: Multiple Realization Flowchart

In Figure 7 the variables previously defined are the same. The following variables were added: B ≡ 100 training exemplars with four features, C ≡ 100 training exemplars with four features, and D ≡ 100 training exemplars with four features. The symbol $C_1(B,T)$ signifies that C_1 was a linear discriminant classifier that was trained on data set B and tested on data set T. The same data sets T and V were used with both methods. The symbol R(V) shows that the posterior probabilities from validation set V were fused using the optimal ROC thresholds from the "within" fusion rule. Single and multiple realization of a data set are one way to test sensor fusion models. Another way to test these models is using multiple feature sets.

3.2.2 Multiple Feature Sets

Unlike the previous case, multiple feature sets were generated for the next step. This was done to incorporate different levels of correlation between features. This data set was designed to correlate features during the fusion process, but not to affect the individual classification efforts. There are two main types of correlation when working with multiple feature sets. These are "intra-correlation" and "inter-correlation". Intra- and inter-correlation can also be categorized as within and across data streams.

3.2.2.1 Intra- and Inter-Correlation

"Within" data correlation is a term used relative to a specific data stream when multiple feature sets are present. There are two types of within data correlation. These are intra-correlation and inter-correlation. Intra-correlation refers to the autocorrelation of a specific data stream. This process is given notionally in Figure 8 below.

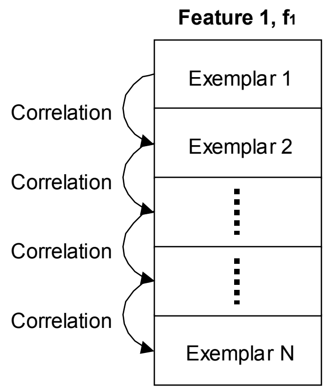

Figure 8: Intra-Correlation of One Feature

The second type of within data correlation is inter-correlation. Inter-correlation is the correlation between features in a given set. This type of correlation is shown notionally in Figure 9 below. This type of correlation is also called "across" correlation.

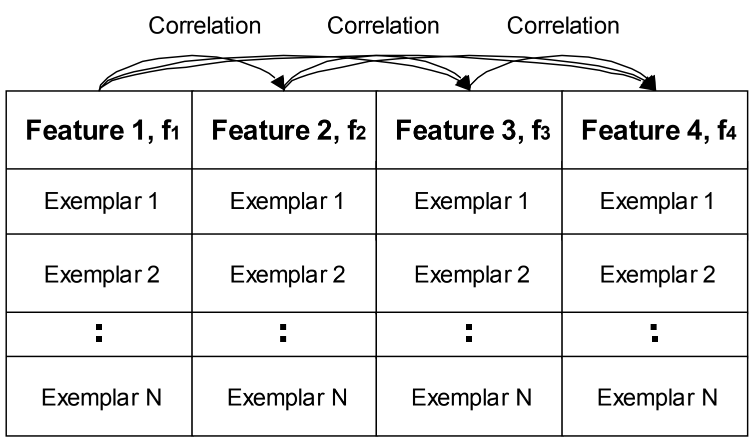

Figure 9: Inter-Correlation of Multiple Features

Within correlation of the intra-correlation type is not considered in this research.

3.2.2.2 Setup

From this point forward "within" correlation refers to intra-correlation and "across" correlation refers to inter-correlation. A second set of data was generated to test inter-correlation of the data across two data sets. Let $F = F_1 \times F_2 \subset \Re^4$ where F_1 is feature set 1 and F_2 is feature set 2. Assume the correlation of the data is given by

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{F_1,F_2} \\ \Sigma_{F_1,F_2} & \Sigma_{2,2} \end{bmatrix}$$

where

$$\Sigma_{1,1} = \Sigma_{2,2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \Sigma_{F_1,F_2} = \begin{bmatrix} 0 & \rho \\ \rho & 0 \end{bmatrix}$$

where $\rho \in \{0, 1/n, \ldots 4/n\}$ and n = 5 and $\Sigma_{1,1}$ is the correlation matrix between the features contained in the feature set F_1 and class 1. If $F_{i,j}$ designates feature set i for class j, let $F_1 = F_{1,1} \cup F_{1,2}$ where

$$F_{1,1} \sim N_2(\mu_{1,1}, \Sigma_{1,1}) \text{ and } F_{1,2} \sim N_2(\mu_{1,2}, \Sigma_{1,2})$$

and where

$$\mu_{1,1} = (0,0)^T \text{ and } \mu_{1,2} = (0.95, 0.95)^T.$$

Let $F_2 = F_{2,1} \cup F_{2,2}$ where

$$F_{2,1} \sim N_2(\mu_{2,1}, \Sigma_{2,1}) \text{ and } F_{2,2} \sim N_2(\mu_{2,2}, \Sigma_{2,2})$$

and where

$$\mu_{2,1} = (0,0)^T \text{ and } \mu_{2,2} = (1.15,1.15)^T.$$

In this case the inter-correlation between the features in a specific set is zero. After the data was generated, it was analyzed in a similar manner as the single feature set data. This can be seen in Figure 10 below.

When there are multiple feature sets, each classifier is trained and applied to a different feature set. In this case the classifier 1 was trained and tested with realizations of F_1 (A_1 and T_1) while classifier 2 was trained and tested with realizations of F_2 (A_2 and T_2) as shown in Figure 10 below. The two classifiers that were used were C_1- linear discriminant analysis and C_2 – quadratic discriminant analysis. Once these classifiers were trained, the validation set was applied to the results, and the posterior probabilities from this validation set were fused using the ISOC fusion method, the ROC "Within" Fusion, and the PNN fusion method.

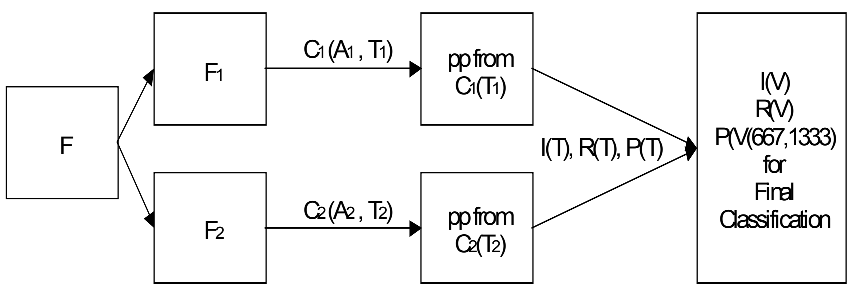

Figure 10: Multiple Feature Set Flowchart.

In Figure 10 the following variables were used: $F_1 \equiv$ Feature Set 1 with two features, $F_2 \equiv$ Feature Set 2 with two features, A ≡ 2000 training exemplars with two features (f_1, f_2 in F_1 and f_3, f_4 in F_2), T ≡ 2000 test exemplars with two features, V ≡ 2000 validation exemplars with two features, I ≡ ISOC Fusion Application, R ≡ ROC Fusion Application, and P ≡ PNN Fusion Application. The symbol $C_1(A,T)$ signifies that classifier 1, linear discriminant analysis, was trained on data set A and tested on data set T. The symbol I(V) shows that the posterior probabilities from the validation set V were fused using the optimal ISOC rule. The symbol P(V(667,1333)) defines that in the PNN fusion, 667 posterior probability exemplars from the validation set were used for training the neural net, while 1333 exemplars were used for application of the PNN. Once this data was generated, an experiment was designed to test the fusion models against correlation.

3.3 Experimental Design

The experiment in this thesis was designed to study the three fusion models; ISOC, ROC and PNN, with both a single feature set and multiple feature sets. When multiple feature sets are present, additional tests were run to determine the effect of correlation on the fusion models. The variable designations from Section 3.2 still apply to the following explanation of our experimental design.

3.3.1 ISOC Application

The ISOC fusion model is designed to find an optimal rule for a given data set. In this research, the classifiers were trained with one data set and tested on another set. The posterior probabilities from this classifier were then used to determine an optimal ISOC rule as outline in Section 2.5.1. The optimal ISOC rule was then applied to the validation data set. The methodology used with the ISOC fusion model is shown in Figure 11.

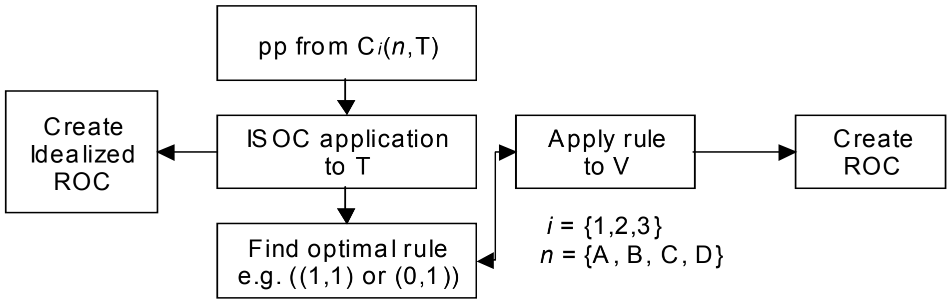

Figure 11: Application of ISOC Fusion Model

The above methodology was applied to a single feature set data using both one realization of the data set and multiple realizations of the data set, as explained in Section 3.2.1. This methodology was also applied to multiple feature sets when inter-correlation of the data was present. Six levels of inter-correlation were tested where $\rho \in \{0, 1/n, 2/n, 3/n, 4/n, 9/2n\}$ and n=5. Varying the three individual classifier thresholds simultaneously from 0 to 1 as shown below created the ROC curves

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} t \\ t \\ t \end{pmatrix} \text{ where } t \in \{0, 0.1, \ldots 1\}$$

and where $t_1 = C_1$ classification threshold, $t_2 = C_2$ classification threshold, and $t_3 = C_3$ classification threshold. These results were plotted against one another to determine the ISOC models robustness in the face of inter-correlation. After the ISOC model was applied, the next step of the experiment was to apply the ROC fusion model.

3.3.2 ROC Application

The ROC fusion model is designed to find the optimal thresholds needed in the individual classifiers to maintain optimal fusion performance. When using ROC fusion, the classifiers were trained with one data set and tested on another set. The posterior probabilities from these classifiers were then fused using the ROC "within" method as outlined in Section 2.5.2.5. After the classifier results were fused, the optimal thresholds for each classifier were found.

To find these thresholds, once the "within" fusion was complete; a given false positive value r* was chosen. The $f_C(r^*)$ is the true positive value for a particular r*, read from the ROC curve f_C. The threshold values we seek are the θ* and ϕ* such that

$$P_{FP}(C_{\theta^*,\phi^*}) = r^* \text{ and } P_{TP}(C_{\theta^*,\phi^*}) = f_C(r^*).$$

Let p^* be the value such that $f_A(p) + f_B(Q(p)) - f_A(p)f_B(Q(p))$ is maximized on [0, r*] where f_A is the ROC curve from classifier 1 and f_B is the ROC curve from classifier 2. Since $Q(p) = (r-p)/(1-p)$ then let $q^*=Q(p^*)$ then $p^*+q^*-p^*q^*=r^*$. Thus, we chose θ* such that

$$P_{FP}(A_{\theta^*}) = p^* \text{ and } P_{TP}(A_{\theta^*}) = f_A(p^*)$$

and we chose ϕ* such that

$$P_{FP}(B_{\phi^*}) = q^* \text{ and } P_{TP}(B_{\phi^*}) = f_B(q^*).$$

A threshold for each classifier was found for fused ROC curve false positive values of r* = {0, .1, .2, …1}. The posterior probabilities from the validation set were then classified using these thresholds. Next the logical "or" rule was used to fuse the classified results. Finally these results were plotted. The methodology used with the ROC fusion model is shown in Figure 12 below.

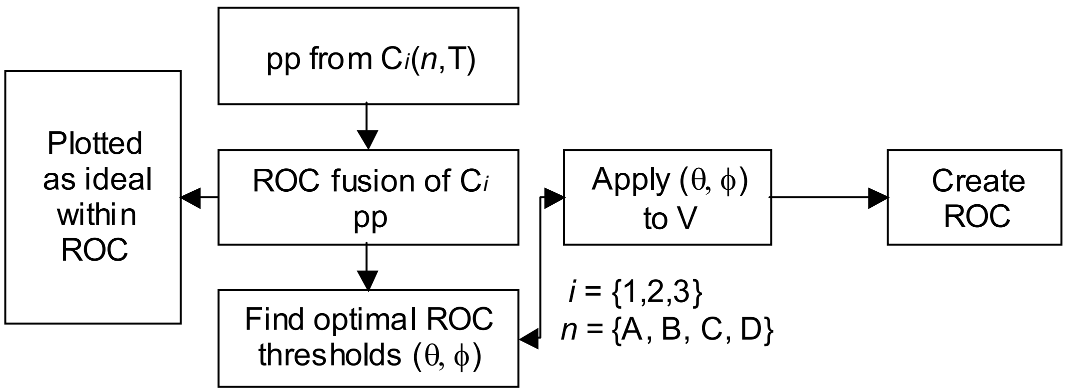

Figure 12: Application of the ROC Fusion Model

As was done with the ISOC fusion model, the above methodology was applied to a single feature set data using both one realization of the data set and multiple realizations of the data set, as explained in Section 3.2.1. This methodology was also applied to multiple feature sets when inter-correlation of the data was present. Six levels of inter-correlation were tested where $\rho \in \{0, 1/n, 2/n, 3/n, 4/n, 9/2n\}$ and n=5. These results were plotted against one another to determine the ROC model's robustness in regards to inter-correlation. After the application of the ROC fusion model, the final step in this experiment was to apply a PNN to the posterior probabilities.

3.3.3 PNN Application

A PNN is designed to classify a target based on a given data set. The PNN fusion model takes the posterior probabilities from two classifiers and uses those probabilities as features. The PNN is then trained on 1/3 of the data points from this posterior probability set and then applied to 2/3 of the validation posterior probabilities. The PNN was employed as shown in Section 2.5.3 of this document. This application methodology can be seen in Figure 13.

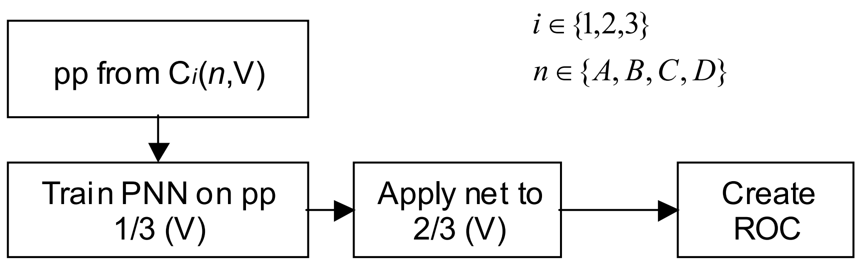

Figure 13: Application of the PNN Fusion Model

As was done with the ISOC and ROC fusion models, the above methodology was applied to a single feature set data using both one realization of the data set and multiple realizations of the data set, as explained in Section 3.2.1. This methodology was also applied to multiple feature sets when inter-correlation of the data was present. Six levels of inter-correlation were tested where $\rho \in \{0, 1/n, 2/n, 3/n, 4/n, 9/2n\}$ and $n=5$. These results were plotted against one another to determine the PNN model's robustness in the face of inter-correlation.

3.4 Conclusion

This chapter discussed the methodology employed in our research effort. The method of data generation was discussed, in both the case of a single feature set and multiple feature sets. The difference between intra-correlation and inter-correlation was explained, and the application of inter-correlation to a data set was exemplified. Next the design of the experiment in both feature set cases was demonstrated. Chapter IV will discuss the results of these experiments.

IV. Findings and Analysis

4.1 Introduction

In this chapter the effects of correlation on three fusion schemes are displayed. Three major cases are discussed. First, the results of single feature set data with one realization are shown. Next, the results of a single feature set with multiple realizations are shown. Finally the results of data with two feature sets at various inter-correlation levels are shown. All individual classifier ROCs are given in the appendix.

4.2 Single Feature Set, One Realization

The results of the simulated data where there was a single feature set and one realization are shown in the figures below. For this analysis, data is developed with a single feature set. This feature set has four features and was generated using a U(0,1) distribution in each feature. The following non-linear mapping function *f(x)* was developed to incorporate the four features.

$$f(x) = x_1 + \pi x_1 x_2 - \frac{1+\sqrt{5}}{2x_2} + \sqrt{x_3} + \frac{x_1}{x_4}$$

where each x_i is uniformly distributed [0,1] with an expected value of 0.5. When the expected value of the U(0,1) is input into this function, $f(x) = -0.2436$. If the result of this function is greater than the mean value, then feature vector **X** is labeled class 0. Otherwise the target is said to be in class 1. After the data was generated, the ISOC method is shown with the addition of “Storm” clouds. Next the idealized fusion models are shown and finally the application of these fusion models are shown.

In this section, the following data sets were used in calculations: $C_1(A, T)$, $C_2(A, T)$, and $C_3(A, T)$ with the F_1 feature set 1 with four features. This is following the notation given in Section 3.2.1 where the symbol $C_1(A,T)$ signifies that classifier 1, linear discriminant analysis, was trained on data set A and tested on data set T. All of the fusion rules were trained on data set T and then fused in three ways: I(V), R(V) and P(V(67,33)).

4.2.1 ISOC Rules – Scatter plot

The first step of the ISOC model is to determine all the possible rule combinations for a particular data set. Next the likelihood ratios are ordered to determine an ordered rule set that maintains the highest possible true positive rate for the lowest possible false positive rate. These results are shown in Figure 14 below. In this figure there are 2^N different rules plotted where N = 8. As shown in Section 2.5.1.3 the number of possible rules is determined by the number of sensors and the number of sensor identification outcomes.

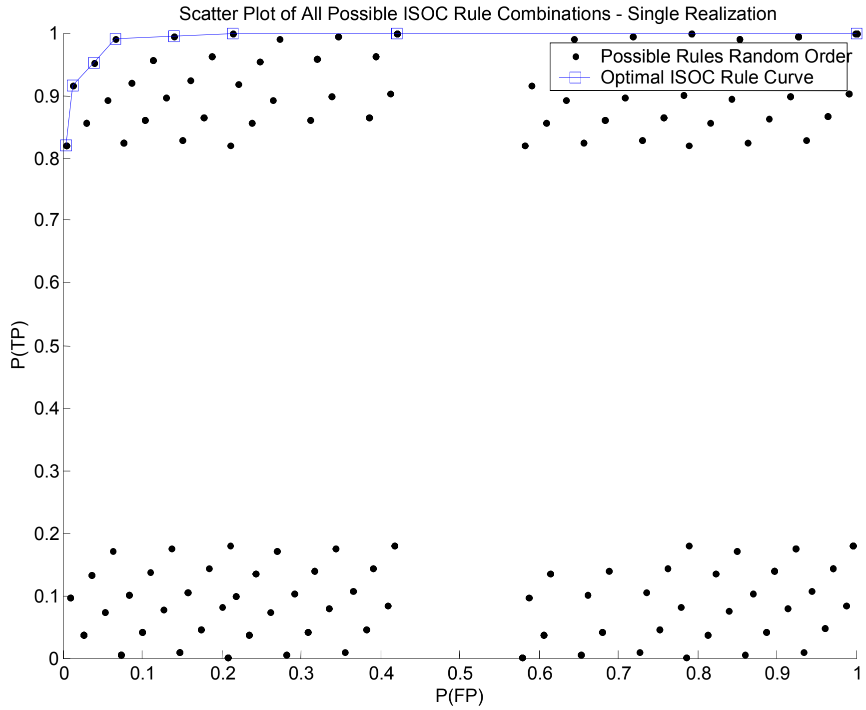

Figure 14: Single Feature Set, One Realization ISOC Rule Scatter Plot

Next the ISOC rules were explored throughout the available threshold space.

4.2.2 Storm Clouds

Storm clouds allow us to explore the two dimensional threshold space for alternative rules. They also give us an idea of the sensitivity of the fusion to changes in the thresholds. Figure 15 shows an example of how varying the thresholds for a fused ISOC classification rule can improve the true positive percentage for a given false positive percentage.

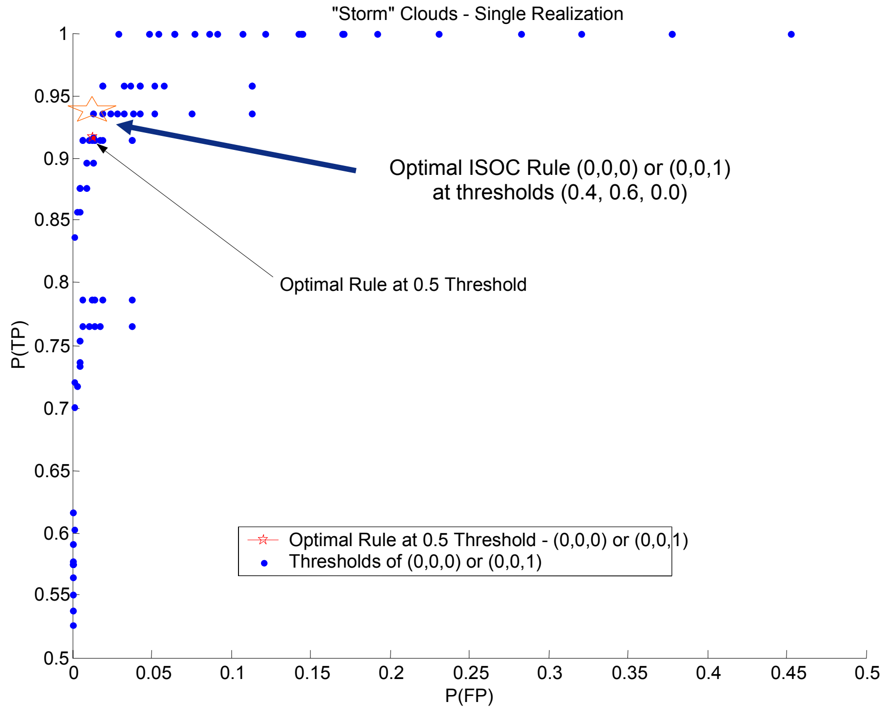

Figure 15: Single Feature Set, One Realization "Storm" Clouds for Optimum Rule

In Figure 15, it is shown, that by varying the thresholds of a given classifier, we can improve the performance of that rule. For example, in this case, a single feature set and one realization, the optimal rule (P(FP), P(TP)) is found at (0.0128, 0.9167) when the 0.5 threshold is used for all three classifiers. When the classifier vector (t_1, t_2, t_3) becomes (0.4, 0.6, 0.0) this optimal rule becomes (0.0128, 0.9362). This figure was produced using the same data analysis process as in Figure 14. The threshold determinations for each classifier were varied about the optimal rule (Section 2.4.1.4). This threshold variation produced the different values of the optimal rule shown above.

4.2.3 Idealized ROC Curves

In the ISOC and ROC fusion models, an optimal ROC curve is determined from the data set T. In this case T is the single feature set data with 100 exemplars. The ISOC optimal rule from the ISOC fusion model was found, and this rule was applied to the training data resulting in an optimal ROC curve through the thresholding method explained in Section 3.3.1. This optimal curve is shown in Figure 16 below.

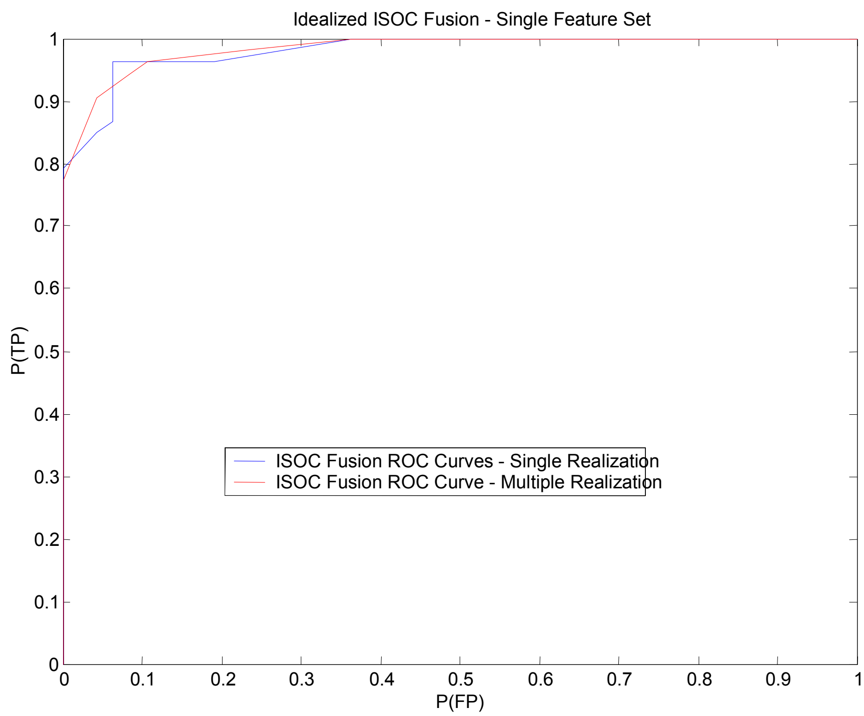

Figure 16: Single Feature Set, Idealized ISOC ROC Curves

We call these ideal ROC curves because the optimal ISOC rule has not yet been validated through the application of the rule to an independent data set. The validation ROC curves are referred to as “applied”. These ideal ROC curves show a prediction of the ISOC operating rule. Both one realization and multiple realization data are shown in

Figure 16; the multiple realization method and results are explained in Section 4.3.3. This procedure was also demonstrated for the ROC within fusion model. These optimal predictive curves can be seen in Figure 17 below.

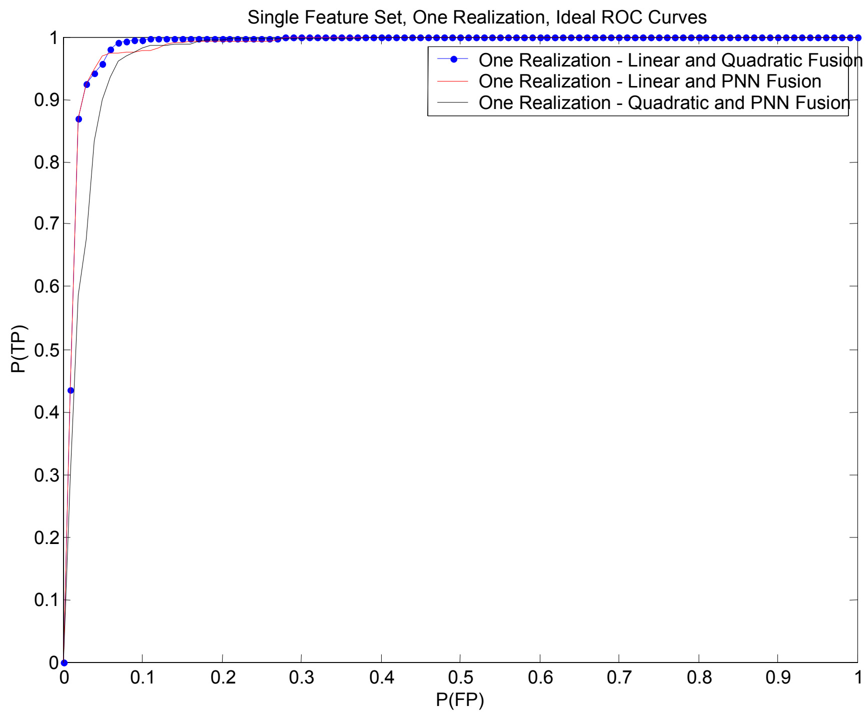

Figure 17: Single Feature Set, One Realization Idealized "Within" ROC Curves

Figure 17 shows the fusion between one realization of the single feature set data, fusing two classifiers at a time. The three classifiers were not fused due to the nature of the data. When these three classifiers are fused, the resulting curve is almost perfect and does not allow the rule to be applied to the validation data. The next step in the ROC within fusion model is to determine the optimal thresholds for each classifier from these ideal fused curves. This methodology is shown in Section 3.3.2.

4.2.4 Threshold Graphs

For each fused ROC curve, there is a set of thresholds that are optimal for the individual classifiers. These thresholds were calculated and the results can be seen in Figure 18. After the optimum thresholds were found for a given false positive rate, they were applied to the validation data to produce the applied ROC curves. The optimum thresholds described in these charts are θ (theta), ϕ (phi), and β(beta). These thresholds were found as described in Section 3.3.2 where θ corresponds to the optimum thresholds for C_1 - linear discriminants classifier, ϕ to C_2 - quadratic discriminants classifier, and β to the C_3 – PNN classifier.

These thresholds were found using the single feature set and one realization of training data set T. The final step in the ROC within fusion process was to apply these thresholds to the validation set V.

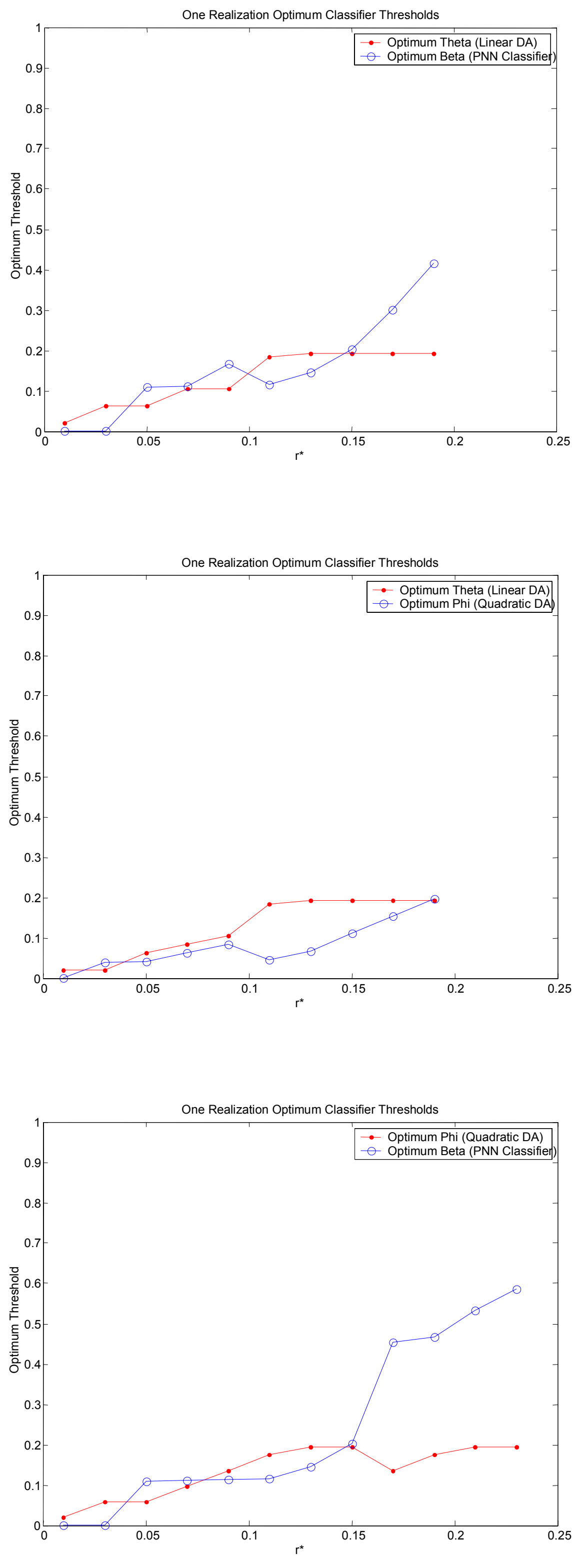

Figure 18: One Realization Optimal ROC Curve Threshold Graphs

4.2.5 Applied ISOC, ROC, and PNN curves

The final step in the ISOC fusion process was to apply the optimal rule to the same validation set. These results can be seen in Figure 19 below.

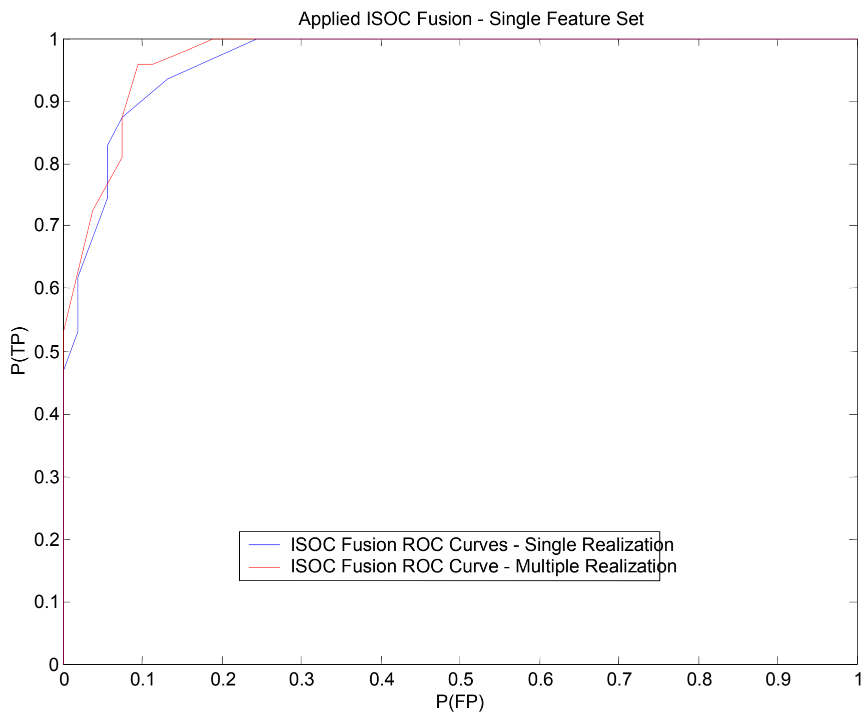

Figure 19: Applied ISOC Fusion, Single Feature Set

The applied ISOC rule curves are similar to the optimal ISOC rule curves. The performance of these curves is degraded slightly, but this fusion method shows a high true positive rate for a given false positive rate. These curves show that at a false positive rate of approximately 0.2, 100% identification of hostiles is possible. Once again, the multiple realization curve is shown and will be explained in Section 4.3.5.

In Figure 20 below, the applied ROC curves are shown. Due to the excellent performance of the classifiers, thresholds from the optimal ROC curves could only be found up to a false positive rate of 0.25. The resulting fused curves are shown below.

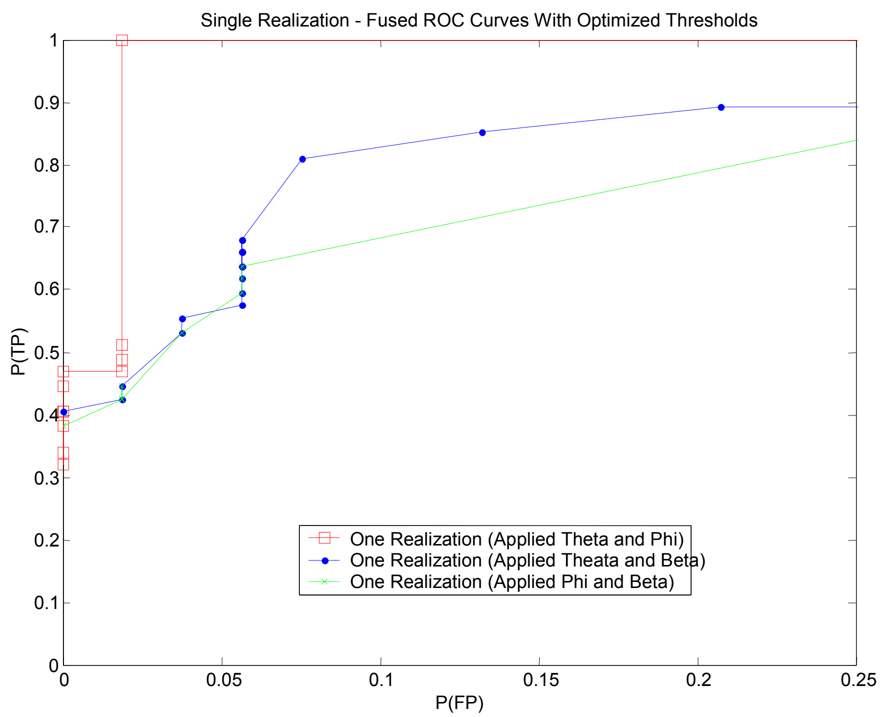

Figure 20: Applied ROC Curves, Single Feature Set, One Realization

While Figure 20 does not show the true positive rate of 100%, it can be seen in Figure 17 from Section 4.2.3, that the individual classifiers reach this rate at a false positive of approximately 0.25. Thus, the thresholding of this space is not necessary.

The third fusion tool that was used was the PNN classifier as a fusion tool. The posterior probability results from the individual classifiers were presented as features to the PNN and the following classification curves were produced. The neural net was trained using P(V(67,33)).

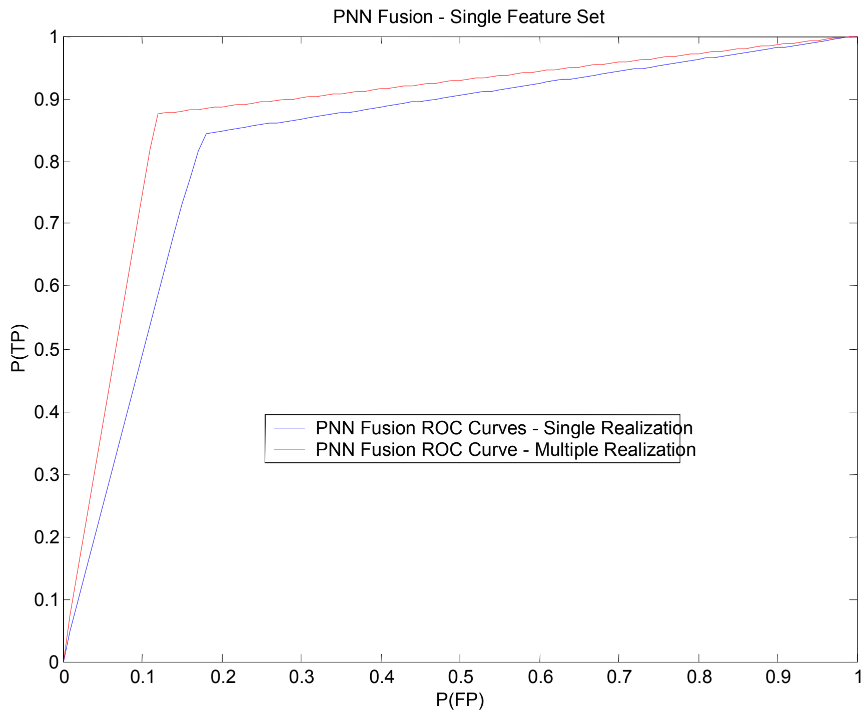

Figure 21: Applied PNN Fusion, Single Feature Set.

As in Figure 19, the PNN curve for multiple realizations is shown in Figure 20. The explanation of this curve can be seen in Section 4.3.5. Figures 19 and 21 show that one realization of a single feature set has a lower classification accuracy than that of multiple feature sets. This is most likely due to the fact that the classifiers are presented with less data for training. The fusion models also have less data available for training. When more data is present, the individual classifiers and the fusion tools maintain better performance. The complete results using multiple feature sets are shown in the following section.

4.3 Single Feature Set, Multiple Realizations

The results of the simulated data where there was a single feature set and multiple realizations are shown in this section. First the ISOC method is shown with the addition of "Storm" clouds. Next the idealized fusion models are shown and finally the application of these fusion models are shown.

In this section, the following data sets were used in calculations: C_1(B, T), C_2(C, T), and C_3(D, T) with the F_1 feature Set 1 with four features. This is following the notation given in Section 3.2.1 where the symbol C_1(A,T) signifies that classifier 1, linear discriminant classifier, was trained on data set A and tested on data set T. All of the fusion rules were found using data set T and then fused in three ways: I(V), R(V) and P(V(67,33)).

4.3.1 ISOC Rules – Scatter Plot

The first step of the ISOC model is to determine all the possible rule combinations for a particular data set. Next the likelihood ratios are ranked to determine an ordered rule set that maintains the highest possible true positive rate for the lowest possible false positive rate. These results are shown in Figure 22. In this figure there are 2^N different rules plotted where N = 8. As shown in Section 2.5.1.3 the number of possible rules is determined by the number of sensors and the number of sensor identification outcomes.

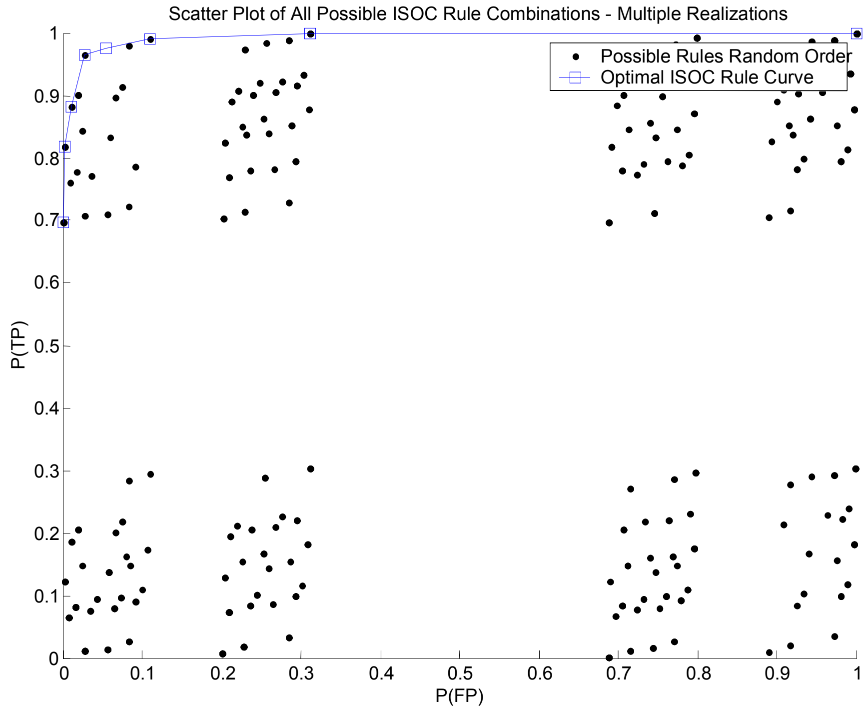

Figure 22: Single Feature Set, Multiple Realizations ISOC Rule Scatter Plot.

4.3.2 Storm Clouds

The optimal rule for multiple realizations is the same optimal rule as that of one realization. In the case of multiple realizations, this rule has a lower true positive rate, but also has a lower false positive rate. The thresholds that produce the storm clouds drastically improve the true positive rate for this optimum ISOC rule in this case. For example, in this case, a single feature set with one realization, the optimal rule (P(FP), P(TP)) is found at (0.0028, 0.8176) when the 0.5 threshold is used for all three classifiers. When the classifier vector (t_1, t_2, t_3) becomes (0.4, 0.4, 0.0) this optimal rule becomes (0.0028, 0.9163).

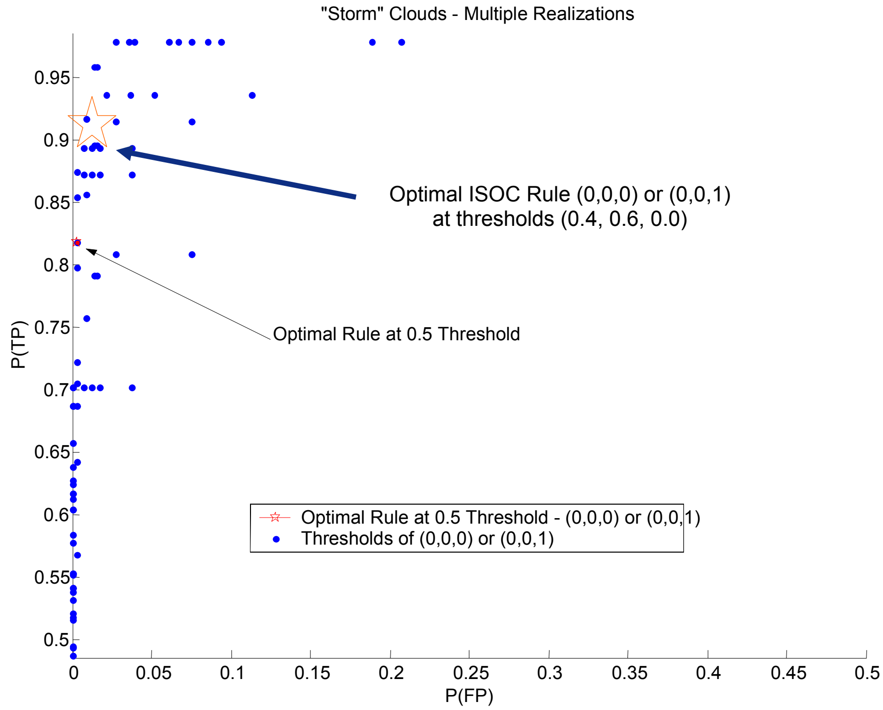

Figure 23: Multiple Realizations "Storm" Clouds for Optimum Rule

4.3.3 Idealized ROC Curves

In the ISOC and ROC fusion models, an optimal ROC curve is determined from the data set T. In this case T is the single feature set data with 100 exemplars. The ISOC optimal rule from the ISOC fusion model was found, and this rule was applied to the training data resulting in an optimal ROC curve. This optimal curve was shown in Figure 16. The idealized ROC within fusion curves are shown in Figure 24.

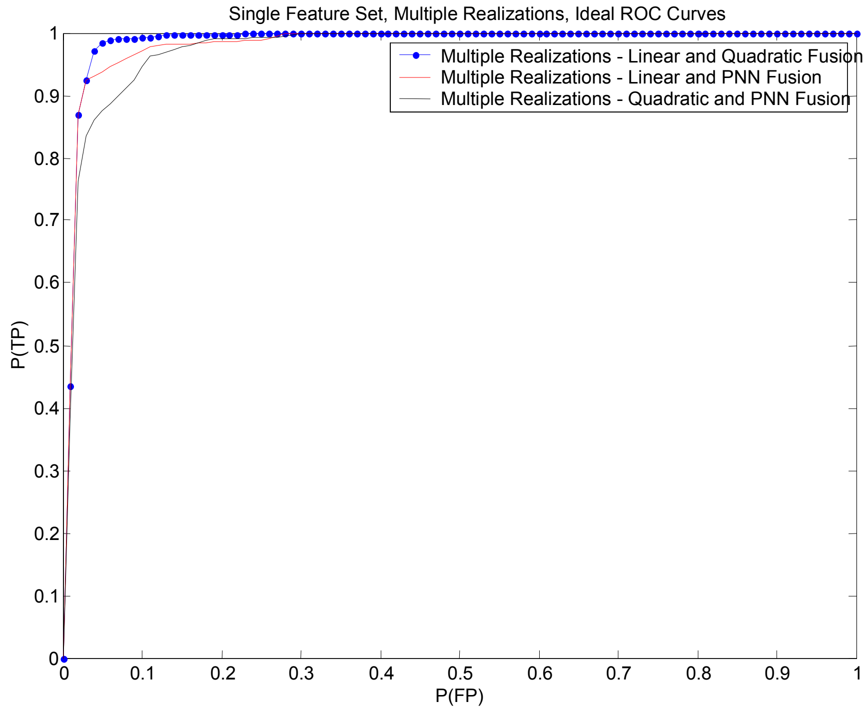

Figure 24: Single Feature Set, One Realization Idealized "Within" ROC Curves

Figure 24 shows the fusion between multiple realizations of the single feature set data, fusing two classifiers at a time. Once again the three classifiers were not fused due to the nature of the data. When these three classifiers are fused, the resulting curve is almost perfect and does not allow the rule to be applied to the validation data. The next step in the ROC within fusion model is to determine the optimal thresholds for each classifier from these ideal fused curves. This methodology is shown in Section 3.3.2.

4.3.4 Threshold Graphs

As explained in Section 4.2.4, the θ (theta) in the following graphs is the optimization of the linear discriminant classifier C_1, ϕ (phi) corresponds to the quadratic classifier C_2, and β (beta) corresponds to the PNN classifier.

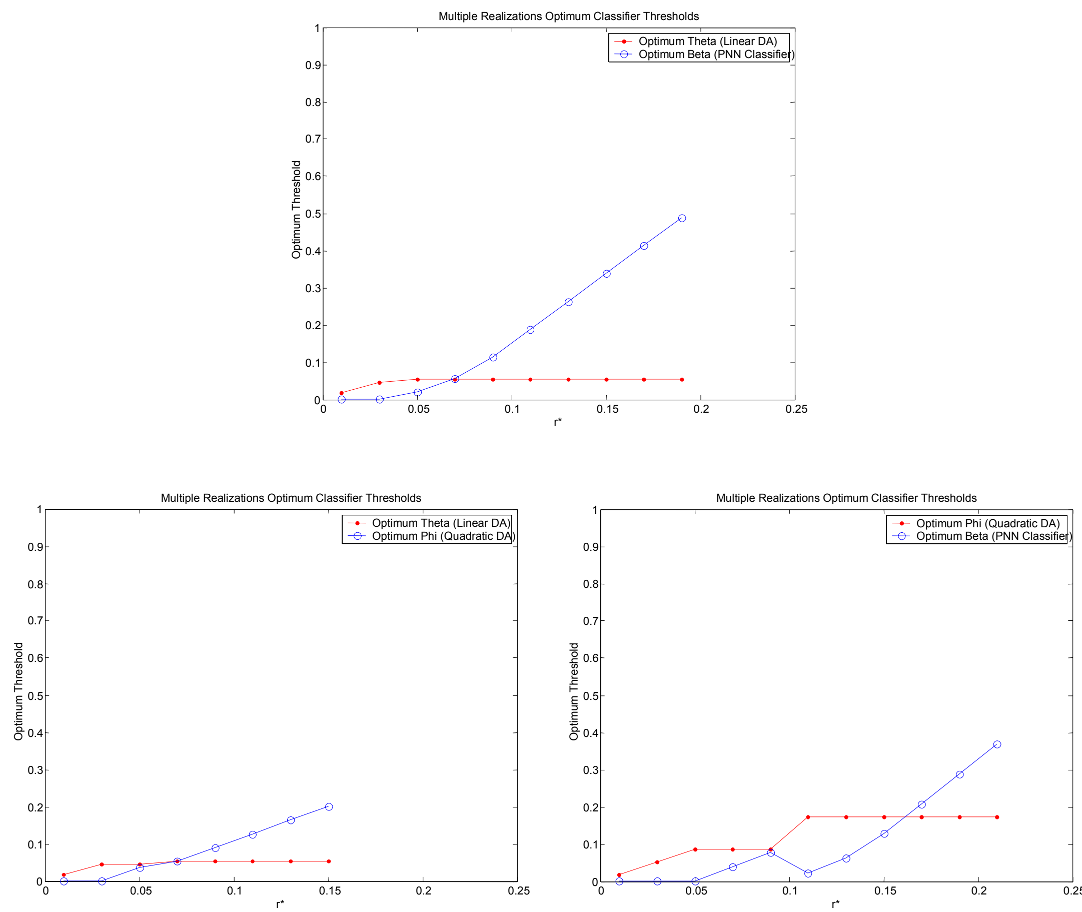

Figure 25: Multiple Realizations Optimal ROC Curve Threshold Graphs

These thresholds were found using the single feature set with one realization training data set T. These graphs show that the linear classifier is the most robust classifier, thus the optimal threshold is small and steady. The final step in the ROC within fusion process was to apply these thresholds to the validation set V.

4.3.5 Applied ISOC, ROC and PNN curves

The final step in the ISOC fusion process was to apply the optimal rule to the same validation set. These results where shown in Figure 19 in Section 4.2.5. The result from the PNN fusion for multiple feature sets was shown in Figure 21 in the same section. In both cases, when there are multiple realizations of one feature set, the classification results are higher overall. This is most likely due to the availability of more training data sets, resulting in higher classification from the individual classifiers.

The results of the applied ROC fusion can be seen in Figure 26. While Figure 26 does not show the true positive rate of 100%, it can be seen in Figure 25 from Section 4.3.3, that the individual classifiers reach this rate at a false positive of 0.15 to 0.2. Thus, the thresholding of this space is not necessary.

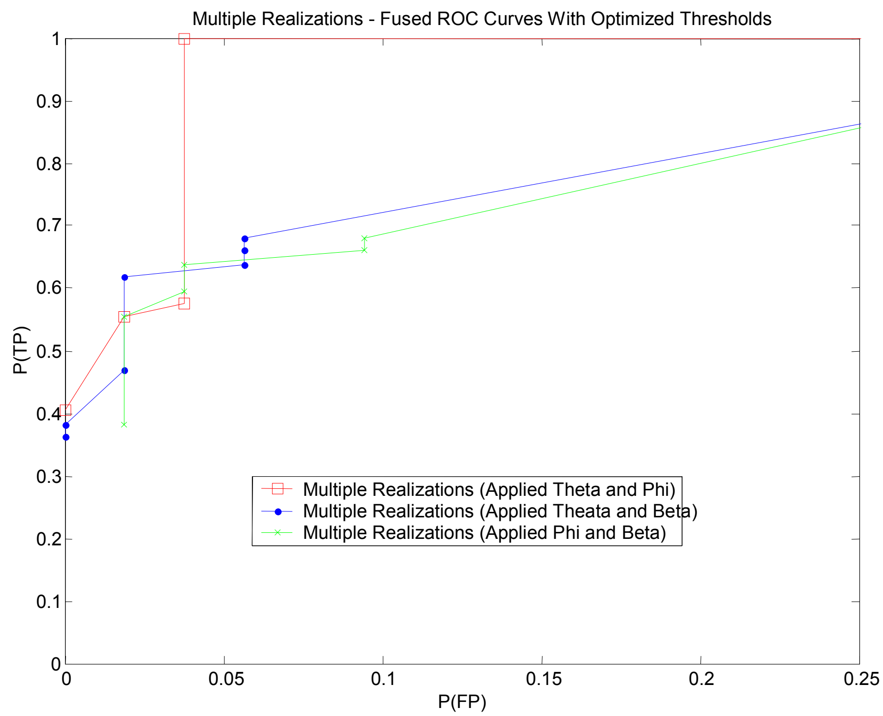

Figure 26: Applied ROC Curve Fusion, Single Feature Set, Multiple Realizations

In Figure 26, the applied ROC curves are shown. Due to the excellent performance of the classifiers, thresholds from the optimal ROC curves could only be found up to a false positive rate of 0.25. At that point, the probability of a true positive identification goes to one.

The single feature set data showed that multiple realizations allowed for greater generalization of the fusion models than one realization. This data also showed that both the ISOC and the within ROC fusion methods were similar in performance. The PNN gave the expected results as the multiple realization curve performed better than one realization. The next step in this research was to analyze the multiple feature set data.

4.4 Multiple Feature Sets

The multiple feature set data consists of two feature sets that vary across correlation. This data set was generated in the following manner. Let $F = F_1 \times F_2 \subset \Re^4$ where F_1 is feature set 1 and F_2 is feature set 2.

where

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{F_1,F_2} \\ \Sigma_{F_1,F_2} & \Sigma_{2,2} \end{bmatrix}$$

and

$$\Sigma_{1,1} = \Sigma_{2,2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \Sigma_{F_1,F_2} = \begin{bmatrix} 0 & \rho \\ \rho & 0 \end{bmatrix}.$$

where $\rho \in \{0, 1/n, \ldots 4/n\}$ and $n = 5$ and $\Sigma_{1,1}$ is the correlation matrix between the features contained in the feature set F_1 and class 1. Let $F_1 = F_{1,1} \cup F_{1,2}$ where

$$F_{1,1} \sim N_2(\mu_{1,1}, \Sigma_{1,1}) \text{ and } F_{1,2} \sim N_2(\mu_{1,2}, \Sigma_{1,2})$$

and where

$$\mu_{1,1} = (0,0)^T \text{ and } \mu_{1,2} = (0.95, 0.95)^T.$$

Let $F_2 = F_{2,1} \cup F_{2,2}$ where

$$F_{2,1} \sim N_2(\mu_{2,1}, \Sigma_{2,1}) \text{ and } F_{2,2} \sim N_2(\mu_{2,2}, \Sigma_{2,2})$$

and where

$$\mu_{2,1} = (0,0)^T \text{ and } \mu_{2,2} = (1.15, 1.15)^T.$$

In this case the inter-correlation between the features in a specific set is zero. In this method, three data sets were generated for each feature set.

In this section, the following data sets were used in calculations: $C_1(A, T)$, $C_2(A, T)$, with the F_1 feature set 1 with two features f_1 and f_2, while the F_2 feature set had two features f_3 and f_4. In this data set, f_1 is correlated with f_4, while f_2 is correlated with f_3. This ensures that there is no correlation present in the individual classifiers, but it is present during the fusion process. This is following the notation given in Section 3.2.1 where the symbol $C_1(A,T)$ signifies that classifier 1, linear discriminant analysis, was trained on data set A and tested on data set T. All of the data sets consist of 2000 exemplars. All of the fusion rules were found using data set T and then fused in three ways: I(V), R(V) and P(V(667,1333)).

The results of the simulated data where there were two feature sets and variable correlations are shown in the following figures. First, the ISOC method is shown with the addition of "Storm" clouds. Next, the idealized fusion models are shown and finally the application of these fusion models are shown.

4.4.1 ISOC Rules – Scatter Plot

The first step of the ISOC model is to determine all the possible rule combinations for a particular data set. Next, the likelihood ratios are ordered to determine the rules that maintain the highest possible true positive rate for the lowest possible false positive rate. These results are shown in Figure 27. In this figure there are 2^N different rules plotted where N = 4. As shown in Section 2.5.1.3 the number of possible rules is determined by the number of sensors and the number of sensor identification outcomes.

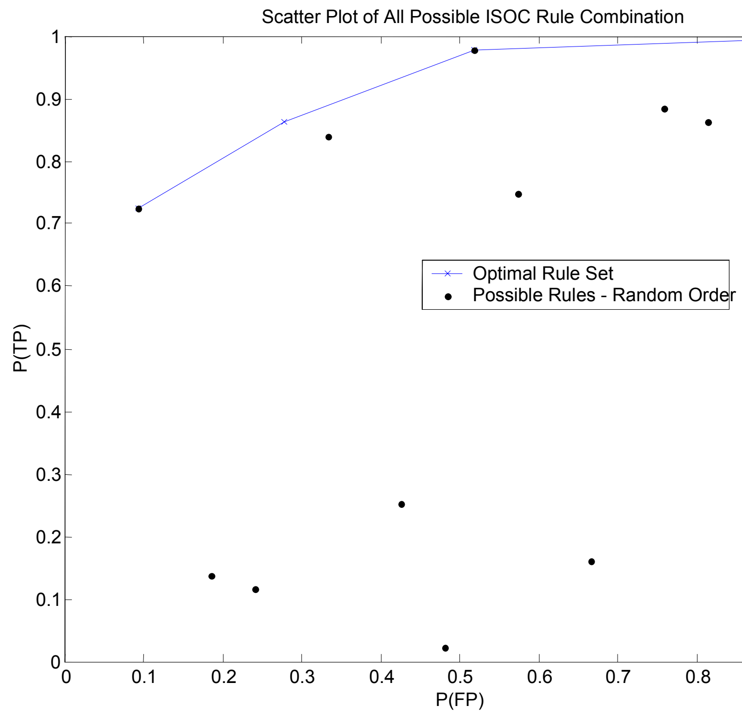

Figure 27: ISOC Possible Rule Sets – Zero Correlation

The number of possible rule sets is decreased drastically due to the removal of one sensor/individual classifier.

4.4.2 Storm Clouds

As in the case of a single feature sets, the true positive rate for a given false positive of an optimal ISOC rule can be improved by varying the thresholds of the individual classifiers. For example, in this case, a single feature set with one realization, the optimal rule (P(FP), P(TP)) is found at (0.0926, 0.7249) when the 0.5 threshold is used for both classifiers. When the classifier vector (t_1, t_2) becomes (0.4, 0.4) this optimal rule becomes (0.0900, 0.7370). The results at zero correlation can be seen below.

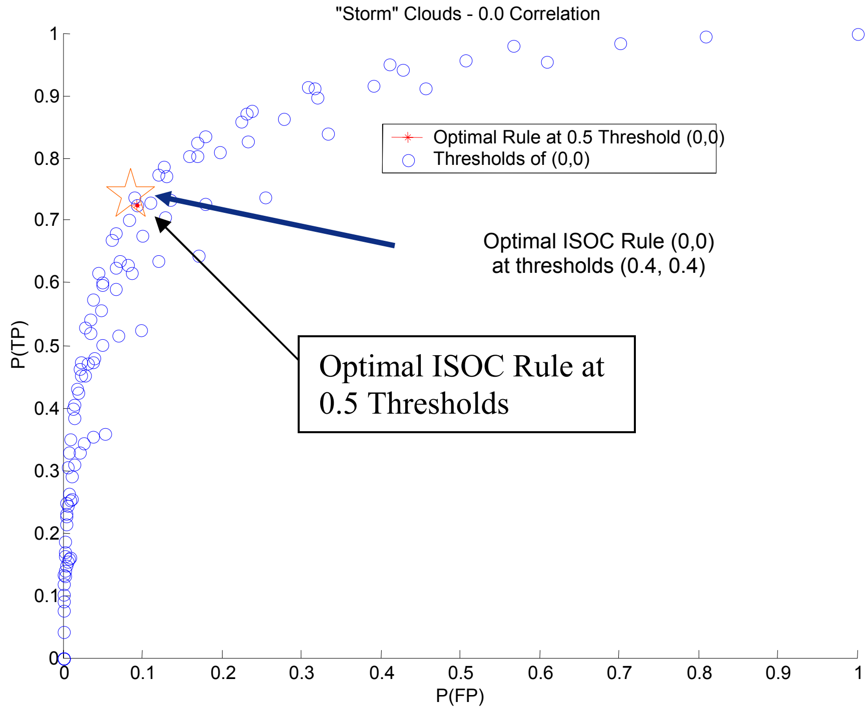

Figure 28: Storm Clouds – Zero Correlation

In Figure 28, it is shown, that by varying the thresholds of a given fusion rule, we can improve the performance of that rule. This figure was produced using the same data analysis process as Figure 27. Once the optimal rule was found, the threshold determinations for each classifier were varied. This threshold variation produced the different optimal rules shown above. In this case the optimal rule for zero correlation is simply the rule that both classifiers determine the target is hostile or (0, 0).

4.4.3 Idealized ROC Curves

In the idealized curves shown in Figures 29 and 30, it can be seen that both the ISOC fusion method and the ROC fusion method are relatively insensitive to correlation. The optimal rule for the ISOC method was chosen by looking at the data for 0.4

correlation. This rule was determined to be (0, 0) or (1, 0). This notation signifies that (C_1 declared the target a hostile or 0 and C_2 declared the target hostile or 0) or C_1 declared the target hostile and C_2 declared the target a friend).

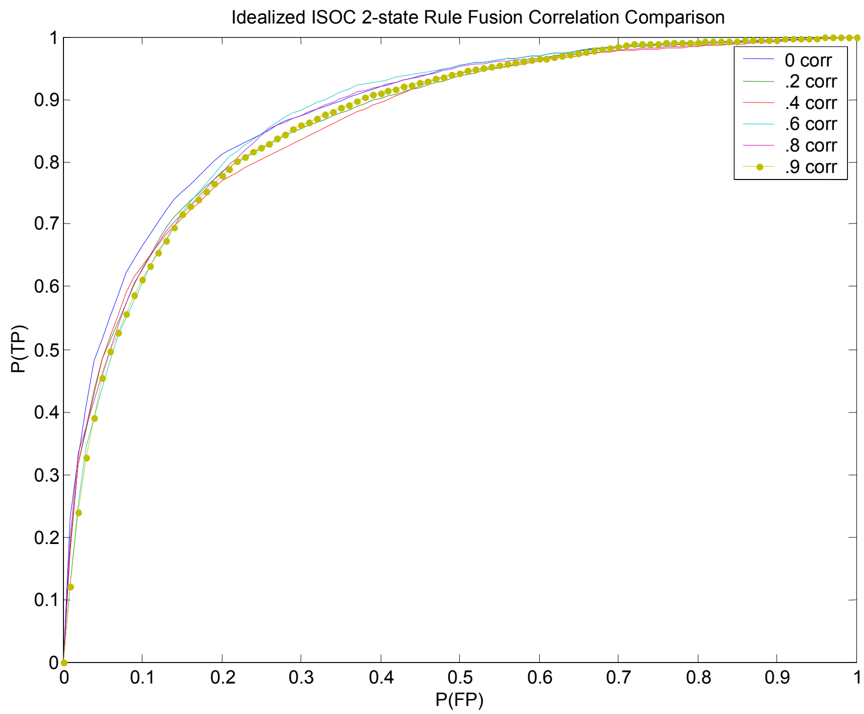

Figure 29: Idealized ISOC ROC Curves for Two Feature, Two-Class Problem

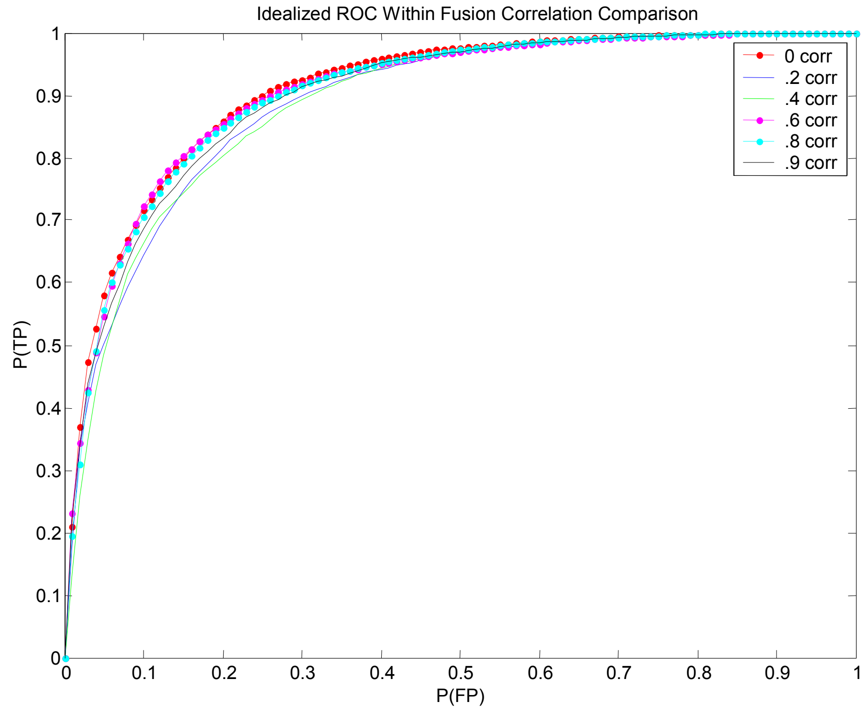

Figure 30: Idealized ISOC ROC Curves for Two Feature, Two-Class Problem

4.4.4 Threshold Graphs

After the idealized results were calculated, the next step in both the ISOC and ROC fusion models is to apply the rules to the validation data set. In order to do this in the ROC fusion case, first the optimal thresholds for each correlation must be calculated. These results are shown in Figure 31 at each correlation level. As before, the fused ROC curves reach a true positive rate of 100% at a low false positive rate. Thus, these results are only shown to that point. It should be noted that as the correlation gets higher these thresholds get closer and closer together.

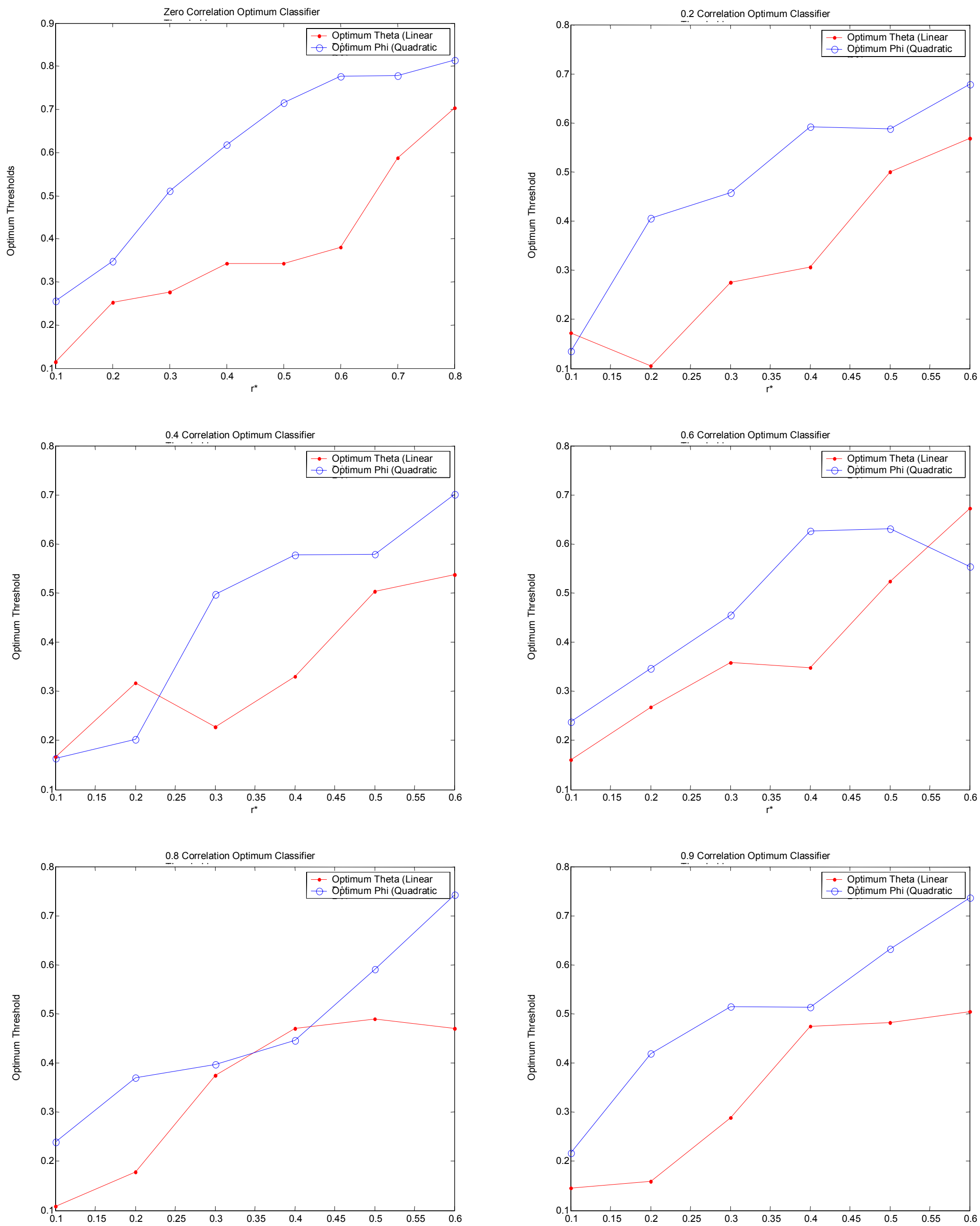

Figure 31: ROC "Within" Fusion Thresholds at Various Correlations

4.4.5 Applied ISOC, ROC and PNN Curves

The final fusion process step is to apply the ISOC and ROC fusion rules to the validation data set. These applications are shown in Figures 32 and 33. As with the idealized curve, the ISOC applied rule is still relatively insensitive to correlation.

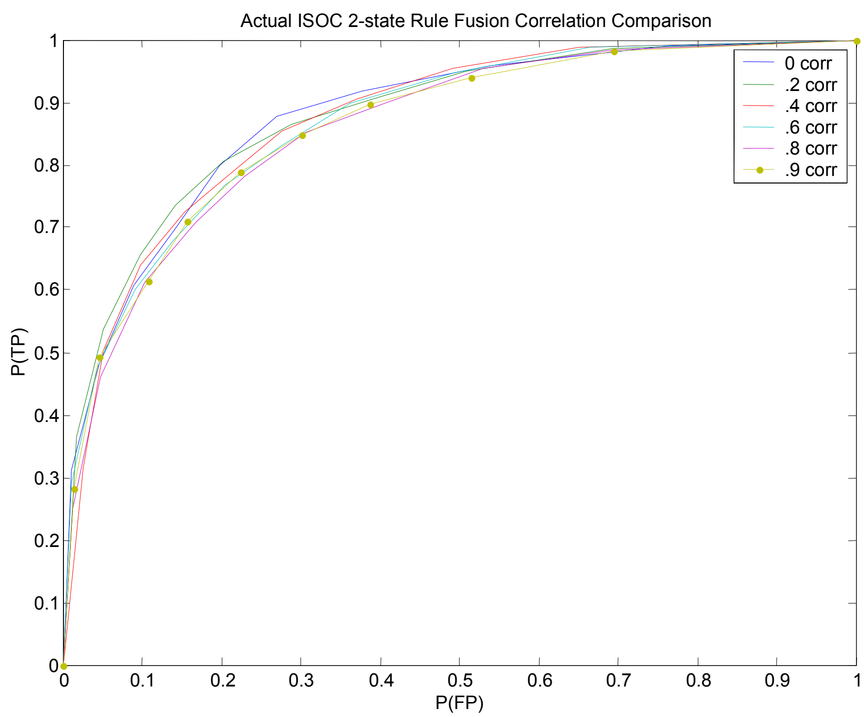

Figure 32: Optimal Rule ISOC Curves for Two Feature, Two-Class Problem

The application of the ROC fusion rule is more sensitive to the correlation between features. This is shown in Figure 33. Once again, since the thresholds for the classifiers were only found up until a FP rate of 0.3, these curves are only plotted to that point.

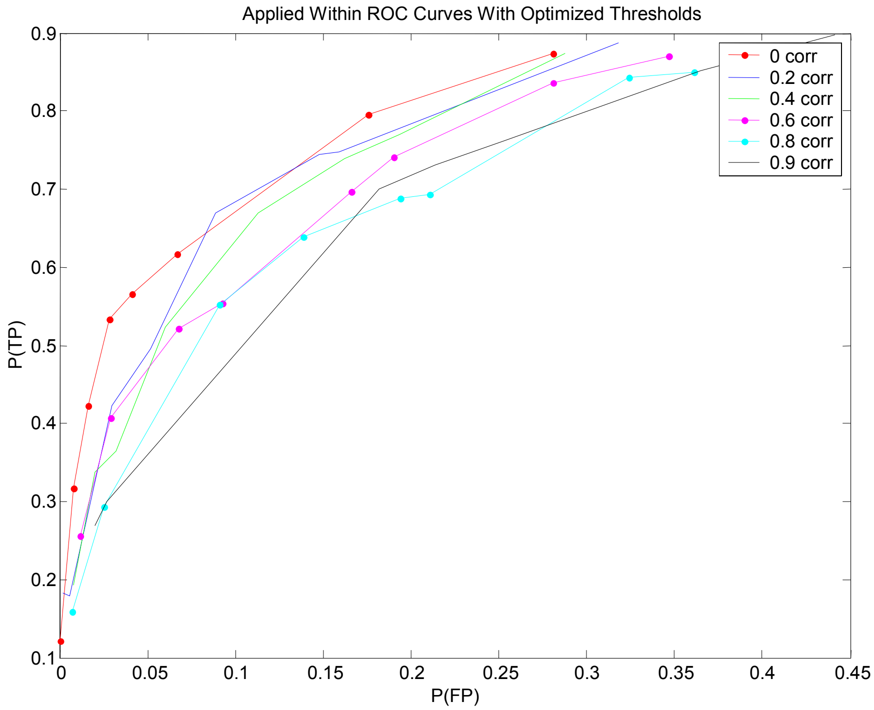

Figure 33: Threshold Applied ROC Curves for Two Feature, Two Class Problem

The PNN fusion was much more sensitive to the correlation between features than the ISOC or ROC fusion. This can be seen in Figure 34.

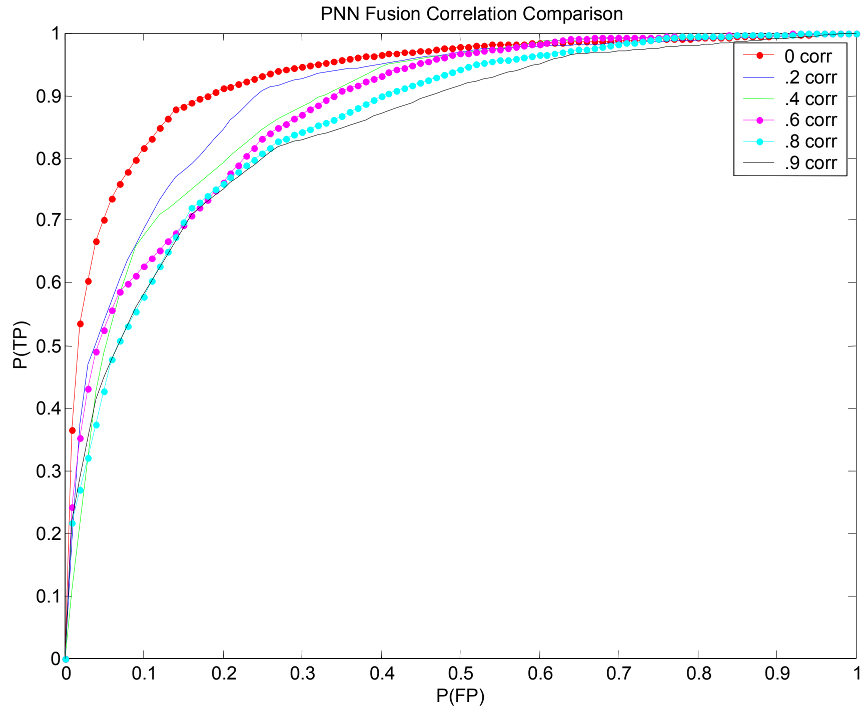

Figure 34: Applied PNN ROC Curves for Two Feature, Two-Class Problem

From the results in Section 4.4 we can see that the ISOC model is more robust to correlation than the other fusion tools. The PNN maintains a higher performance at low correlation, however this performance is degraded at a high level of correlation. The within ROC curve has a similar performance to the ISOC curve at low correlation, but is degraded slightly at a higher correlation. It has been shown that with the ISOC method, the best rule at set thresholds is found, while the within ROC fusion method finds optimal thresholds for a set rule. These conclusions can be seen in Figure 35.

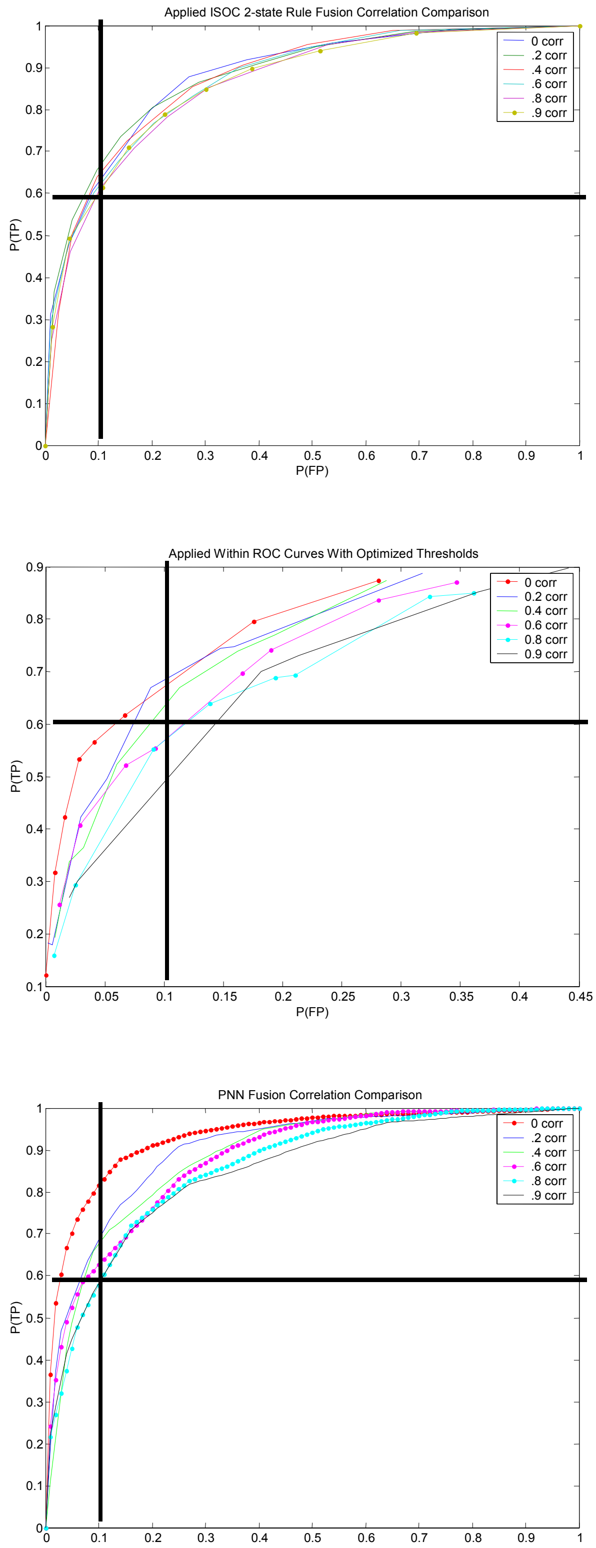

Figure 35: Comparison of Three Fusion Models

4.5 Chapter Summary

The ISOC, ROC and PNN fusion models can be compared and contrasted with interesting data. This chapter showed the results of two major cases of data, single feature set data and multiple feature set data. In the single feature set data, two cases of data analysis were presented, one realization and multiple realizations. The effects of correlation on three fusion schemes were displayed. Test situations were developed to allow the examination of various levels of correlation both between and within feature streams. The effects of training a fusion ensemble on a common data set versus an independent data set were also contrasted. Some incremental improvements to the ISOC procedure were discovered in this process.

Several conclusions can be made regarding these results. First, fusing classifiers trained on independent data sets is generally better than fusing classifiers trained on the same data set. Second, the ISOC method can be improved by searching the parameter space. Third, the ISOC method appears to be the most robust to correlation. Finally, the PNN is an extremely simple, easy to apply method that outperforms the other fusion methods at low correlation levels. This thesis is the first step towards the creation of a synthetic classifier fusion-testing environment. These effects and others appear to be useful to the creators of the next steps in this environment.

V. Conclusion

5.1 Introduction

The goal of this thesis was to exercise several fusion models, on several techniques, across interesting data sets to assess the outcomes. The fusion models explored were the ISOC fusion model, the ROC "Within" fusion model and a probabilistic neural net (PNN) used as a fusion tool. Due to unavailability of real-world data and for control purposes, we generated artificial data for this study.

5.2 Literature Review Findings

Several interesting references were found in this area. It was shown that Air Force Doctrine and targeting guidance requires a specified level of information accumulation. This level of accumulation can be achieved through fusing the information from several sensors. The level of information accumulation that is required is dependent on the specific target, but it can be assumed that this information cannot be collected safely through one information source alone. Several sources are required, which leads to sensor fusion. Due to the complexity of sensor fusion, several models have been developed and assumptions made in those models must be closely inspected.

Next, the latest research in the independence of fusion rules and their dependence on data diversity was discussed. The different fusion models we chose were reviewed. The three models that we chose were the ISOC fusion model, the ROC "Within" fusion model, and a probabilistic neural net as a fusion tool.

5.3 Methodologies Employed

The methodology employed in this thesis involved both data generation and fusion model analysis. The method of data generation was discussed, in both the case of a single feature set and multiple feature sets. The difference between intra-correlation and inter-correlation was explained, and the application of inter-correlation to a data set was exemplified. Next the design of the experiment in both feature set applications was demonstrated.

5.4 Conclusive Results

Several conclusions can be made regarding these results. First, fusing classifiers trained on independent data sets is generally better than fusing classifiers trained on the same data set. Second, the ISOC method can be improved by searching the parameter space. Third, the ISOC method appears to be the most robust to correlation. Finally, the PNN is an extremely simple, easy to apply method that outperforms the other fusion methods at low correlation levels. This thesis is the first step towards the creation of a synthetic classifier fusion-testing environment. These effects and other appear to be useful to the creators of the next steps in this environment.

5.5 Recommendations for Future Research

This thesis explores the effects of correlation on sensor fusion. It is a starting point for many future studies. There are two major areas that are proposed for future study. The first major area involves simulated data. First the issue of intra-correlation should be explored. Second, the inter-correlation of a feature set should be explored.

Finally, in the simulated data, noise should be added to the data and the effects of this noise on the fusion efforts should be explored.

The second major area is real-world data. Due to the unavailability of data and the classification issue, real-world data was not used in this study. The results of this thesis should be validated using actual sensor data from a weapons system to test the accuracy of the models and measurements.

Appendix: Individual Classifier Results

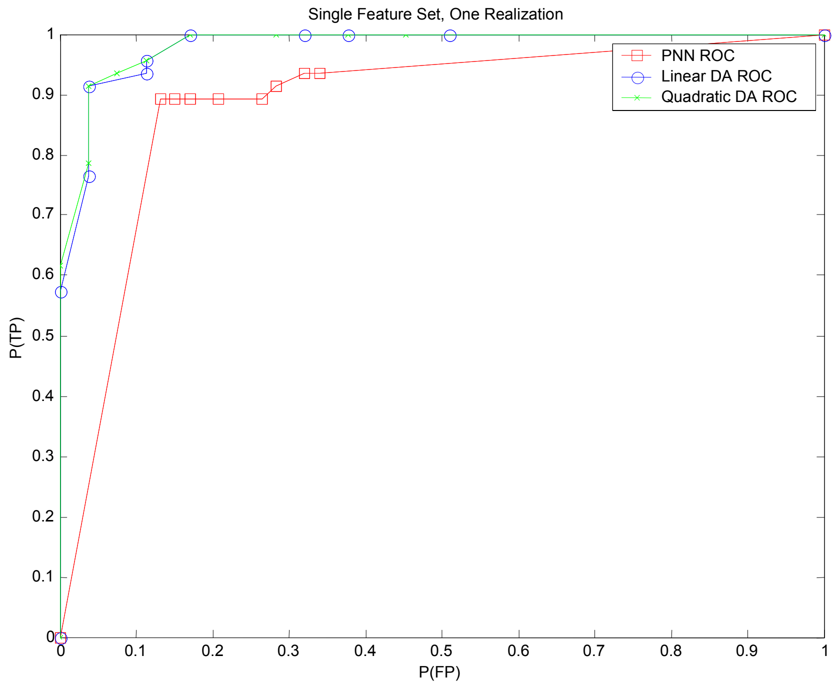

Figure A.1: Classifier Results, Single Feature Set, One Realization

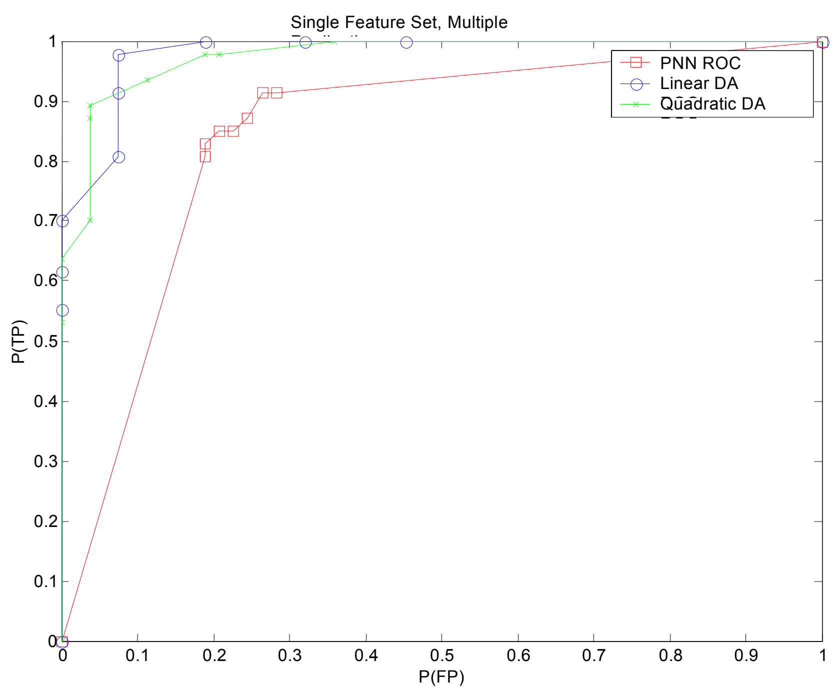

Figure A.2: Classifier Results, Single Feature Set, Multiple Realizations

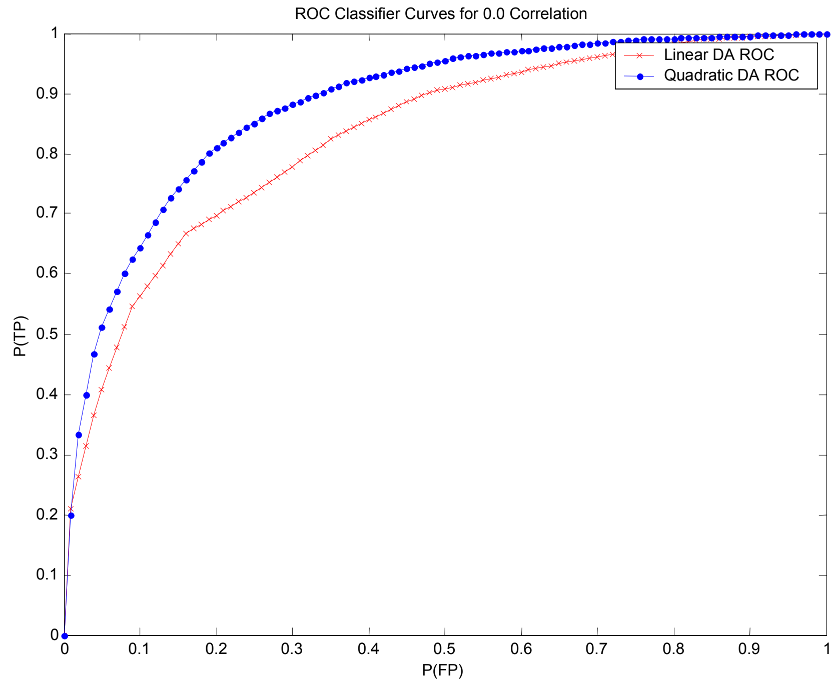

Figure A.3: Classifier Results, Two Feature Sets, 0.0 Correlation

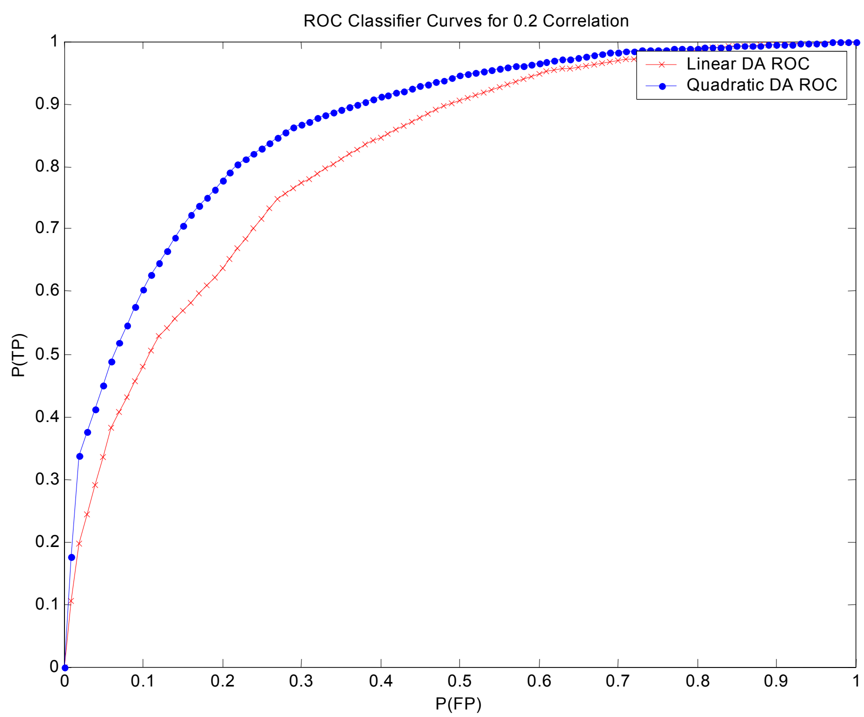

Figure A.4: Classifier Results, Two Feature Sets, 0.2 Correlation

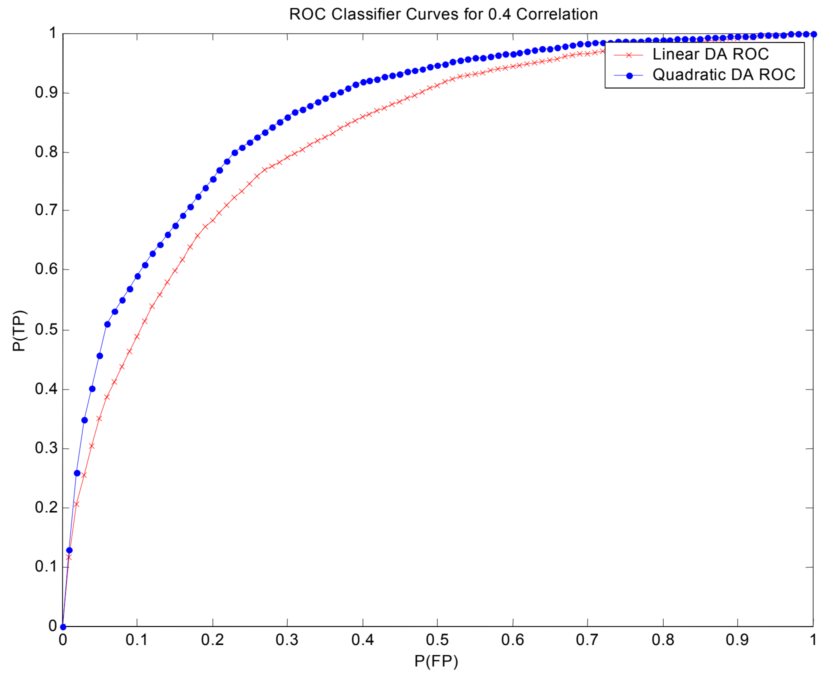

Figure A.5: Classifier Results, Two Feature Sets, 0.4 Correlation

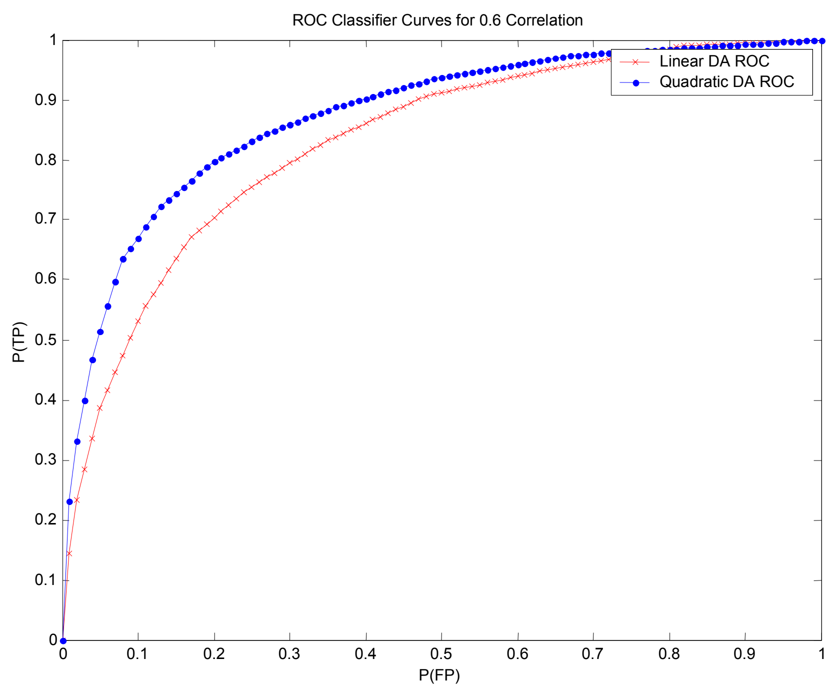

Figure A.6: Classifier Results, Two Feature Sets, 0.6 Correlation

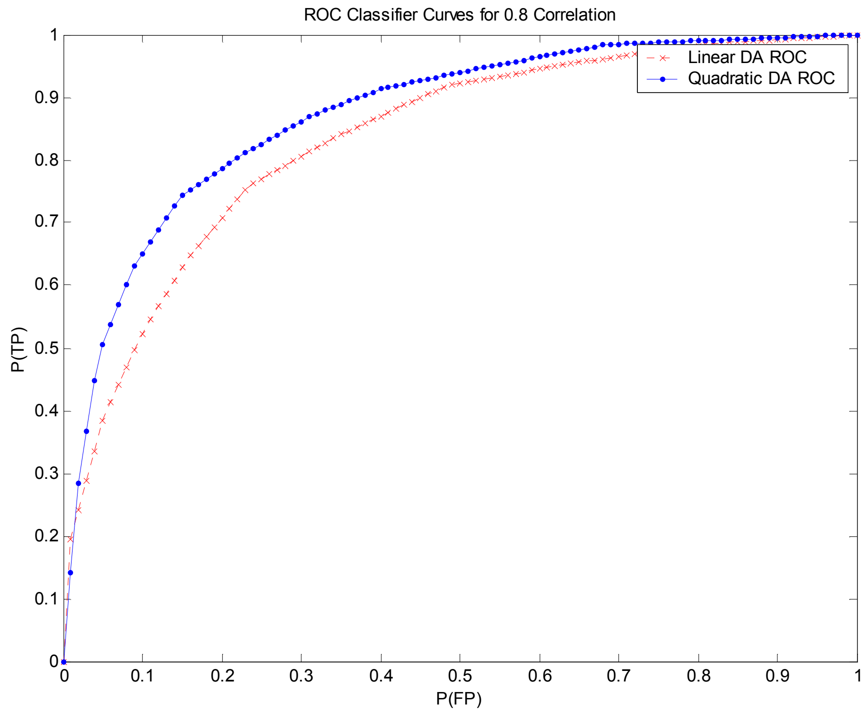

Figure A.7: Classifier Results, Two Feature Sets, 0.8 Correlation

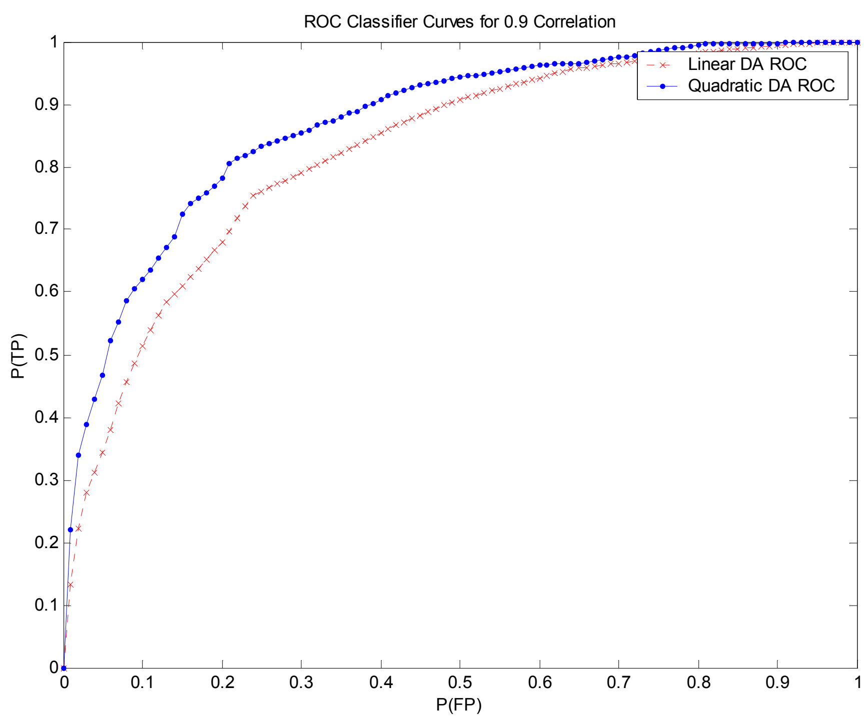

Figure A.8: Classifier Results, Two Feature Sets, 0.9 Correlation

Bibliography

Air Force Doctrine Document 2-1, *Air Warfare*, 22 January 2000.

Air Force Pamphlet 14-210, *USAF Intelligence Targeting Guide*, 1 February 1998.

Haspert, J.K., "Optimum ID Sensor Fusion for Multiple Target Types." *IDA Document D2451*, 2000.

Clutz, T., *A Framework for Prognostics Reasoning*. Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 2002.

Egan, J. P., *Signal Detection Theory and ROC Analysis*, Academic Press: New York, 1975, pp 31-36.

Oxley, M.E., and K. Bauer, "Classifier Fusion for Improved System Performance." *AFIT/ENS Working Document 02-02*, 2002.

Ralston, J.M., "Bayesian Sensor Fusion for Minimum-Cost I.D. Declaration." *1998 Joint Service Combat Identification Systems Conference on Requirements, Technologies and Developments*, (CISC-98), Volume 1 – Technical Proceedings.

Shipp, C.A and L.I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers." *Information Fusion*, vol 3, iss 2: 2 June 2002, pp. 135-138.

Wackerly, D.D., W. Mendenhall III, and R. L. Scheaffer, *Mathematical Statistics with Applications*, 6th edition, Duxbury: Australia, 2002.

Wasserman, P., and V. Nostrand, *Advanced Methods in Neural Computing.* 1993.

Willet, P., P.F. Swaszek, and R.S. Blud, "The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise." *IEEE Transactions on Signal Processing*, vol 48, iss 12: 12 December 2000, pp. 3266-3279.

Vita

Captain Susan A. Storm graduated from Heber Springs High School in Heber Springs, Arkansas. She entered undergraduate studies at the United States Air Force Academy (USAFA) where she graduated with a Bachelor of Science Degree in Space Physics in May 1998. She was commissioned through USAFA with a Reserve Commission.

Her first assignment was at Kirtland AFB as a research physicist for the Air Force Research Laboratory, Advanced Optics and Imaging Division, Surveillance Technologies Branch. In August 2002, she entered the Graduate School of Engineering and Management, Air Force Institute of Technology. Upon graduation, she will be assigned to Randolph AFB, Texas.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
03-2003	**Master's Thesis**	Jun 2002 – Mar 2003

4. TITLE AND SUBTITLE	
AN INVESTIGATION OF THE EFFECTS OF CORRELATION IN SENSOR FUSION	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER** QAF185025203203
Storm, Susan, A., Capt., USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-7765	AFIT/GOR/ENS/03M-22
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR Attn: Major Juan R. Vasquez 801 North Randolph Street, Room 933 (703) 696-8431 Arlington, VA 22203-1977 e-mail: Juan.Vasquez@afosr.af.mil	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFOSR **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This thesis takes the first step towards the creation of a synthetic classifier fusion-testing environment. The effects of data correlation on three classifier fusion techniques were examined. The three fusion methods tested were the ISOC fusion method (Haspert, 2000), the ROC "Within" Fusion method (Oxley and Bauer, 2002) and the simple use of a Probabilistic Neural Network (PNN) as a fusion tool. Test situations were developed to allow the examination of various levels of correlation both between and within feature streams. The effects of training a fusion ensemble on a common dataset versus an independent data set were also contrasted. Some incremental improvements to the ISOC procedure were discovered in this process.

15. SUBJECT TERMS

Sensor Fusion, Classification, Probabilistic Neural Network, ISOC Fusion, ROC Curve Fusion

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Kenneth W. Bauer, AFIT/ENS
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-6565, ext 4328; e-mail: Kenneth.Bauer @afit.edu
U	U	U	UU	94	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18