AN INVESTIGATION OF THE OPTIMAL
SENSOR ENSEMBLE FOR SENSOR FUSION

**THESIS**

Paul P. Clemans, Captain, USAF

AFIT/GOR/ENS/04-03

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**
# *AIR FORCE INSTITUTE OF TECHNOLOGY*

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

AFIT/GOR/ENS/04-03

AN INVESTIGATION OF THE OPTIMAL
SENSOR ENSEMBLE FOR 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

Paul P. Clemans, BS

Captain, USAF

March 2004

AFIT/GOR/ENS/04-03

AN INVESTIGATION OF THE OPTIMAL
SENSOR ENSEMBLE FOR SENSOR FUSION

Paul P. Clemans, BS
Captain, USAF

Approved:

_____// Signed //_____          ___5 March 2004__
Kenneth W. Bauer (Chairman)                                        date


_____// Signed //_____          ___5 March 2004__
Mark E. Oxley  (Member)                                              date

**Abstract**

This thesis continues the research begun by Storm, Bauer and Oxley in 2003 into the fusion of classifiers. It examines the fusion of up to three correlated classifiers using three different fusion techniques. The overall objective was to determine the optimal ensemble of classifiers to maximize the expected classification accuracy. The ISOC fusion method (Haspert, 2000), the ROC "Within" fusion method (Oxley and Bauer, 2002) and a Probabilistic Neural Network were the three fusion techniques employed in these set of experiments. Performance of the classifiers and the fusion methods is measured via ROC curves. Two possible configurations of feature correlations were examined. The expected true positive value relative to a prior distribution of correlation levels for each configuration was then used to compare the classifier and the fused classifiers performance and thereby allowing for the selection of an optimal ensemble.

*To my wife*

# Acknowledgments

I would like to thank my faculty advisor, Dr. Kenneth Bauer, for his guidance, and insight throughout the course of this thesis effort. More importantly, I would like to express my sincere appreciation for his constant encouragement and assurance. I always left his office believing I could achieve this.

To my companion in this journey, Capt Nate Leap, I would convey my deepest gratitude for his patience, for sharing his understanding with me time and again, and for checking endless lines of code with me. If it were not for him, this thesis would probably be much shorter. I count myself fortunate to consider him a friend.

I would also like to thank Maj Trevor Laine and Maj Tim Albrecht for their willingness to take the time to share their understanding with me. Your indomitable spirit and encouragement really changed really encouraged me on many occasions.

To my wife and daughter, your abiding love, active support and constant understanding made this last year bearable. I do not know how to express my gratitude for the countless unspoken sacrifices you made over the last year and a half. My greatest regret at AFIT has been the time I missed with you, but now the thing is done!


Paul P. Clemans

# Table of Contents

# List of Figures

**List of Tables**

AN INVESTIGATION OF THE OPTIMAL

SENSOR ENSEMBLE FOR SENSOR FUSION


## I.       Introduction

**Background**

Effective Command and Control (C2) depends in large part on the ability to accurately identify all of the hostile, friendly and neutral entities in the battlespace referred to as Combat Identification (CID).  Accurate CID hinges on the ability to effectively process data to build a three-dimensional picture of the battlespace. This in turn permits real-time application of tactical options so weapons can be employed at optimal ranges against the most critical enemy targets (Peters and Ryan, 1998).  In other words, commanders require accurate CID to obtain a situational awareness of the battlespace which allows them to effectively prosecute their operations.

Across the Department of Defense (DoD) reliable CID in operations has consistently proven to be an elusive capability.  Thirty-five Americans were killed and 72 wounded due to "friendly fire" or fratricide during the Gulf War (Report to Congress, 1992).  Approximately 68 percent of these incidents appeared to be the result of target misidentification and/or coordination problems (Report to Congress, 1992).  Since the Coalition controlled the battlespace in every aspect of the war these casualties represented a need for better situational awareness (i.e., identification) of forces in the battlespace.  Three years later, two F-15E aircraft shot down two UH-60 Blackhawks over Iraq in Operation Provide Comfort.  This tragedy also illustrates a breakdown in situational awareness/combat identification in that the F-15E pilots coordinated with an Airborne Warning and Control System (AWACS) aircraft before firing. Misidentification continued to be a problem even through the recent events in Operation

Iraqi Freedom.  For example, on 25 March 2003 a Patriot surface-to-air battery in Iraq came under mortar fire.  The crew engaged the batteries automatic systems and took cover.  The system's radar misidentified a local F-16 as hostile and locked-on to the aircraft.  The F-16 responded by firing a High-speed Anti-Radiation Missile (HARM) at the battery, destroying its radar dish (Weisman, 2003).  More cases could be mentioned but these serve to illustrate that misidentification and subsequent fratricide is a very real problem for the U.S. military.  In fact, Lt. Gen. Conway, who led over 85,000 Marines into Operation Iraqi Freedom, has said that the amount of fratricide was probably his biggest disappointment of the war (Conway, 2003).

Combat identification can be divided into four mission areas: air-to-air, air-to-ground, ground-to-air and ground-to-ground.  Each of these four mission areas has their own architecture.  There is no overarching architecture for CID (GAO Report, 2001).  For example, U.S. aircraft often use Identification Friend or Foe (IFF) to identify other aircraft.  Vehicles on the ground; however, might use thermal plates or thermal tape to identify friendly forces.  Not only are the sensors in each of these architectures different but the decision makers in each case varies as well.  Frontline soldiers use their training and understanding of the Rule Of Engagement (ROE) to make friend or foe decisions.  The air forces, on the other hand, usually coordinate with an air operations controller.  Such varied environments and architectures have lent themselves only to partial solutions, so that there is no one general solution for CID across the DoD community.

The Joint Combat Identification Advanced Concept Technology Demonstration office partitioned target determination into four basic CID system concepts.  A majority of CID systems, if not all, fall into one of the four system concepts.

- Some systems align a sensor with the weapon sight.  The sensor interrogates the target.  A reply from the target identifies it as friendly, otherwise the target is unknown.

- "Don't shoot me" systems use Global Positioning Systems (GPS) or location systems. A weapon system sends out an interrogation in all directions containing the targeted position. Friendly systems in the area return a "don't shoot me" response.

- Situational awareness systems receive periodic position updates from friendly forces. From these updates, C2 systems are able to de-conflict friendly fire.

- Non-cooperative target recognition systems find a signature in acoustic signals, thermal and electromagnetic emissions and other data sources. The signature is then compared to a database to determine if the signature is indicative of a hostile, friendly or neutral target (Garamone, 2003).

Air Force CID systems predominately use situational awareness and non-cooperative target recognition concepts when identifying air-to-air and air-to-ground targets. A few examples will illustrate the diversity of applications these concepts have. The AWACS uses the situational awareness concept when its radar tracks friendly aircraft through its airspace. An Unattended Ground Sensor (UGS), on the other hand, uses target signatures to identify vehicles on the ground. Some systems use more than one concept like the Joint Surveillance Target Attack Radar System (JSTARS). It tracks friendly forces while at the same time using target recognition to identify nearby hostile forces. Fighters of all kinds use IFF systems. Finally, Intelligence, Reconnaissance and Surveillance (ISR) platforms, like the U-2, use a variety of Electro-Optical (EO), Infrared (IR) and radar sensors among others on a single platform to perform non-cooperative target recognition. Such an array of sensors on so many aircraft requires a focal point to fuse the sensor's data, to analyze the intelligence, to identify contacts as hostile or friendly, to form a situational awareness picture and, ultimately, to direct air operations. This focal point is called an Air Operations Center (AOC).

The AOC is the weapon system by which the Joint Forces Air Component Command (JFACC) commands and controls aerospace forces in a theater of operations (AOC CONOPS, 2001). Within the AOC there is one cycle, or process, for finding and

prosecuting targets. That cycle is Find, Fix, Track, Target, Engage and Assess (F2T2EA). Briefly, all sensor data, or selected sensor data, is sent through the AOC and potential targets are found. The potential target is then identified as hostile or friendly and located. Surveillance assets are then used to track targets until action can be taken. Targeteers in the AOC determine the target's priority per JFACC guidance and the appropriate action. If targeteers assign a weapons platform to attack the target, then the weapons system engages the target. Finally, after the engagement Battle Damage Assessment (BDA) is performed on the target to determine if it was destroyed.

Combat identification functions occur in the "Fix" cycle step. In this step, sensor data has already arrived at the AOC indicating that there is an unknown contact. In some instances, such as IFF, the sensor will give decision makers a positive or false identification of the target. In other cases, such as U-2 imagery, an analyst will be required to make a determination. In either case, analysts may fuse the intelligence from the sensors with other intelligence sources and the AOC's situational awareness. The analyst may also use automated target recognition and target cueing tools to help them identify targets. After the initial analysis, an analyst can declare the target hostile, friendly, or unknown. If the target is still unknown after the initial analysis, intelligence collection managers can cross-cue a different sensor to the same target to take advantage of complementary sensors and increase the analyst's target identification confidence. The result is that the intelligence analyst, the decision maker, gives the commander target identification.

In the case of Time Critical Targeting (TCT), intelligence analysts are required to make a target determination in minutes. The AOC CONOPS looks to information technologies to improve analyst's ability to analyze sensor intelligence and confidently identify. Furthermore, the document suggests that "information technology could provide the decision-making tools, decision support systems, and simulations to enable

4

commanders to make better and quicker decisions" (AOC CONOPS, 2001). These decision tools, decision support systems and simulations are all based on statistics and probability.

The Air Force has tasked Air Combat Command (ACC) to study various combinations of combat identification sensors through modeling, simulation and analysis. Among the various efforts, ACC has supported basic research at the Air Force Institute of Technology (AFIT) in sensor fusion. Last year, Capt Storm performed research for ACC entitled "An Investigation of the Effects of Correlation in Sensor Fusion" (Storm, 2003). Her research encompassed three fusion methods, the Identification Operating System Characteristic (ISOC) method, the Receiver Operating Characteristic (ROC) method and Neural Fusion; an example of which is fusion via Probabilistic Neural Networks (PNN). This thesis builds upon her research.

**Problem Statement**

Decision makers, or analysts, are required to declare a target as friendly or hostile using the available sensor data. If a determination can not be made, the decision makers must decide which additional sensor(s) to task to improve their probability of making a correct identification. Additional sensor taskings are usually based on the decision maker's prior experience with an expectation of improving their probability of correct target identification. However, decision makers can not prove which sensor, or sensor ensemble, has the best probability of correctly identifying the target nor quantify by how much that probability will improve.

**Research Objective**

This research seeks to determine the optimal sensor ensemble and fusion technique combination across differing prior correlation distributions. To support this objective, the research will develop both an implementation methodology to perform three-classifier fusion and a reasonable optimality criterion to measure ensemble performance. Secondly, an empirical study will be conducted using the proposed methodology. Lastly, the research will test the viability of creating posterior probabilities from a modified radial basis function neural network.

*Assumptions/Limitations.*

Sensor data was not readily available for this effort; hence, sensors' feature data was simulated using a Matlab program.

*Terminology.*

During the course of the research it became apparent that many terms in the operational world are synonymous with terms in the statistical world. Often writers used these terms interchangeably. Encapsulated here are some synonymous terms to help the reader. Sensors create feature data. A set of feature data comes from each sensor and because each data set is associated with one sensor sometimes the terms are used interchangeably. Automatic Target Recognition (ATR) software or human operators examine the feature data looking for targets. In statistical terms, the software and operators are considered classifiers of the data. Here again the operators will identify the targets, while statisticians classify exemplars. A hostile target misidentified as a friend is called a "leaker" in operations. Misidentifying a friend as a hostile does not have a

specific term but leads to fratricide.  Statistically, a hostile classified as hostile is called a true positive.  A friendly classified as a hostile is called a false positive.

   *Implications.*

   If expected sensor identification accuracies can be defined and quantified, decision makers will be able to make more informed decisions regarding sensors taskings. This will lead efficient use of high-demand/low-density ISR assets.  More accurate, timely decisions can be made regarding target identification.

**Introduction**

       This chapter reviews literature relevant to this research. It begins with a brief overview of the sensor fusion process to provide the context in which the literature will be applied. The rest of the chapter then discusses the literature in the order in which it pertains to the fusion process. The fusion process begins with the data. The data's statistical dependence most affects this research, so a discussion of the statistical dependence is included. Next, a review of posterior probabilities and neural networks leads to a discussion of the modifications made to the radial basis function neural network classifier. The fusion methods follow the classifiers in the fusion process, so a description of the three fusion methods, the ISOC, ROC and PNN, comes next. Lastly, a review of methods used to measure classifier and/or fusion performance completes the literature review.

**Fusion Process Overview**

       The objective of the fusion methods is to yield a better classification than the single best classifier alone. Training data is used to train the individual classifiers. The classifiers then classify two other data sets called, fusion training and testing. The result of this classification is two sets of posterior probabilities: fusion training posterior probabilities and testing posterior probabilities. The two sets of posterior probabilities are then sent to the fusion methods. The fusion methods train with the fusion training posterior probabilities and are tested with the testing posterior probabilities. All experimental designs used in this research are derived from this basic construct.
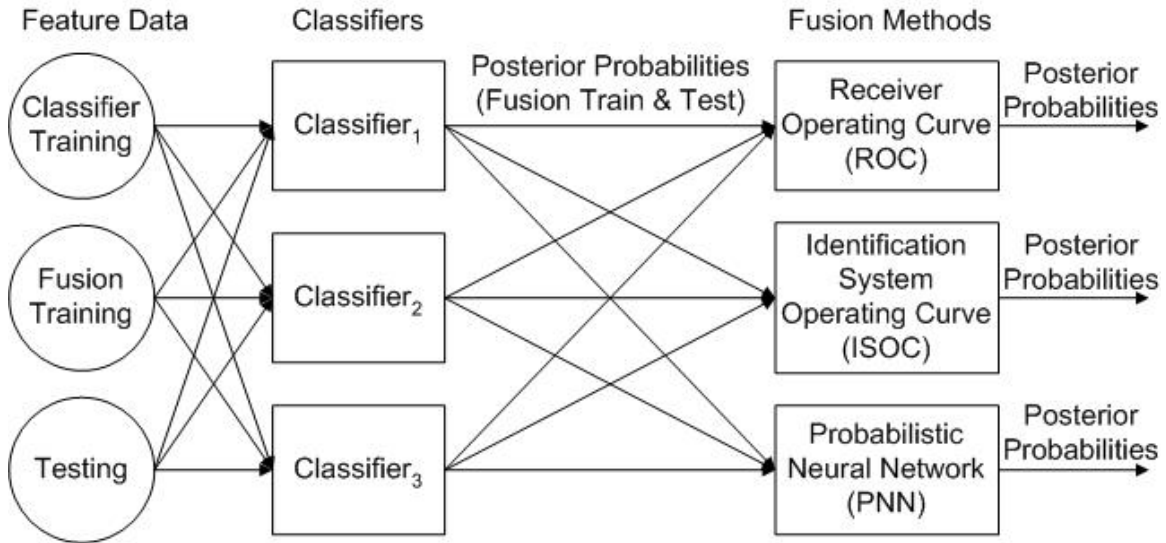
Figure 1: Fusion Overview

**Statistical Independence**

Both the ISOC and the ROC methods assume that the data coming from the

sensors are independent (Haspert, 2000; Oxley and Bauer, 2002). Statistically

independent data provides more new information than dependent data. A fusion method

should classify test data better given more training data. Some have attempted to find a

fusion method rule to apply to statistically dependent data. One attempt used two

classifiers on a bivariate Gaussian surface the simplest fusion case possible (Willet,

2000). Three fusion rules were applied logical "AND", "OR" and "XOR" to three

partitions of a Gaussian mean-shifted space. The "AND" rule could always classify data

with one threshold in one partition, never classify the data with one threshold in the

second partition and no consistent classification could be determined in the third partition

(Willet, 2000). Basically, even the simplest problem failed to yield a consistent fusion

9

rule for correlated data. With that in mind, this research seeks to determine how resilient ISOC and ROC fusion rules are to correlated data.

**Discriminant Classifiers**

Given a new exemplar and the probability density functions of multiple classes, discriminant analysis compares the ratio of the probability density functions at the new exemplar. The ratio is then used to determine which class the exemplar belongs to. In this particular experiment, the classes have a bivariate normal distribution so that the discriminant calculates the ratio of the class one probability density function to class two probability density function. The following equation represents the multivariate probability density function:

$$f_i(X_0) = \frac{1}{(2\pi)^{(p/2)}|\Sigma|^{(1/2)}} \exp\left[-\left(\frac{1}{2}\right)(X_0 - \mu_i)'\Sigma^{-1}(X_0 - \mu_i)\right]$$

where $X_0$ represents a new exemplar, $\mu_i$ represents the mean of class $i$ and $\Sigma$ represents the pooled covariance of the classes. For this research both classes are normally distributed and have equal covariance matrices. Bayes' rule is applied to the disciminant analysis to produce posterior probabilities. The following equation then produces the posterior probability for each class.

$$P(\pi_j \mid X_0) = \frac{\exp[-(1/2)*(X_0 - \mu_j)'\Sigma^{-1}(X_0 - \mu_j)]}{\sum_{i=1}^{K}\exp[-(1/2)*(X_0 - \mu_i)'\Sigma^{-1}(X_0 - \mu_i)]}$$

The quadratic discriminant operates exactly like the linear discriminant except that it does not assume equal covariance matrices among the classes. By using normal distributions

and applying Bayes' rule, we arrive at different equation which produces the posterior

probability for each class.

$$P(\pi_j \mid X_0) = \frac{\exp[(-1/2) * \ln|\Sigma_j| - (1/2) * (X_0 - \mu_j)'\Sigma_j^{-1}(X_0 - \mu_j)]}{\sum_{i=1}^{K} \exp[(-1/2) * \ln|\Sigma_i| - (1/2) * (X_0 - \mu_i)'\Sigma_i^{-1}(X_0 - \mu_i)]}$$

where $\Sigma_j$ represents the covariance matrix of class $j$. In this way, posterior probabilities
were created for the linear and quadratic discriminant classifiers.

**Modified Radial Basis Function Network**

The standard radial basis function network classifies new exemplars based on the

sum of their weighted distances to exemplars of known classes. This method outputs

weights for each class given a new exemplar. It does not produce posterior probabilities.

Since the fusion methods require posterior probabilities, the standard radial basis function

was modified.

It has been shown that the outputs of a multilayer perceptron network

approximate the *a posterior* probability function of the classes for any number of layers

and any type of activation functions (Ruck, et al., 1990). The network's backpropagation

achieves this by minimizing its mean squared-error approximation to the Bayes optimal

discriminant function (Ruck, et al., 1990). The mean squared-error approximation is

represented by the following equation:

$$\varepsilon^2(w) = \int_{\chi} [F(x, w) - g_0(x)]^2 \, p(x) dx$$

where $F(x, w)$ represents the perceptron output for new exemplar $x$ and weights $w$,

$g_0(x)$ represents the Bayes optimal discriminant function and $p(x)$ represents the density

function. Since a multi-layer perceptron network approximates the Bayes optimal

discriminant, it appears reasonable that a network utilizing a single perceptron output node and a non-perceptron hidden layer would exhibit some of the same qualities. So the modified radial basis function outputs are treated as posterior probabilities in this research.

The modified radial basis function architecture consists of a hidden layer of radial basis neurons and a single log-Sigmoid output node. The network uses *n* number of training exemplars each with *m* number of features to establish the weight vector for each neuron in the hidden layer, shown in Figure 2. The bias into hidden layer is established empirically by varying the neuron's spread. The relationship between spread and bias was as follows:

$$bias = 0.8326 / spread$$

Both the weight vector and the bias are used in the activation function for each radial basis neuron is as follows:

$$h_i = e^{-(\|w-p\|*b)^2}$$

where $w - p$ represents the distance between the neuron's weight vector and the new exemplar's feature vector and $b$ represents the neuron's bias.

The output node utilizes a Log-Sigmoid transfer function. The transfer function requires weighted inputs and a bias to allow for training, so weights and bias are initialized the Nguyen-Widrow initialization method. The weighted activations and the bias are then summed at the node and sent through the Log-Sigmoid transfer function. The function takes their sum and maps the output to a [0,1] range.

$$\log sig(n) = 1/(1 + e^{-n})$$

The weights and bias are adjusted using updates from a gradient descent momentum method and an adaptive learning rate. The output yields activations for class 1 only. Class 0 would be its complement. Figure 2 illustrates the modified radial basis function architecture.
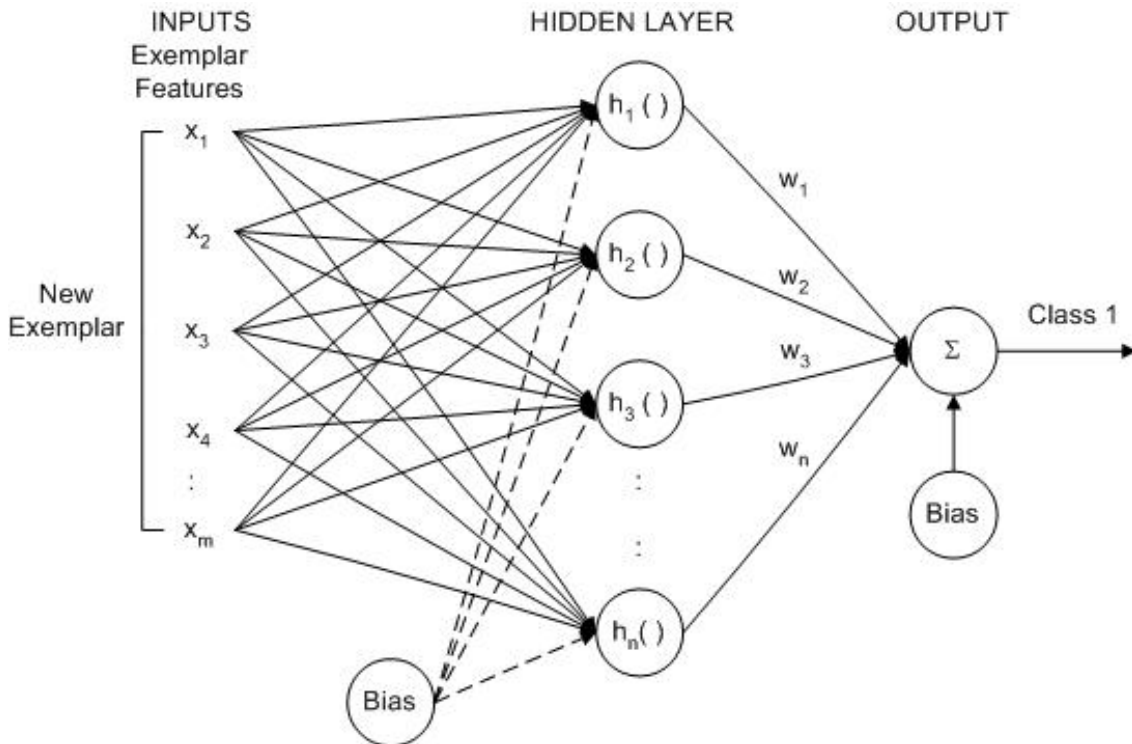


Figure 2: Modified Radial Basis Function Network Architecture

**General Regression Neural Networks**

General Regression Neural Networks subsume all other radial basis functions (Wasserman, 1993). In fact, the GRNN network topology is identical to a normalized radial basis function network (Wasserman, 1993) as shown in Figure 3. The GRNN assigns the target values as the weights. Because zero and one are used as the target values (class values), this has the effect of separating the two classes. The GRNN then

ignores any inputs from other classes until final probability calculation. The result is that the GRNN approaches an optimal estimator in the mean-squared-error sense (German, Bienenstock and Doursat, 1992). Further research by Richard and Lippman in 1991 has shown that if a classifier is optimal in the mean-squared-error sense then given enough data it will approach a Bayes optimal classifier. Finally, if the classifier approaches the Bayes optimal classifier then its output will very closely approximate the posterior probabilities. For purposes of this research, the posterior probabilities are desirable for fusion.
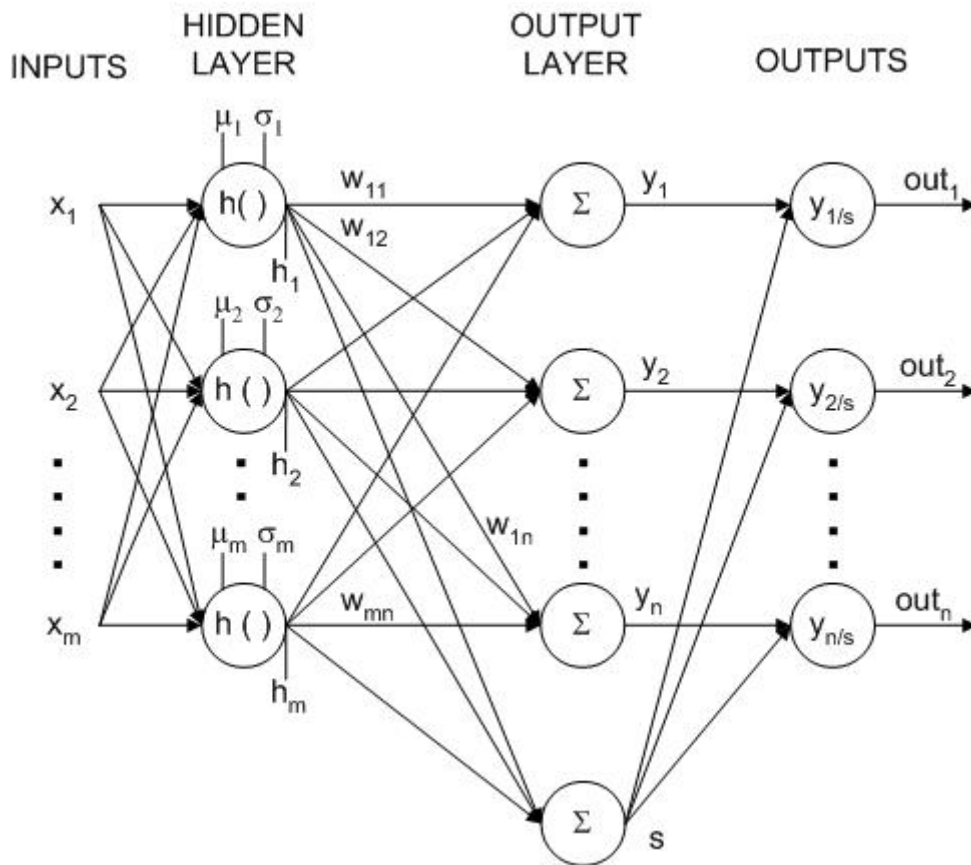


Figure 3: General Regression Neural Network

**Identification System Operating Characteristic Fusion Method**

The Identification System Operating Characteristic (ISOC) method provides an optimal sensor fusion rule to identify new targets.  The process to identify this optimal rule incorporates the ability to adapt the rule to the current environment.  Briefly, the method develops all possible boolean sensor fusion rules from a training data set, applies the costs and probability of misidentification to the rules and selects the optimal fusion rule based on the minimum cost.  If the targets encountered change, the training data set and, hence, the set of possible sensor fusion rules will change.  Likewise, if the cost and/or probability of misidentification changes then the set of all the possible rules remains the same but the optimal rule will change. This leads to a quantifiable adaptation of the sensor fusion rule to the anticipated targets and operating environments. Current adaptation of sensor fusion identification rules usually involves some subjective and often intuitive judgment or declared policy about the relative reliability or priority of different identification sensors and procedures, the consequences of making identification errors, the nature of the threat environment (Ralston, 1998).  The ISOC approach quantifies the key subjective factors to produce the optimal sensor fusion rule leading to the lowest total expected costs.  Figure 4 illustrates the ISOC process.
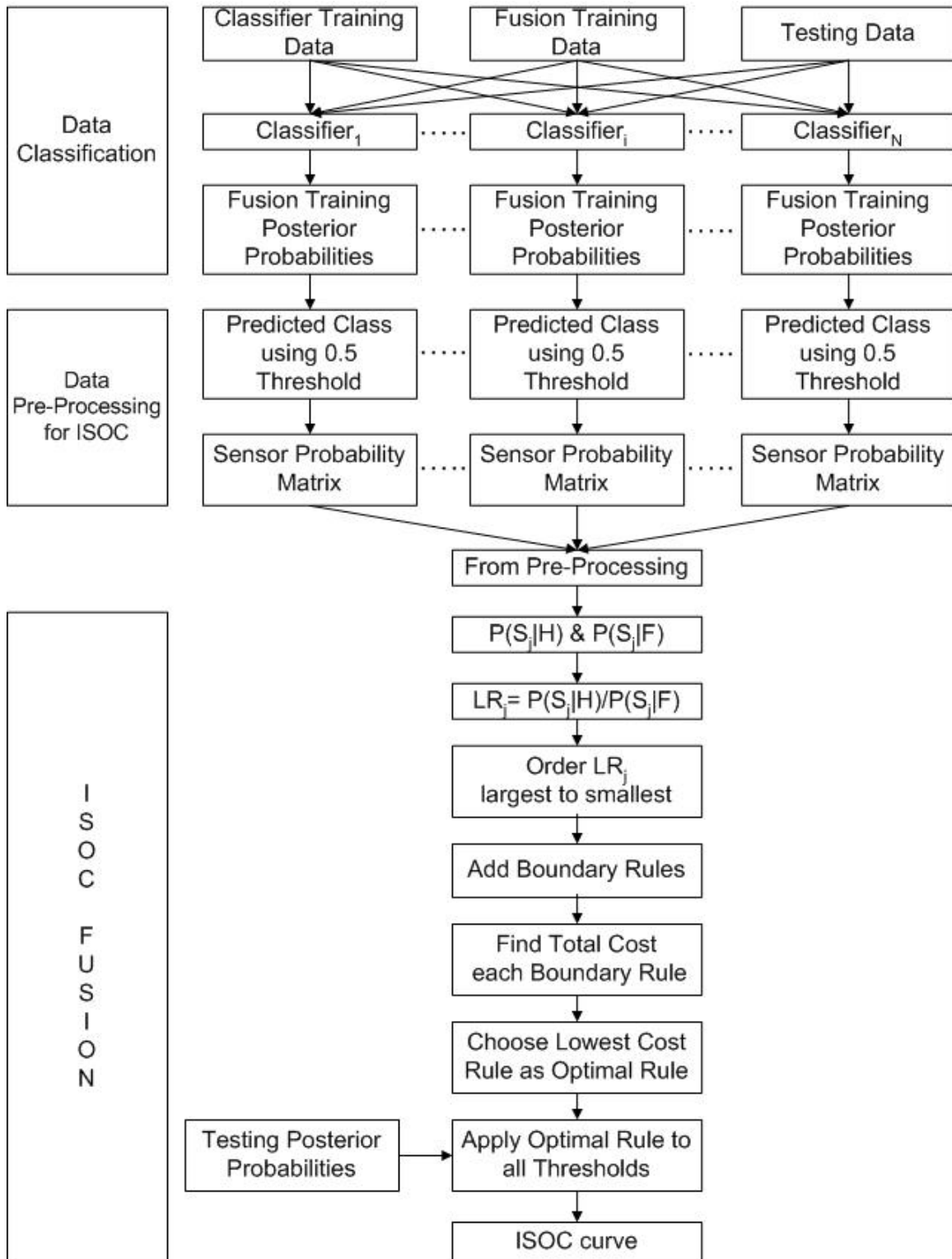
Figure 4: ISOC Fusion Process

*a. Sensor Performance Matrices*

The classification performance of each sensor can be expressed in a performance matrix similar to Table 1. Table 1 shows an example of a two class, two output state example. The matrix number of true classes and output states can be expanded if needed.

Table 1: Sensor Performance Matrix

| Output State | True Class | |
|---|---|---|
| | H | F |
| "H" | P("H"|H) | P("H"|F) |
| "F" | P("F"|H) | P("F"|F) |

*b. Combat Identification System States*

To develop the Combat Identification System (CIS) states, the system of sensors must utilize two separate indexing schemes, one for the sensors and one for each sensor's output states. The sensors' index scheme will be $1 \leq i \leq N_S$, where $N_S$ represents the total number of sensors and i represents a particular sensor in that system.

Table 2: Sensor Performance Matrix with Indices

| Sensor (i) | Friend | Hostile |
|---|---|---|
| output state($1_i$) | p($1_i$|F) | p($1_i$|H) |
| • • | • • | • • |
| output state($k_i$) | p($k_i$|F) | p($k_i$|H) |
| • • | • • | • • |
| output state($n_i$) | p($n_i$|F) | p($n_i$|H) |

The second index scheme for the individual sensor's output states will be $1 \leq k_i \leq n_i$, where $n_i$ represents the total number of sensor output states for the $i^{th}$ sensor and $k_i$ represents the specific output state of the $i^{th}$ sensor. Table 2 illustrates the indexing

schemes for the ith sensor given two target types.  The number of output states may be

different for each sensor.  However, the number of target types should remain the same

across all sensors.

The Combat Identification System (CIS) is the system of sensors used to identify

a particular target.  So, one combination of the sensors' output states defines one CIS

state (i.e., configuration).  Under this definition, each sensor can only assume one output

state for any given CIS state.   Let each CIS state be designated as $S_j$ and define j as the

index that runs over all N states of the system, then $1 \leq j \leq N$ (Ralston, 1998).  It will

form a vector where $S_j = (s_1^j, s_2^j, \ldots, s_2^j, \ldots, s_{Ns}^j)$.  So that $s_i^j \equiv$ sensor output state of the ith

sensor in the jth CIS state as illustrated in Table 3.

Table 3:  CIS States (Sensors' Output State Combinations)

| j | $S_j$ |
|---|---|
| 1 | $(s_1^1, s_2^1, \ldots, s_i^1, \ldots, s_{Ns}^1)$ |
| : | : |
| j | $(s_1^j, s_2^j, \ldots, s_i^j, \ldots, s_{Ns}^j)$ |
| : | : |
| N | $(s_1^N, s_2^N, \ldots, s_i^N, \ldots, s_{Ns}^N)$ |

There will be N distinct configurations of the overall CIS given by (Ralston, 1998)

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

Assuming the sensors are independent, the probability of each CIS state given a

target type can be found by multiplying the probabilities of each sensor's output state

given the same target type (T):

18

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

*c.   Identification Fusion Rules*

Some sensors' output states within any particular CIS state will inevitably yield conflicting identifications.  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).

Each fusion rule can be expressed as vector $R = (r_1, r_2, \dots, r_j, \dots, r_N)$, where j represents the CIS state index and $r_j \in \{0,1\}$ represents either the inclusion ($r_j = 1$) or exclusion ($r_j = 0$) of a particular CIS state in the fusion rule.  The total number of distinct possible fusion rules is $2^N$ (Ralston, 1998).  The probability that a particular fusion rule will correctly identify different target types can be found by multiplying the rule by the CIS states' conditional probabilities.  The following equation (and an alternate form) holds, where $P(h \mid H)$ represents the probability of classifying a target as "hostile" given that it is truly hostile.

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

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

Conversely, the following probability represents probability of classifying a friendly as a "hostile."

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

The conditional probabilities for each target type (i.e., Hostile, Friend, etc.) are found for

each fusion rule. The problem is to choose the fusion rule R to maximize declaring a

hostile as a hostile while minimizing declaring a friend as a hostile (Ralston, 1998).

Unfortunately, there are too many fusion rules to test them all. Figure 5 illustrates the

number of possible fusion rules resulting from nine sensors, which are $2^9$ or 512 fusion

rules.
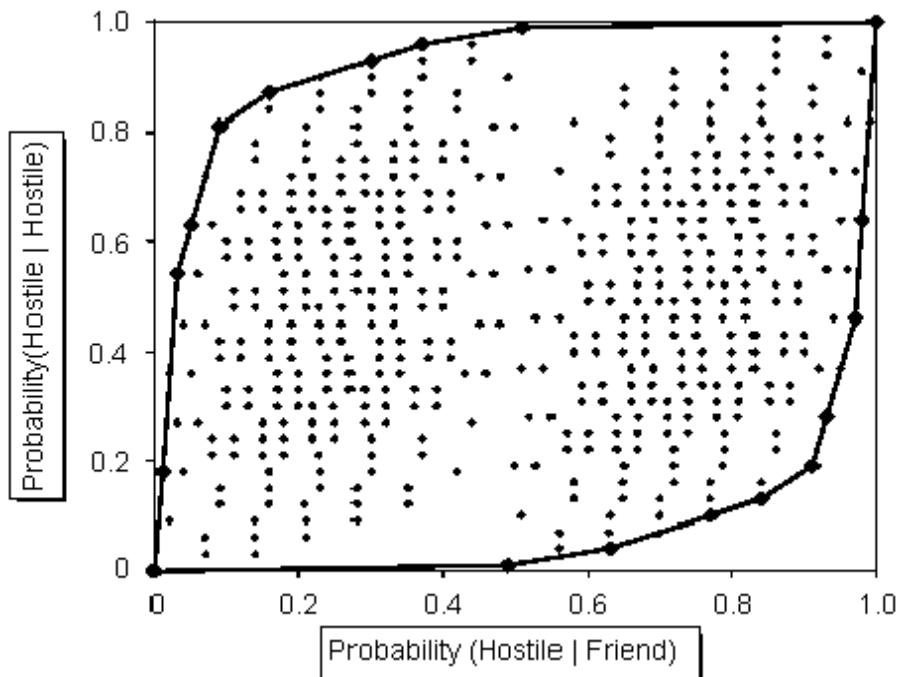


Figure 5: Possible Fusion Rules (Haspert, 2000)

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).  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).  In other words, the likelihood ratios P(j|H)/P(j|F) are ordered.  The plot of the cumulative probabilities of P(j|H) vs P(j|F) according to the likelihood ratio order yields an Identification System Operating Characteristic (ISOC).  As might be expected, each point on the ISOC is a complete identification fusion rule (Ralston, 1998).  However, no point (i.e., fusion rule) appears better than any other to serve as the best operating point.  To determine that, requires additional information about the anticipated ratio of encountering true-friends and true-hostiles in the theater of operations and about the costs of making identification errors of different kinds (Ralston, 1998).

### d.  Cost of Identification Errors

The costs of misidentification reflect a trade-off between the relative undesirability of allowing enemy leakers versus incurring fratricide of friendly platforms (Haspert, 2000).  This trade-off uses a cost function to select an ISOC operating point (i.e., fusion rule) that minimizes cost.  The cost function is as follows:

$$C_T = C_{FN} * P_H * P_{FN} + C_{FP} * P_F * P_{FP}$$

where        $C_T$ =  expected cost of misidentification

$C_{FN}$ = cost of not identifying hostile as hostile (e.g., a potential leaker)

$C_{FP}$ = cost of declaring a friend as hostile (e.g., fractricide)

$P_H$ = *a priori* probability of hostile

$P_F$ = *a priori* probability of friend

$P_{FN}$ = probability hostile not declared hostile

$$P_{FP} = \text{probability friend declared hostile}$$

where $P_{FN}$ and $P_{FP}$ can also be written as

$$P_{FN} = 1 - P_{TP}$$

$$P_{TP} = P(j \mid H)$$

$$P_{FP} = P(j \mid F)$$

The *a priori* probability of hostile and friendly are proportional to the relative number of hostile targets, $N_H$. Hence $P_H \sim N_H$ and $P_F \sim N_F$ (Haspert, 2000). The Command authority must make a subjective determination regarding the cost figures. The total cost of misidentification is calculated for each ISOC operating point. The operating point with the lowest total cost is determined to be the optimal operating point. The fusion rule associated with that operating point is declared to be the optimal sensor fusion identification rule.

**Receiver Operating Characteristic Fusion Method**

Whereas the ISOC method finds an optimal rule to fuse two or more classifiers, the Receiver Operating Curve (ROC) model finds the optimal thresholds needed in the individual classifiers to maintain optimal fusion performance for a fixed fusion rule (Storm, 2002). The ROC model discussed in this paper can also be called a ROC "within" model because two classifiers (sensors) are applied to the same feature set (Oxley and Bauer, 2002). The classifiers map the feature set into two different label sets. These label sets are then combined or fused via optimal thresholds and the logical "or" rule into a single system label set. Figure 6 illustrates the ROC "within" fusion process.

| Data Classification | Classifier Training Data | Fusion Training Data | Testing Data |
|---|---|---|---|

Classifier$_1$ ...... Classifier$_i$ ...... Classifier$_N$

**Data Classification**

Fusion Training Posterior Probabilities — Fusion Training Posterior Probabilities ...... Fusion Training Posterior Probabilities

**Data Processing for ROC "within"**

Predicted Class For Every Threshold 0 - 1 — Predicted Class For Every Threshold 0 - 1 ...... Predicted Class For Every Threshold 0 - 1

Sensor Probability Matrix for every Threshold — Sensor Probability Matrix for every Threshold ...... Sensor Probability Matrix for every Threshold

ROC Curve — ROC Curve ...... ROC Curve

From Pre-Processing — From Pre-Processing — From Pre-Processing

**ROC "WITHIN" FUSION**

Fusion

Vary "r" 0 - 1

Max
$P_{TP}{}^A + P_{TP}{}^B - P_{TP}{}^A * P_{TP}{}^B$
s.t. $0 < p < r$
$r = p + q - p*q$

Find $p^*, \theta^*, q^*, \phi^*$

Create ROC curve w/$\theta^*, \phi^*$ and "OR" rule

Fusion

Fuse $(i-1)^{th}$ classifiers' ROC w/ $i^{th}$ ROC curve

Fusion

Fuse $(N-1)^{th}$ classifiers' ROC w/$N^{th}$ ROC curve

Testing Posterior Probabilities

Apply Optimal Thresholds & Logical "OR" Rule

ROC Curve

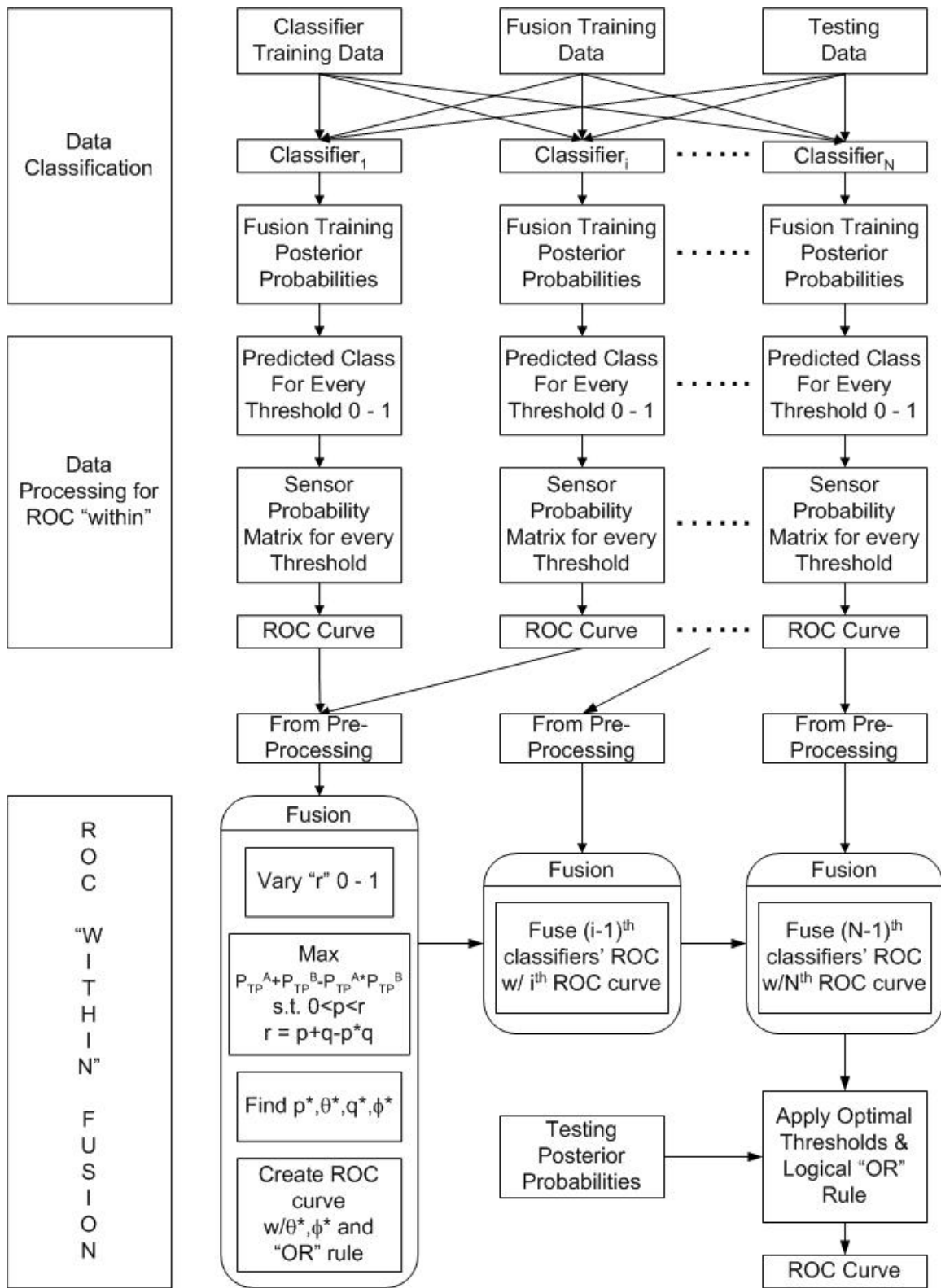Figure 6: ROC "Within" Fusion Process

*e. Classifier Parameters*

Let two classifiers $A_\theta$ and $B_\phi$, where $\theta \in \Theta$ and $\phi \in \Phi$ are the two parameter sets,

act upon a set of feature vectors (i.e., exemplars) and label each set of features according

to a label set. In this case, the label set $L = \{ L_h , L_f)$, where $L_h$ means hostile label and $L_f$

means friend label. Further, let X be the complete set of feature vectors so that $X_h$

represents the true set of feature vectors representing the hostile class and $X_f$ represent the

true set of feature vectors representing the friend class. The normal definitions of true

positive (TP), false positive (FP), true negative (TN) and false negative (FN) are

illustrated in Table 4. The probabilities of these conditions for each classifier are

determined.

Table 4: Class Definitions

|  | True Class | |
|---|---|---|
| Classified as: | **H** | **F** |
| "H" | TP | FP |
| "F" | FN | TN |

Let $P^A_{TP}$ represent the probability of classifier $A_\theta$ correctly labeling a hostile exemplar,

true positive. The following defines $P^A_{TP}$ as well as the other possible probabilities.

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

The definitions for $B_\phi$ are similar (Clutz, 2002). This concept along with a varied

threshold $(\theta)$ generates a ROC curve. One such threshold could be the probability level

that determines if a set of features is hostile or friendly. For example, if the threshold is

0.5 and the set of features has a 0.6 probability of being hostile, then the set of features is

declared hostile. As the threshold varies from zero to one, the probability of TP and FP change. The plot of the FP vs TP as the threshold changes defines the ROC curve, so that each point on the curve represents a threshold.

Once the hostile and friendly probabilities are determined for each classifier they will be combined for all possible label sets. A conditional probability table for classifiers $A_\theta$ and $B_\phi$ is given in Table 5.

Table 5: Conditional Probabilities for Two Classifiers

| | | (A$_\theta$,B$_\phi$) Reports as: | | | |
|---|---|---|---|---|---|
| | | **"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** | $P^A_{TP}P^B_{TP}$ | $P^A_{TP}P^B_{FN}$ | $P^A_{FN}P^B_{TP}$ | $P^A_{FN}P^B_{FN}$ |

*f. Classifier Fusion*

Allow $C_{\theta,\phi}$ to be defined as the concatenation of classifiers $A_\theta$ and $B_\phi$. As a result of the concatenation, $C_{\theta,\phi}$ yields a concatenated label from two labels $l_1$ and $l_2$. Some rule or method will be required to reconcile the two labels should there be any conflicts. The conditional probabilities in Table 5 are the starting point to resolving that conflict. The table can also be titled as the "Conditional Probabilities for $C_{\theta,\phi}$" as it already concatenates the two classifiers. Using the logical "or" rule to label hostile, $C_{\theta,\phi}$ will only be declared friendly when both classifiers are friendly. The general form of this equation is $P_{TP} = 1 - P_{FN}$. When applied to this case,

$$P^C_{TP} = 1 - P^C_{FN}$$

$$= 1 - (P^A_{FN})(P^B_{FN})$$

$$= 1 - (1 - P^A_{TP})(1 - P^B_{TP})$$

$$= 1 - (1 - P^A{}_{TP} - P^B{}_{TP} + (P^A{}_{TP})(P^B{}_{TP}))$$

$$= 1 - 1 + P^A{}_{TP} + P^B{}_{TP} - (P^A{}_{TP})(P^B{}_{TP})$$

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

Applying the same logic to the hostile label using the logical "or" rule, $C_{\theta,\phi}$ will be labeled hostile if both or either classifiers $A_\theta$ and $B_\phi$ are hostile. Since $P_{FP} = 1 - P_{TN}$, then $P^C{}_{FP} = 1 - P^C{}_{TN} = 1 - (P^A{}_{TN})(P^B{}_{TN})$. This results in $P^C{}_{FP} = P^A{}_{FP} + P^B{}_{FP} - (P^A{}_{FP})(P^B{}_{FP})$.

The maximum $P^C{}_{TP}$ associated with the each $P^C{}_{FP}$ value must now be found to develop an optimal ROC curve. Let $r$ represent $P^C{}_{FP}$, $p$ represent $P^A{}_{FP}$ and $q$ represent $P^B{}_{FP}$. So that $P^C{}_{FP} = P^A{}_{FP} + P^B{}_{FP} - (P^A{}_{FP})(P^B{}_{FP})$ can be re-stated as $r = p + q - (p)(q)$. Solving for $q$ yields $Q(p) = (r - p)/(1 - p)$. Now let $r$ vary across all false positive values, $r \in [0,1]$. For every $r$ value, let p vary so that $p \in [0,r]$. Then for every $p$ value calculate a corresponding $q$ value from the equation. The result is a $(p, q)$ vector for every $r$ value. We seek the optimal pair (p,q) that maximizes the $P^C{}_{TP}$ given by:

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

which becomes:

$$P^C{}_{TP}(r) = \underset{p \in [0,1]}{Max}[f_A(p) + f_B(Q(p)) - f_A(p)f_B(Q(p))]$$

The ROC curve of classifier A is a function and yields a true positive value (i.e., $f_A(p)$) for every false positive value $p$. Likewise, the ROC curve of classifier B is a function and yields a true positive value (i.e., $f_B(q)$) for every false positive value $q$. Since the value of $q$ is derived from the $p$ value, $f_B(q)$ is seen as the composition $f_B(Q(p))$ in the equation. In this way, each set of $(p,q)$ values generates a unique $P^C{}_{TP}$ value, therefore we get a function, $f_C(r)$. The pair maximizing the equation are denoted as $p^*$ and $q^*$ and

26

the associated point is denoted as $(r^*, f_C(r^*))$. As $r$ varies over its range, a complete set of fusion points are generated.

This process may also be described in terms of thresholds. Briefly, every point on a ROC curve represents a true positive and false positive pair. These declaration pairs are generated by comparing the original classifier posterior probabilities to different thresholds values. So each declaration pair is associated with a particular threshold.

**Probabilistic Neural Network Method**

The Probabilistic Neural Network (PNN) can be considered a deterministic network that approaches a Bayesian optimality given a large training data set (Wasserman, 1993). Some other advantages include the network's instantaneous training and its robustness to noise (Wasserman, 1993). Its main drawback is that the hidden layer is proportional to the number training exemplars. So the hidden layer's activation calculations may become excessively large.
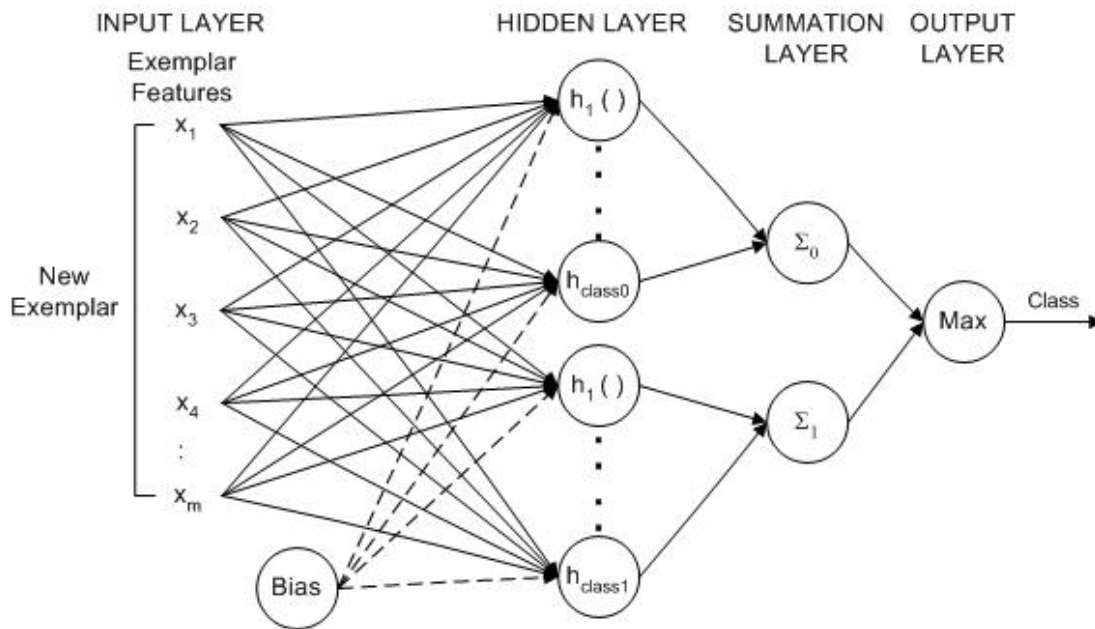


Figure 7: Probabilistic Neural Network Architecture

In this implementation, as shown in Figure 7, the PNN receives fusion training posterior probabilities to establish the network and testing posterior probabilities to test its classification abilities. The method described here assumes that both the training and testing posterior probabilities are normalized. This simplifies the Euclidean distance between the mean of radbas neurons and the new exemplar. Given a new exemplar $X = (x_1, x_2 ... x_m)$ and radbas neuron's weights $X_{Ri} = (x_{R1}, x_{R2} ... x_{Rm})$, the activation out of the radbas neuron can be represented as:

$$Z_{ci} = \exp[-(\|X - X_{Ri}\| * bias)^2]$$

Here $c$ represents the class and $i$ represents the pattern layer neuron. The implication is that the training data groups the hidden layer neurons into classes. The bias to the hidden layer is calculated as follows, where the spread is determined in the experiments.

$$bias = 0.8326 / spread$$

The summation layer simply sums the activations associated with a given class (Wasserman, 53). It can be represented as follows:

$$S_c = \sum_{i=1} \exp[-(\|X - X_{Ri}\| * bias)^2]$$

The output layer compares the sums of activations from the various classes. The class having maximum value receives a one; the other classes receive a zero. Figure 8 illustrates the PNN process. Since the PNN outperforms the other fusion methods, it trains only two-thirds of the fusion training posterior probabilities and tests on the remaining one-third of the fusion training posterior probabilities.
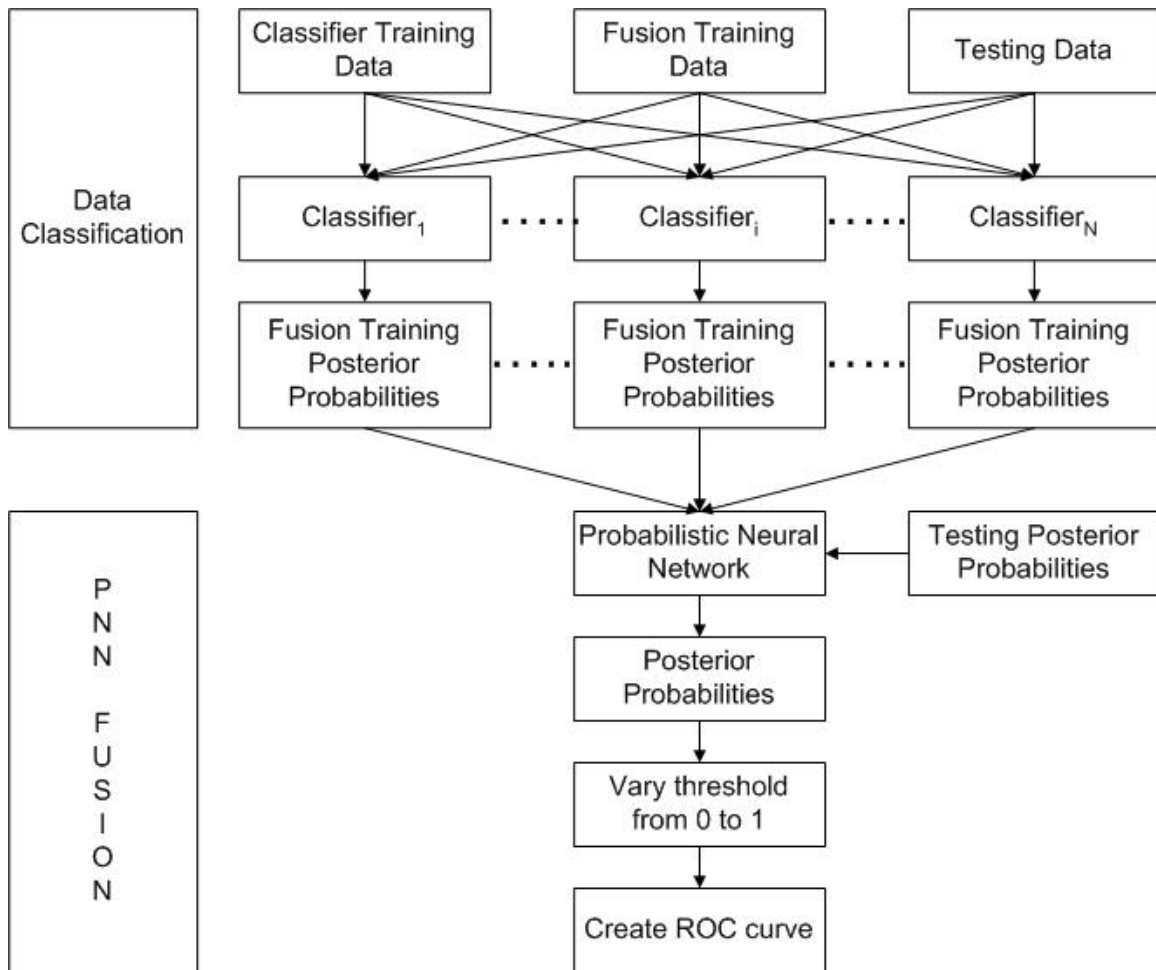
Figure 8: Probabilistic Neural Network Fusion Process

**Classifier Fusion Considerations**

There are two different approaches to combining a set of classifiers. The first approach selects a classifier from the set that is an "expert" in a particular local area of the feature space (Kuncheva, 2001). Feature vectors drawn from this local area are classified by the "expert" classifier. Sometimes more than one "local expert" can be used for different local areas of the feature space. The second approach, classifiers fusion, assumes that all classifiers are trained over the whole feature space, and are thereby considered as competitive rather than complementary (Kuncheva, 2001).

Fusion is useful only if the combined classifiers are mutually complementary, that is, classifiers should make different classification errors over the feature space (Roili, 2002). So, one tries to choose classifiers that are optimal in different regions of the feature space (Roili, 2002). Given the relationship between optimality and classification error, some have considered just the classifier ensemble's error diversity over the feature space. Unfortunately, while this is intuitive, there is some evidence that suggests that the diversity of classifiers does not affect the overall accuracies of the combination methods and their improvement over the single best classifier (Kuncheva, 2002). An experiment using ten different measures of classifier diversity and ten different combination methods found very little evidence of any correlation between the measures of diversity and the classifier combination performance (Kuncheva, 2002).

**Performance Measurements**

There are several methods of measurement used to determine classifier performance. ROC curves, for example, are commonly used for summarizing the performance in automatic target recognition when classification accuracy alone is not sufficient (Alsing, 2000). The ROC curve depicts the relationship between the detection rate (i.e., probability of true positive) and false alarm rate (i.e., probability of false positive) as a decision threshold is varied. The decision threshold usually varies between a high, or conservative, threshold where no targets are detected to a low, or aggressive, threshold where all targets are detected and all non-targets are labeled as targets. Figure 9 illustrates the probability density function of the two classes as they relate to the features. The thresholds are varied, for example, the illustration shows a threshold of 20 percent.

By determining the true positive and false positive rates for each threshold, we are able to construct the ROC curve, illustrated on the right.
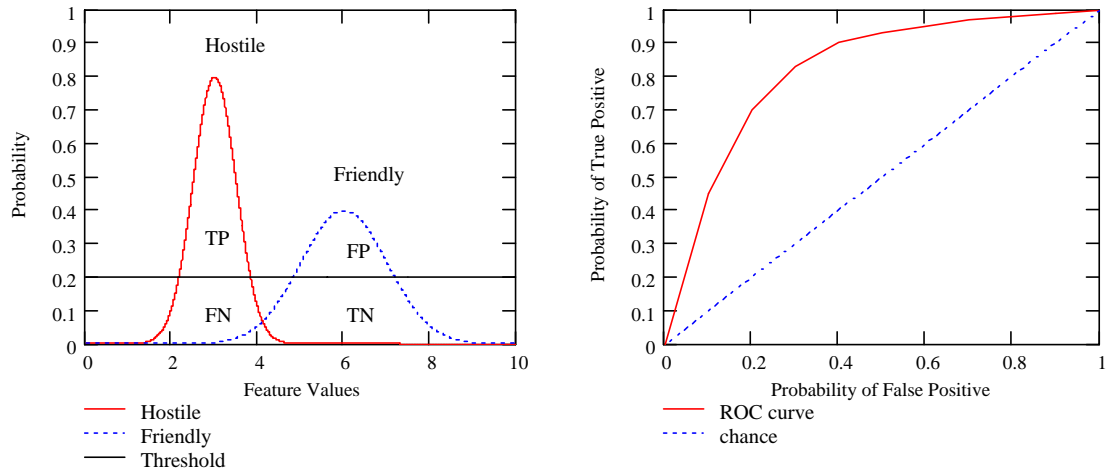


Figure 9: Receiver Operating Characteristics Curve

## III.    Methodology

**Overview**

This research consists of four classifier fusion experiments and one related excursion. The main distinction between the four experiments is that each one uses a different data set. The four data sets are a simple Gaussian, a XOR (close) Gaussian, a XOR (spread) Gaussian and a "domino" Gaussian. The fusion process applied to each dataset remained same across all the experiments. The linear discriminant classifier, the quadratic discriminant classifier and the radial basis function were used as the classifiers. Then the ROC, ISOC and PNN methods were used to fuse all combinations of two and three classifiers. The ensemble results were then measured using two techniques: ROC curves and expected true positive classification values. While the ROC curves are described in the literature review, the expected true positive classification rates were devised as part of this research and are described in the experiments' methodology. Since the classifiers, fusion methods and measures of performance are identically implemented in each of the four experiments, their methodology is described only once in the first experiment. Since the data sets vary from experiment to experiment, the data sets will be described in the methodology for each experiment.

An assumption was made based upon related work that the modified radial basis network would approximate a Bayes optimal classifier closely enough that the output could be treated as a posterior probability. This excursion compares the results of the modified radial basis function to results from the GRNN given the same data. The GRNN has been proven to approximate the Bayes optimal classifier so that if the radial

basis function approximates the GRNN then the assumption is reasonable.  For clarity, a list of the experiments and excursion are as follows:

Table 6: List of Experimental Designs

| Experiment # | Experiment Description |
|---|---|
| 1 | Simple Gaussian Distribution (3 Good Classifiers) |
| 2 | XOR (close) Gaussian Distribution (No Good Classifiers) |
| 3 | XOR (spread) Gaussian Distribution (2 Good Classifiers) |
| 4 | Complex Gaussian Distribution (1 Good Classifier) |
| 5 | XOR Gaussian Distribtuion RBF vs GRNN |

**Experiment 1: Simple Gaussian Distribution**

*Data Generation*

A feature data set represents all the data collected by various sensors on the same targets.  Any given sensor only generates a couple features of that data set.  The classifiers represent the algorithm a sensor uses to determine the probability of a target in its feature data.  So data set features are broken out of the data set and sent to the appropriate classifier to generate posterior probabilities of a target.

Let $F_1$ represent the set of features from sensor 1 and let the aggregation of the feature sets be represented as $F = F_1 \times F_2 \times \cdots \times F_i \times \cdots \times F_N$, where $N$ represents the number of feature sets.  The terms "feature set" and "sensor" are used interchangeably because a particular feature set is associated with a particular sensor.  Each set of features

33

may contain any number of features. So let $f_j^{\ i}$ represent the number of features $j$ in feature set $F_i$. Table 1 illustrates the relationship of the exemplars, features and feature sets.

<p align="center">Table 7: Exemplars, Features and Feature Sets</p>

| Exemplars | $F_1$ | | ...... | $F_i$ | | ...... | $F_N$ | |
|---|---|---|---|---|---|---|---|---|
| | $f_1^{\ 1}$ | $f_2^{\ 1}$ | | $f_1^{\ i}$ | $f_2^{\ i}$ | | $f_1^{\ N}$ | $f_2^{\ N}$ |
| 1 | 0.199 | 0.923 | | 0.155 | 0.87 | | 0.029 | 0.211 |
| 2 | 0.626 | 0.76 | | 0.365 | 0.678 | | 0.401 | 0.492 |
| : | : | : | | : | : | | : | : |
| : | : | : | | : | : | | : | : |
| m | 0.327 | 0.648 | | 0.024 | 0.478 | | 0.878 | 0.695 |

Here we have assumed two features per sensor. If the features, $f_i$, are independent their correlation matrix reduces to an identity.

$$\Sigma_{F_i,F_i} = I$$

The data generation process creates several feature sets. A generalized correlation matrix would be constructed as follows with potential correlations within and between feature sets.

$$\Sigma = \begin{bmatrix} \Sigma_{F_1,F_1} & \cdots & \Sigma_{F_1,F_N} \\ \cdots & \Sigma_{F_i,F_i} & \cdots \\ \Sigma_{F_N,F_1} & \cdots & \Sigma_{F_N,F_N} \end{bmatrix}$$

Inter-correlation between features in different feature sets causes the two feature sets (sensors) to become dependent. Figure 10 below illustrates inter-correlation between feature sets.

Figure 10: Inter-correlation between Feature Sets

In this research, a single classifier is used to classify single feature set.  The posterior

probabilities between the classifiers reflect the statistical independence or dependence

between the feature sets.  Both the ROC and ISOC fusion methods assume that the

classifiers' posterior probabilities, and by extension the feature sets, are independent.  By

making the features sets dependent, we violate this assumption and can quantify the

sensitivity of the fusion methods to this assumption.

Some fusion methods fuse two feature sets and then fuse a third feature set.  The

correlation matrix must maintain the correct feature independence and correlation with

regard to its sub-matrices so that the fusion method works correctly.  Having defined the

features and their correlations, we now define the distributions of the two classes of

features.  Let $F^0$ represent the feature sets from class zero and $F^1$ represent the feature

sets from class one.  All feature sets for each class are generated at the same time using a

multivariate normal distribution.  The only difference between $F^0$ and $F^1$ is that the

mean of their distributions are different $\mu^0 \neq \mu^1$ so that $F^0 \sim N(\mu^0, \Sigma)$ and

$F^1 \sim N(\mu^1, \Sigma)$.  Once the features are generated for both classes, the features are

concatenated so that $F = F^0 \cup F^1$.  All exemplars are then randomized.  This data

35

generation process is repeated to provide the required three independent data sets. If

multiple correlation levels are required, this process generates three data sets for each

correlation level. The data is then presented to the first level classifiers.

For this particular experiment, the number of exemplars will vary so that

$N_{ex} \in \{25,50,100\}$. The feature sets have independent features; however, the features are

correlated between feature sets. The correlation will vary from 0 to 1 where

$\rho \in \{0,0.2,0.4,0.6,0.8,1.0\}$. Let $F_i$ represent the set of features, where $i \in \{1,2,3\}$. This

implies $F = F_1 \times F_2 \times F_3$ and the feature set space $F \subset \Re^6$. All feature sets have two

independent features. This leads to a correlation matrix within each feature set.

$$\Sigma_{F_1,F_1} = \Sigma_{F_2,F_2} = \Sigma_{F_3,F_3} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Equation shows a possible two feature set correlation matrix using feature set notation

and features. This structure must be preserved in a three feature set correlation matrix for

the fusion methods.

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

The three-feature set correlation matrix is presented here in terms of features.

$$\Sigma = \left[ \begin{array}{cccc|cc} 1 & 0 & 0 & \rho & 0 & \rho \\ 0 & 1 & \rho & 0 & \rho & 0 \\ 0 & \rho & 1 & 0 & \rho & 0 \\ \rho & 0 & 0 & 1 & 0 & \rho \\ \hline 0 & \rho & \rho & 0 & 1 & 0 \\ \rho & 0 & 0 & \rho & 0 & 1 \end{array} \right]$$

Figure 11:  Feature Correlation Matrix

It can be seen that $f_1^1$ and $f_2^1$ are statistically independent as are the other feature pairs.

Further, it can be seen that correlation is induced between feature sets, in that, $f_1^1$, $f_2^2$

and $f_2^3$ are correlated, as are $f_2^1$, $f_1^2$ and $f_1^3$. The feature correlation matrix, shown in

Figure 11, was used to generate data for both classes; however, the two classes used

different means. Let $\mu^0 = \{0,0,0,0,0,0\}$ be the mean for class 0 and $\mu^1 = \{1,1,1,1,1,1\}$ be

the mean for class 1. Further, let $F^0$ be class 0 data distributed as $F^0 \sim N(\mu^0, \Sigma)$ and

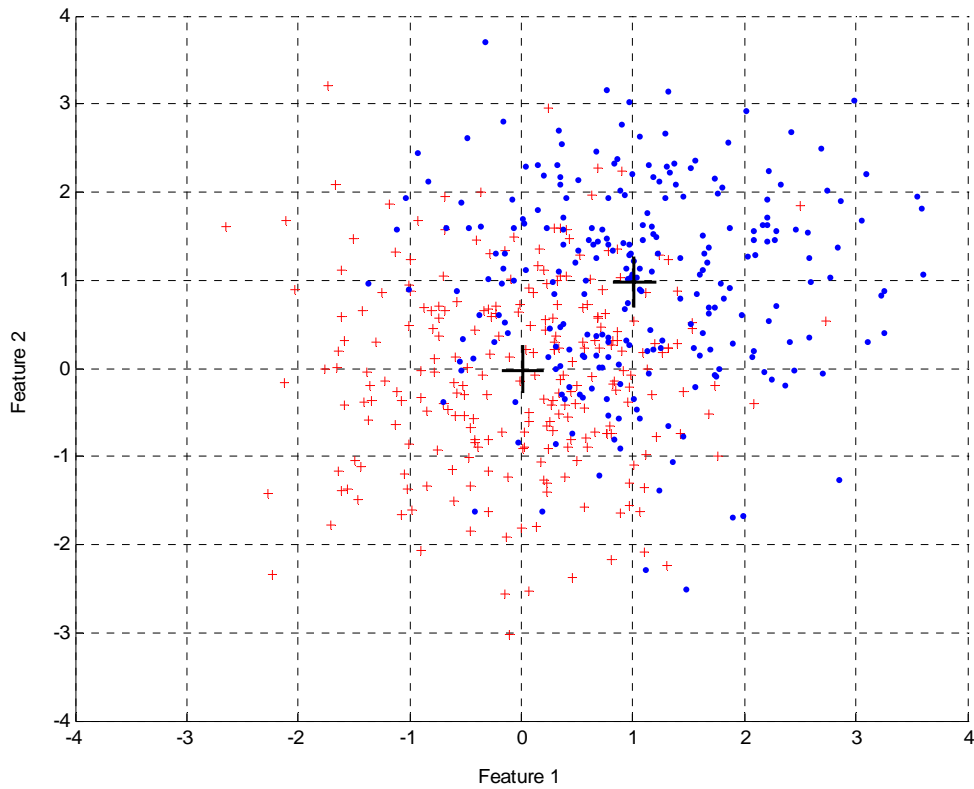$F^1$ be class 1 data distributed as $F^1 \sim N(\mu^1, \Sigma)$.



Figure 12: Two-Class Simple Gaussian Distributed Data

The two class data set are aggregated $F = F^0 \cup F^1$ and presented to the classifiers. There are forty repetitions of each sample size, so that data will be generated forty times for each sample size. Figure 12 illustrates the distribution of one feature set with two classes of data.

*Classifiers*

Three classifiers were used in this experiment. They were the linear discriminant, the quadratic discriminant and the modified radial basis function. Each classifier received its own feature set from the training data, fusion training data and the testing data. The data within the feature sets were independent. The feature sets themselves were correlated. The classifiers were trained with the training data's feature set. Then the classifiers assigned posterior probabilities to each exemplar in the fusion training data and testing data. These two sets of posterior probabilities were then sent to each of the three fusion methods.

*Fusion Methods*

Some decisions were made in applying the fusion techniques. The ROC method requires that the ROC curves from two classifiers be fused first to form a two-classifier fused ROC curve. This new ROC curve is then fused with the ROC curve from the third classifier which produces the final three-classifier fused ROC curve. In this research, the linear and quadratic discriminant classifiers were chosen as the first two classifiers and the modified radial basis function was chosen as the third classifier. The fusion training posterior probabilities were used to create the three-classifier fused ROC curve. The final ROC curve was then tested using the test posterior probabilities.

The other two fusion methods are scalable to accept greater or fewer number of classifiers so no decisions were required. The ISOC method required only the addition of the third classifier's posterior probabilities to the Combat Identification System (CIS) states. The PNN method only required additional hidden layer nodes. Once both methods had been trained with the fusion training posterior probabilities, they were tested using the test posterior probabilities.

*Optimal Ensemble / Fusion technique combination*

All ensemble / fusion technique combination classification performances were measured using two techniques. ROC curves are applied as described in the literature review. However, the second technique, expected true positive classification rates was devised for this research to serve as an optimality criterion.

The expected true positive classification rates were calculated for all ensemble / fusion technique combinations at the 0.1 false positive level. Each ensemble / fusion technique combination classified data correlated at six different levels. The features correlated at the different correlation levels could be assigned a distribution so that each correlation level would have a discrete probability density value. Two distributions were used in this research, a uniform and a linear distribution. Their values from 0 to 0.9 correlations are as follows:

$$\text{Uniform:} \quad p(x) = \{0.167, 0.167, 0.167, 0.167, 0.167, 0.167\}$$

$$\text{Linear:} \quad p(x) = \{0.05, 0.09, 0.14, 0.19, 0.24, 0.28\}$$

When the probability densities are applied to the true positive value of each correlation at the 0.1 false positive level, then the expected true positive classification rate can be calculated. The expected true positive classification rate is calculated for each ensemble /

fusion technique combination.  The ensemble / fusion technique combination yielding the highest classification rate according to this optimality criterion was then determined to be the best.  A mathematical description might appear as follows:

$$\underset{S \in R}{\text{ArgMax}} \;\; E_\rho(TP \mid S)$$
$$\text{s.t.}$$
$$FP(S) \leq 0.1$$

where, $\quad \overset{fp=0.1}{E_\rho}(TP) = \sum_{i=1}^{6} (TP \mid \rho = x_i) * P(\rho = x_i)$

and;      R represents the set of all possible single sensor and sensor fusions, where ISOC, ROC and PNN fusion techniques are applied to all subsets with 2 or more sensors.

**Experiment 2: XOR (close) Gaussian Distribution**
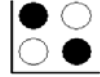
*Data Generation*

The data generated in experiment two required two distribution means per class. This resulted in four total distributions of the feature data in an XOR pattern.  The means used were as follows:

Class 0:  $\mu_1^0 = \{0,0,0,0,0,0\}$ and $\mu_2^0 = \{1,1,1,1,1,1\}$

Class 1: $\mu_1^0 = \{0,1,0,1,0,1\}$ and $\mu_2^0 = \{1,0,1,0,1,0\}$

The Gaussian distributions possess equal variances which are equal to one for all distributions.  Figure 13  represents the two feature data set to the linear classifier.
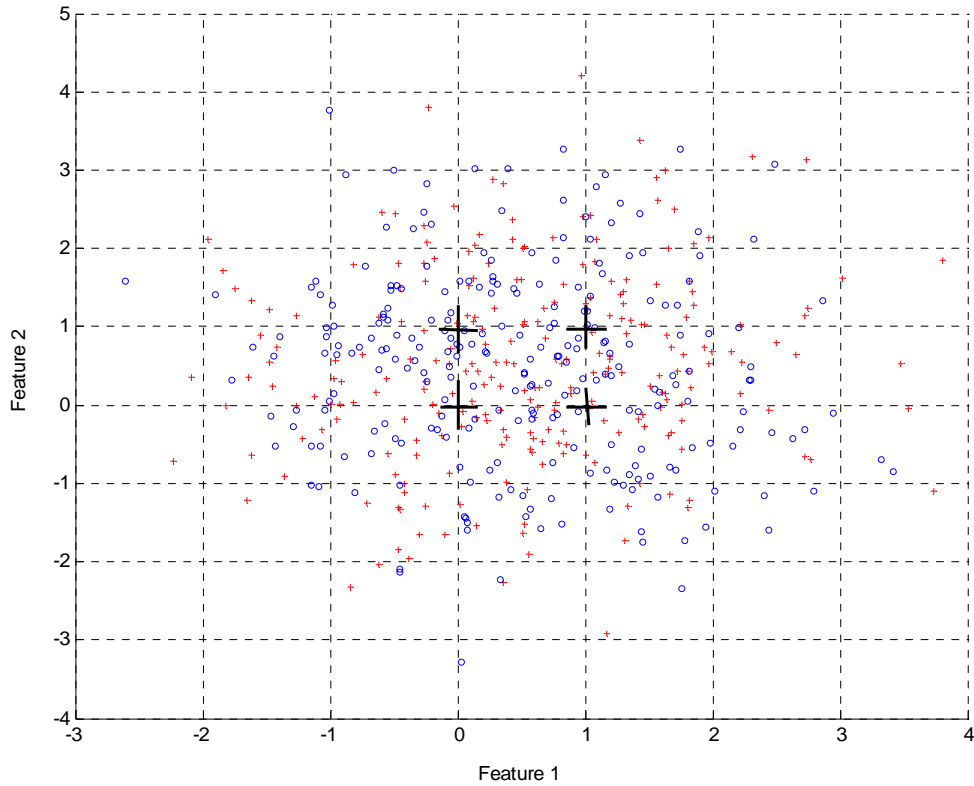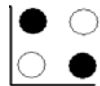
Figure 13: Two-Class XOR (close) Gaussian Distributed Data

## Experiment 3: XOR (spread) Gaussian Distribution

*Data Generation*

The data generated for the third experiment used same XOR pattern but increased the distance between the individual Gaussian distributions. This experiment has the same number of distributions as the second experiment.

Class 0: $\mu_1^{\,0} = \{0,0,0,0,0,0\}$ and $\mu_2^{\,0} = \{2.5,2.5,2.5,2.5,2.5,2.5\}$

Class 1: $\mu_1^{\,0} = \{0,2.5,0,2.5,0,2.5\}$ and $\mu_2^{\,0} = \{2.5,0,2.5,0,2.5,0\}$

Here again, Gaussian distributions possess equal variances which are equal to one for all distributions. Figure 14 represents the two feature data set to the linear classifier.
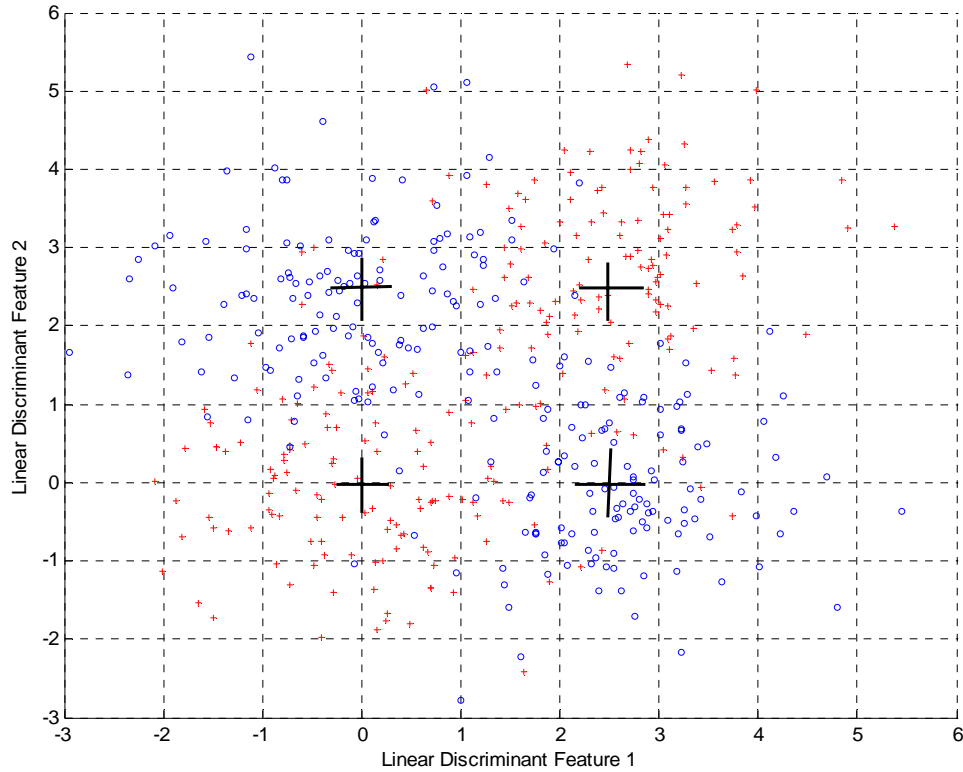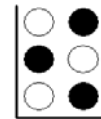


Figure 14: Two-Class XOR Gaussian Distributed Data

**Experiment 4: "Domino" Gaussian Distribution**

*Data Generation*

The data generated for the fourth experiment added a distribution to each class to change the pattern into a "domino." The means were as follows:

Class 0: $\mu_1^0 = \{0,0,0,0,0,0\}$, $\mu_2^0 = \{2.5,2.5,2.5,2.5,2.5,2.5\}$ and $\mu_2^0 = \{0,5,0,5,0,5\}$

Class 1: $\mu_1^1 = \{0,2.5,0,2.5,0,2.5\}$, $\mu_2^1 = \{2.5,0,2.5,0,2.5,0\}$ and $\mu_2^1 = \{2.5,5,2.5,5,2.5,5\}$

All variances were set equal to one for all distributions. Figure 15 represents the two feature data set to the linear classifier.
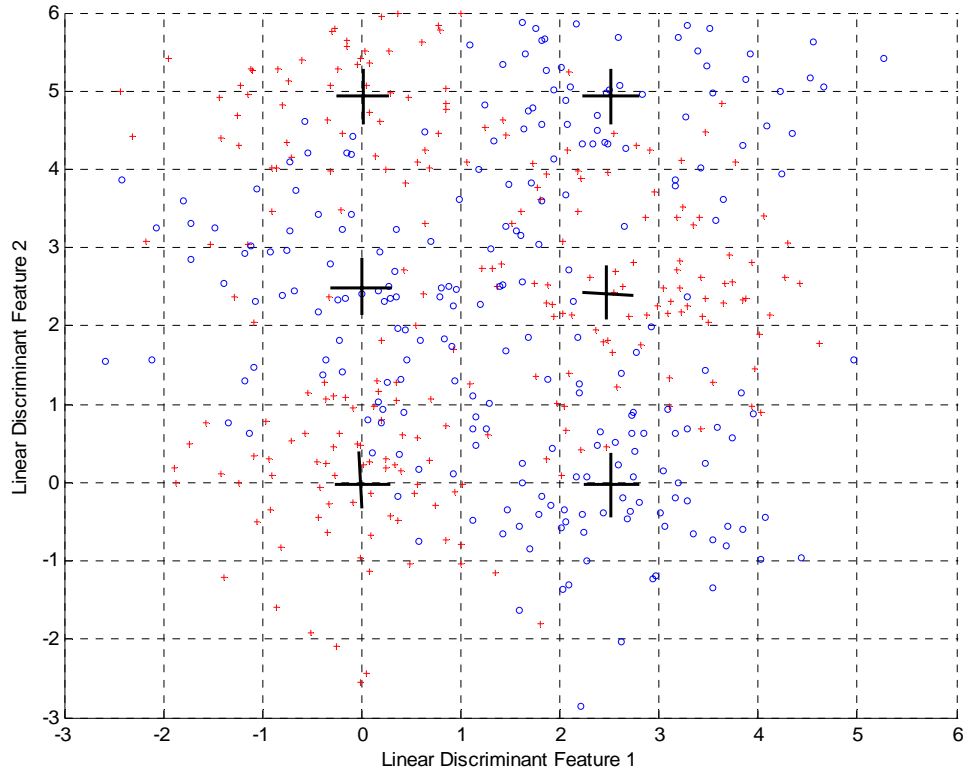


Figure 15: Two-Class "Domino" Gaussian Distributed Data

## Excursion 1: Radial basis function vs general regression neural network

*Data Generation*

Both the simple Gaussian distributed data and the XOR Gaussian distributed data were used for this experiment. However, the means used for the distributions were changed to allow for greater separation of the data and better classification results. For the simple Gaussian distributed data, let the class 0 have $\mu^0 = \{0,0,0,0,0,0\}$ and the class

43

1 have $\mu^1 = \{2.5,2.5,2.5,2.5,2.5,2.5\}$. For the XOR Gaussian distributed data, let class 0

have $\mu_1^0 = \{0,0,0,0,0,0\}$ and $\mu_2^0 = \{2.5,0,2.5,0,2.5,0\}$ and let class 1

have $\mu_1^1 = \{2.5,2.5,2.5,2.5,2.5,2.5\}$ and $\mu_1^1 = \{0,2.5,0,2.5,0,2.5\}$.

*Performance Measurement*

ROC curves generated from each classifiers results were used for comparison. Classification accuracy is not the primary concern in this comparison. The main concern is how well the radial basis function approximates the general regression neural network. If the ROC curves from each classifier approximate one another, then the posterior probabilities assigned approximate one another and the two classifiers are approximating similar mean-squared-errors.

## IV. Findings and Analysis

**Result 1: Two-Class Simple Gaussian Data Experiment**

The methodology was applied to the two-class simple Gaussian dataset. In analyzing the results, the finding will step through the ensembles' ROC curve performance. Then the analysis will compare the ensemble results using the expected true positive classification rates. The two separate performance measures were necessary. It is significant to note that while an ensemble may do well based on the overall ROC curve, it may perform poorly at the particular threshold of interest.

All single classifier ensembles had comparable performances. The data was reasonably separated so that the linear classifier was able to distinguish between the two classes.
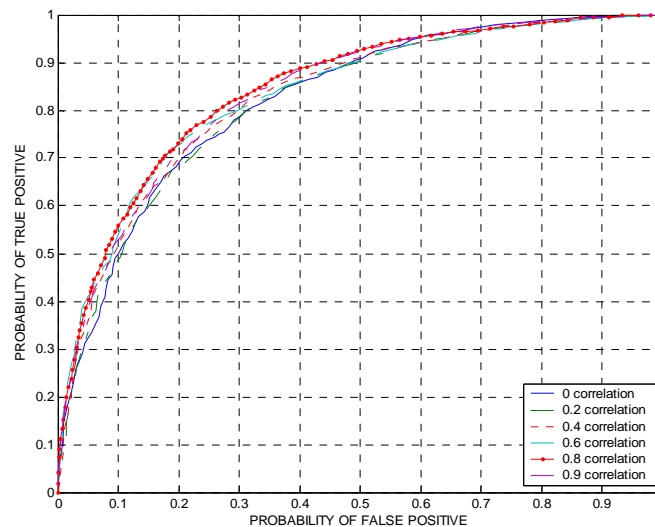


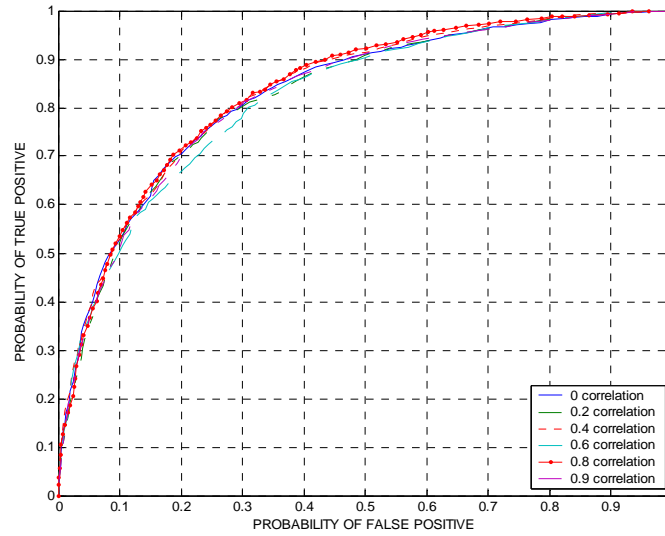Figure 16: Linear Classifier ROC Curve (SG)

Figure 17: Quadratic Classifier ROC Curve (SG)

Radial basis and quadratic classifiers were also able to distinguish between the classes. The quadratic classifier's performance is shown in Figure 17. The radial basis classifier's performance was slightly worse than the quadratic classifier.

The PNN two-classifier ensemble reflects the posterior probabilities of the fused classifiers. The linear-quadratic ensemble, having the two best classifiers, offers the best classification performance. The linear-radial and quadratic-radial ensembles have comparable, slightly degraded classifications performances since they include the worst classifier, the radial basis function neural network. The three-classifier ensemble performs slightly better than the linear-quadratic ensemble at the lower correlation levels but poorer at the higher levels of correlation. This can be attributed to the fact that the PNN uses posterior probabilities from all three classifiers and ignores no information. Lower correlation levels offer the PNN more data to train on so that the three-classifier ensemble offers more data than just two-classifier ensemble.
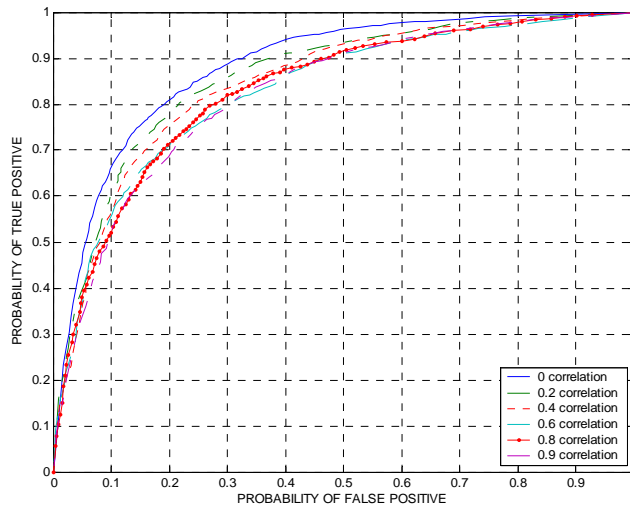
46

Figure 18: PNN Fusion of Linear & Quadratic Classifiers (SG)

At the higher correlation levels, the three-classifier ensemble performs worse than the two-classifier ensemble. This appears to be due to the idea that while three-classifier ensemble offers more data relative to the two-classifier ensemble, it also includes more errant classifications. In all cases, the PNN exhibits sensitivity to correlation levels.
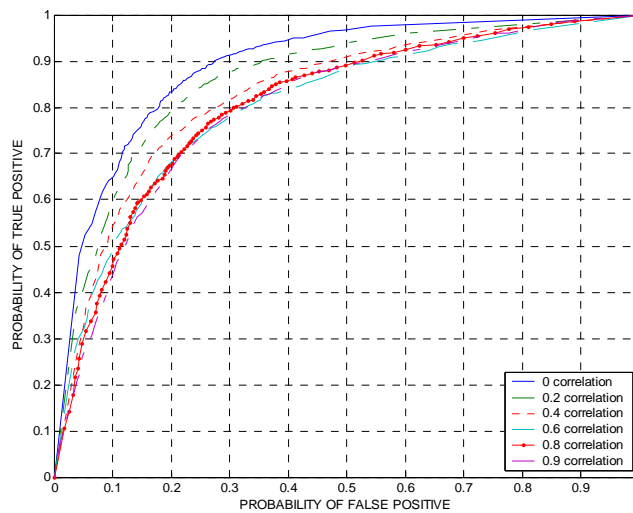


Figure 19: PNN Fusion of Linear, Quadratic & Radial Basis Classifiers (SG)

47

The ISOC two-classifier ensembles exhibited "robustness" to correlation. Two-classifier ensembles which included the radial basis classifier performed slightly worse.
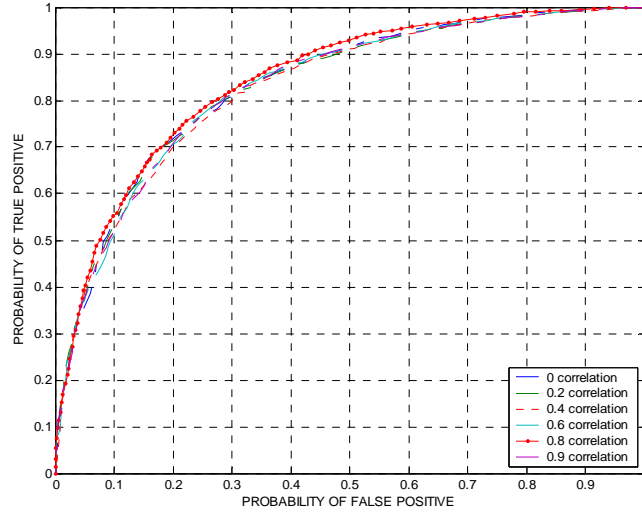


Figure 20: ISOC Fusion of Linear & Quadratic Classifiers (SG)

However, a ROC curve of the three-classifier ensemble provides remarkable separation between the correlation levels despite the fact that neither the single classifiers nor the two-classifier ensembles exhibited any separation.
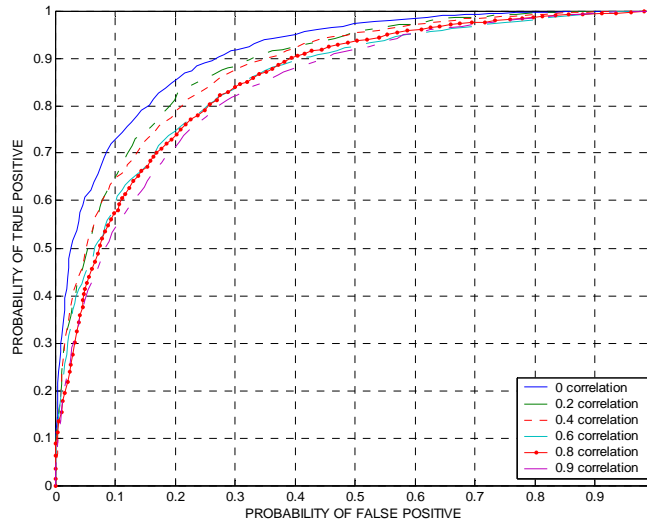


Figure 21: ISOC Fusion of Linear, Quadratic & Radial Basis Classifiers (SG)

Examination of the optimal ISOC rule for the ensemble revealed that the rule reduced to a majority voting method. For Table 8, shown below, the ones represent a "hostile" classification and zeros represent a "friendly" classification.

Table 8: Optimal ISOC Rule 3-Classifier Fusion

| Linear | Quadratic | Radial Basis | Optimal ISOC Rule |
|--------|-----------|--------------|-------------------|
| 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 |

Comparing the optimal ISOC rule with the ROC curves, it appears that the three classifiers are making slightly different errors. The majority voting method style-rule eliminates the errors a single classifier is making different from the other two classifiers. So, the result is that the fusion rule keeps only the errors common to all three classifiers and this increases the ensemble's performance. The separation of the correlation levels can be explained by realizing that correlation represents the amount of new information available to the ensemble. At the zero correlation level, the ensemble has the most new information available and yields the best classification performance. As the correlation increases the ensemble has less new information to classify new exemplars with and the ensemble's performance deteriorates.

The ROC two-classifier ensembles performed as expected. The linear-quadratic ensemble, having the two best classifiers, offered the best classification performance. The linear-radial and quadratic-radial ensembles have comparable, slightly degraded classifications performances since they include the worst classifier, the radial basis.
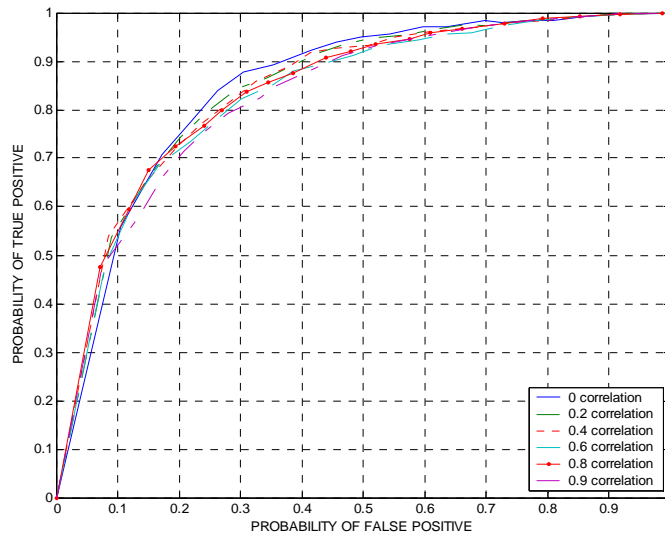
49

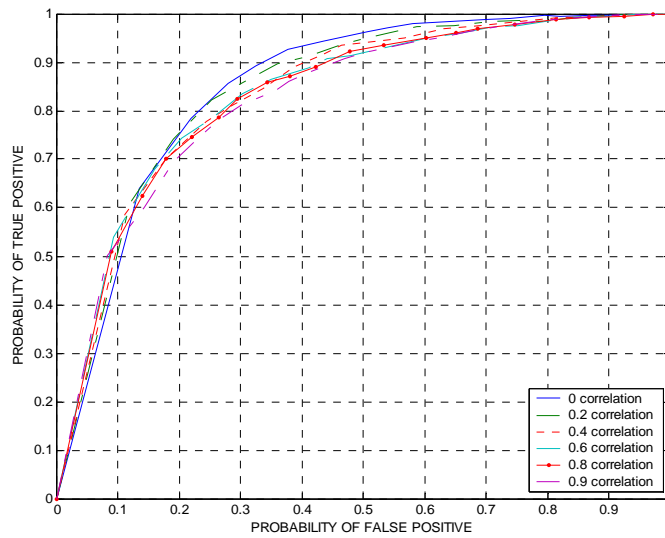Figure 22: ROC Fusion of Linear & Quadratic Classifiers (SG)



Figure 23: ROC Fusion of Linear, Quadratic & Radial Basis Classifiers (SG)

The three-classifier ensemble's performance fell between the best and worst two-classifier ensembles because it included all three classifiers.

The methodology to calculate the expected true positive classification rates was then applied to each ensemble's results. The rates are shown in Figure 24.
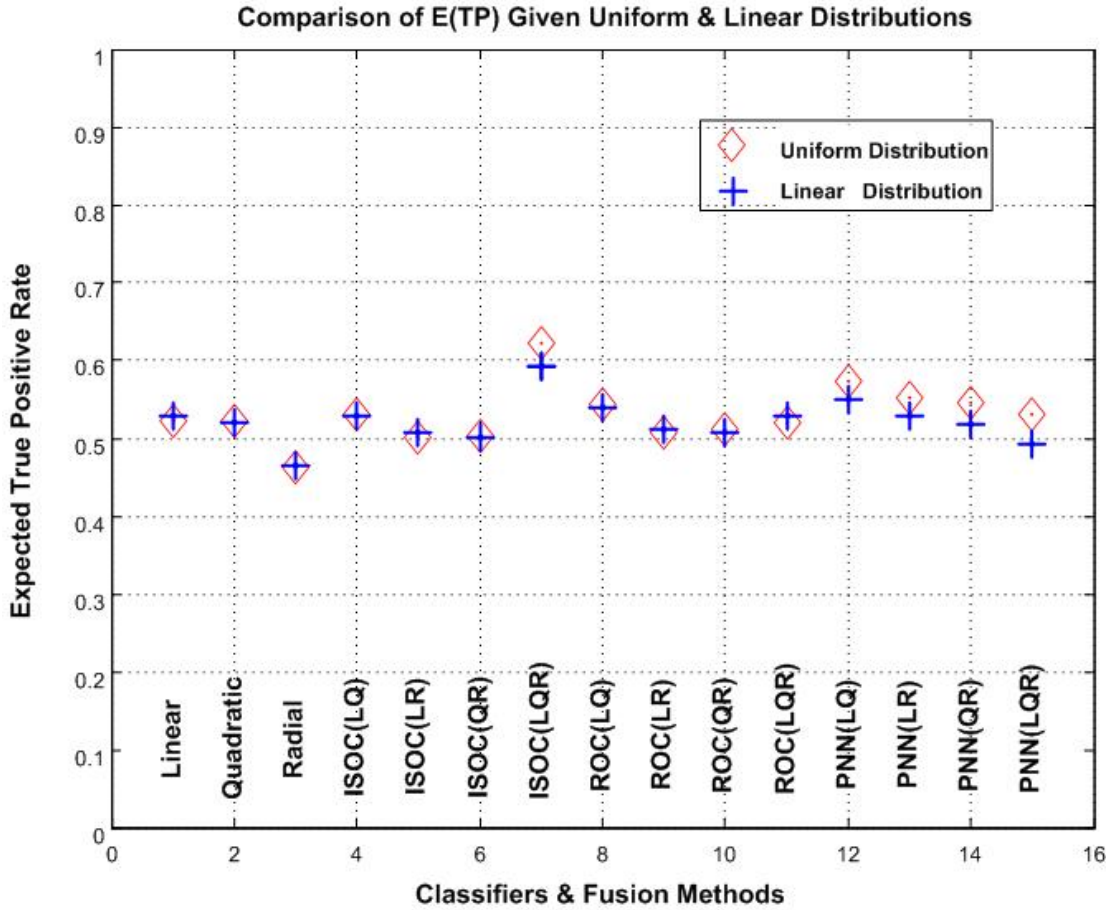


Figure 24: Simple Gaussian Expected True Positive Classification Rates

Of the three classifiers, the linear and quadratic outperformed the radial basis classifier. This fact becomes significant as the fusion methods use the posterior probabilities from these classifiers to form ensembles. ISOC two-classifier ensembles have an expected true positive classification rate that follows the performance of the classifiers being fused. However, the ISOC three-classifier ensemble's performance is better than any single classifier by approximately ten percent. This separation between the single best classifier and the best ensemble was the greatest separation found across

all experiments.  It is interesting to note that fusion, in the cases studying, only provided a marginal increase in classification accuracy.  The three-classifier ISOC ensemble performance reflects the analysis described for the ROC curves.  Briefly, the classifiers are making different errors and the optimal ISOC fusion is able to identify some of the individual classifiers error.  The ROC ensembles' performance follows the performances of their member classifiers, but never worse than the worse classifier.  In the PNN two-classifier ensembles, the ensembles follow the performances of their member classifiers. For reasons discussed in the PNN three-classifier ensemble, the three-classifier ensemble performs poorest of all PNN ensembles.  The PNN ensemble performances are always outperform most of the other ensembles.  In this case, even the poorest PNN ensemble performs approximately as well as the best single classifier.  Finally, the prior linear distribution of correlation levels shows that the distribution only affects classification when the ensemble is sensitive to correlation.

**Result 2: Two-Class XOR (close) Gaussian Data Experiment**

All single classifier ensembles had comparable performances.  The data was poorly separated so that the classifiers had some difficultly distinguishing between the two classes.  The poor classification across all the classifiers caused all ensembles to perform poorly as well.  Figure 25 shows the linear classifier ROC curve.  It is representative of all the classifier and ROC and PNN ensembles.  Even the PNN ensembles, which are usually sensitive to correlation levels, showed no distinction between 0 correlation and 0.9 correlation levels.  The ISOC ensembles provided the only notable variation among the ensembles.
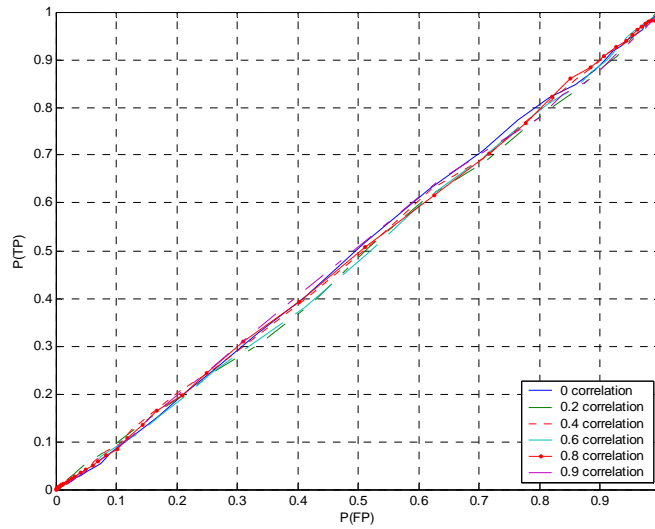
Figure 25: Linear Classifier ROC Curve

The ISOC ensembles can not construct a complete ROC curve under the given

conditions. All three two-classifier ISOC ensembles and the three-classifier ISOC

ensemble failed to construct ROC curves over the full 0 to 1 false positive range.
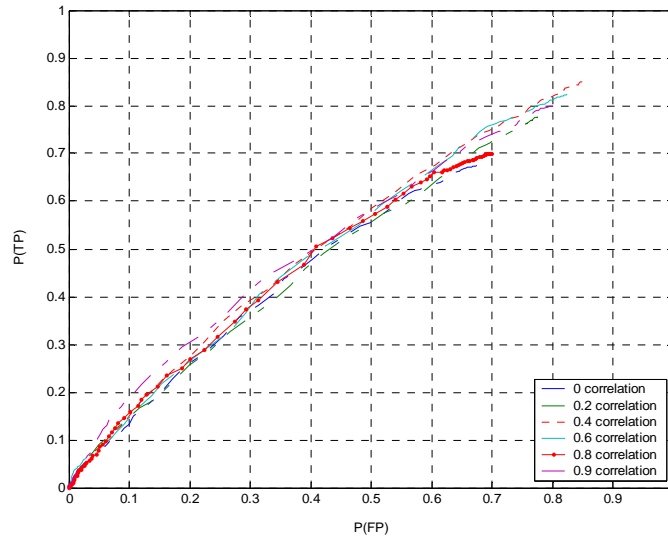


Figure 26: ISOC – Linear/Quadratic Ensemble ROC Curve

The reason for this failure can be seen in the optimal ISOC rule and its application to the classifiers posterior probabilities.

Table 9: Optimal ISOC Rule

| Linear | Quadratic | Optimal ISOC Rule |
|--------|-----------|-------------------|
| 1 | 1 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 0 | 0 |

If a poor classifier has a condition where its $P(fp) > P(tp)$ at the threshold used to develop the ISOC rule, then the "all hostile", or (1,1), sensor output state will not be the first introduced into the optimal ISOC rule. Furthermore, if the lowest cost ISOC rule only contains one sensor output state, then the "all hostile" sensor output state will be excluded from the optimal ISOC rule. When constructing the ROC curve, one expects to find an increasing number of hostile indications with lower threshold values. And hence, the ROC curve will reach 100 percent probability of finding all true positive and false positive exemplars when the threshold reaches zero. However, given the optimal ISOC rule in Table 9, when the threshold approaches zero and the "all hostile" sensor output state becomes prevalent, the optimal ISOC rule labels the exemplars as "friendly" and returns the ROC curve to the origin. Appendix A provides more details.

The expected true positive classification rates were determined for each ensemble and provide no new insight. All ensembles performed poorly. The single best classifier's performance and the best ensemble were approximately the same. Among the ensembles there appears to be very slight improvements and degradation of performance depending

54

upon the relative performance of the single classifier members. Figure 27 shows the expected true positive classification rates.
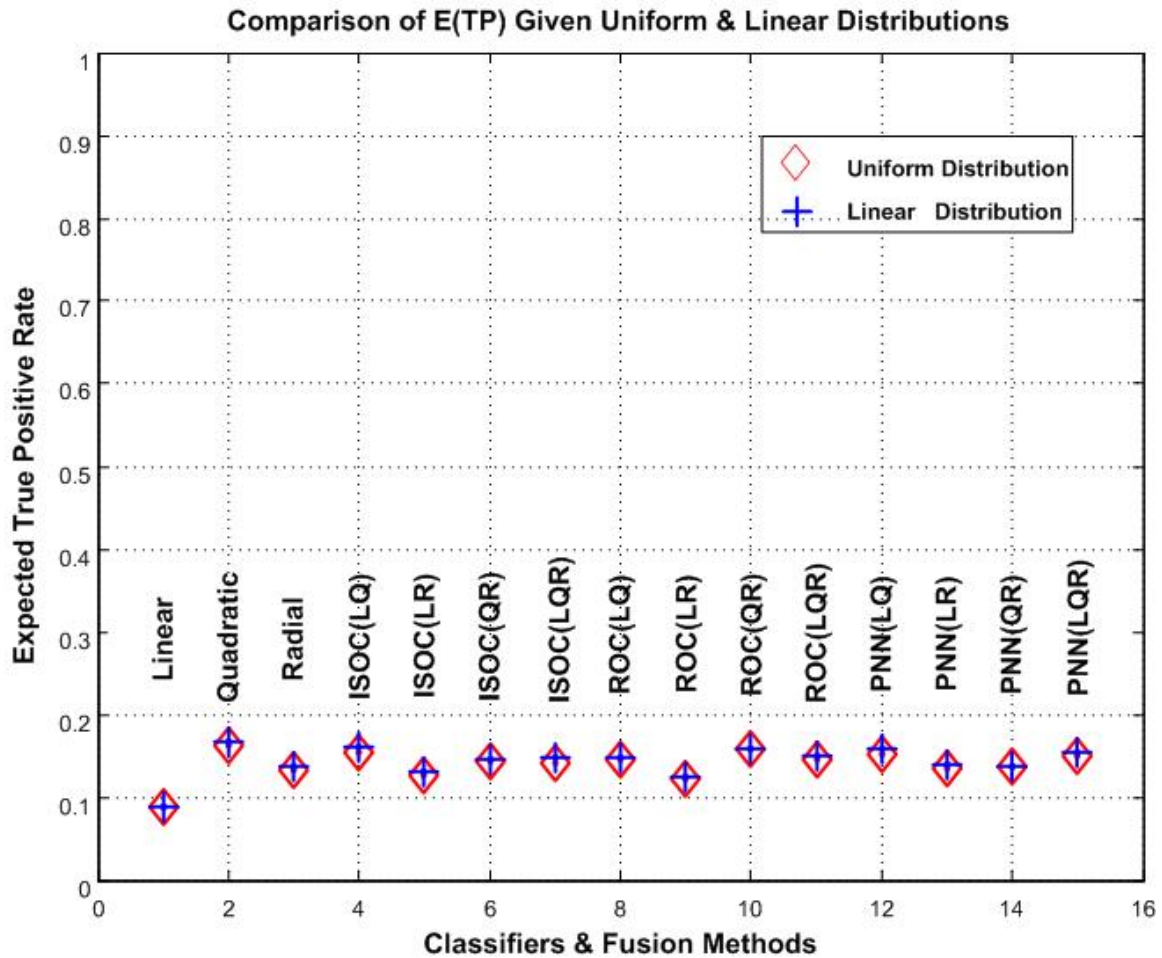


Figure 27: XOR (close) Gaussian Expected True Positive Classification Rates

### Result 3: Two-Class XOR (spread) Gaussian Data Experiment

This experiment represents the case where one classifier fails while two other classifiers perform well. The linear classifier performed about as well as chance, while the quadratic and modified radial basis classifiers performed well. The two good classifiers have nearly identical performances. Figure 28 shown below illustrates the

linear performance and Figure 29 illustrates the quadratic and modified radial basis
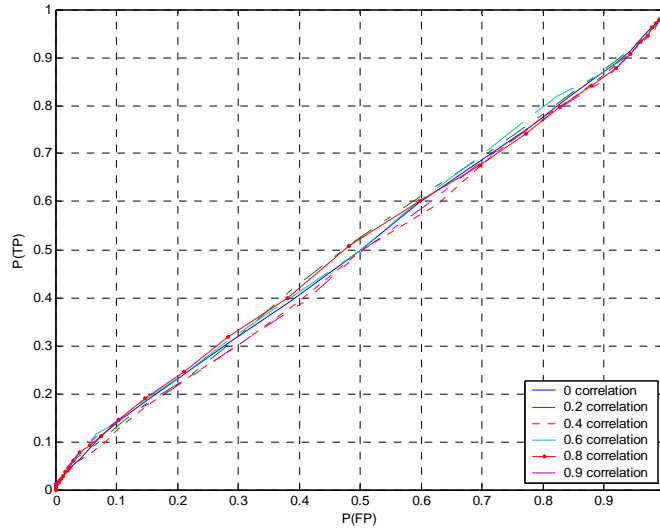
performances.



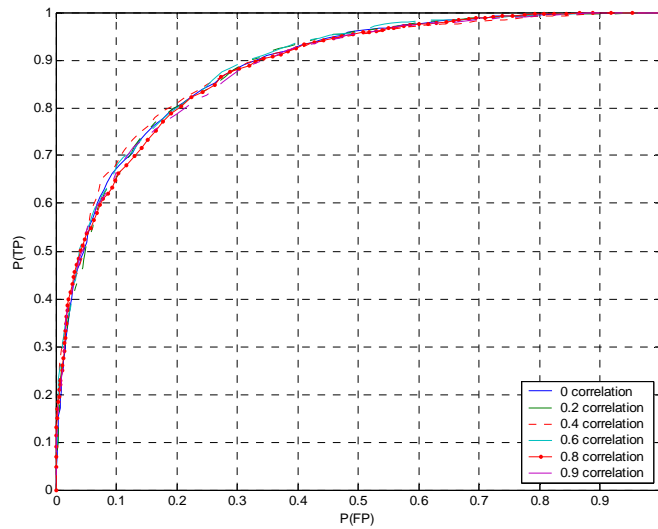Figure 28: Linear Classifier ROC Curve



Figure 29: Quadratic Classifier ROC Curve

The PNN's two-classifier ensembles exhibited two notable characteristics. If

either of the good classifiers (i.e., quadratic or radial basis) were fused with the linear

classifier, the ensemble's ROC curve results were very similar to the better classifier's results. In essence, the linear classifier was ignored and since the remaining had a very tight correlation level range the ensemble produced very tight correlation level range.
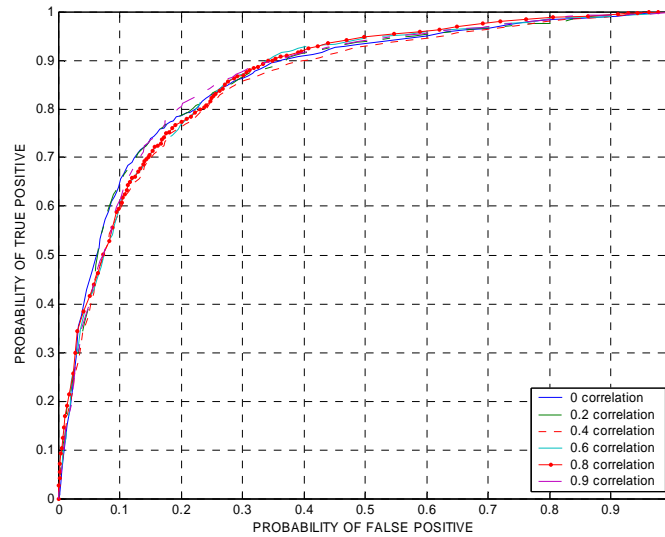


Figure 30: PNN Fusion of Linear & Radial Basis Classifiers (XOR)

The quadratic-radial basis ensemble showed the best performance. Here again, the PNN demonstrates better classification at lower correlations due to the increase in new data from more independent features. Also, while both classifiers had very tight correlation level ranges, they made different errors so that the ensemble classification performance was better than either one classifier's performance.

Since the linear classifier was extremely poor the two-classifier PNN ensemble and the three-classifier PNN ensembles are nearly identical. In fact, the data has some variation so that the two ROC curves could be considered identical.

Figure 31: PNN Fusion of Quadratic & Radial Basis Classifiers (XOR)
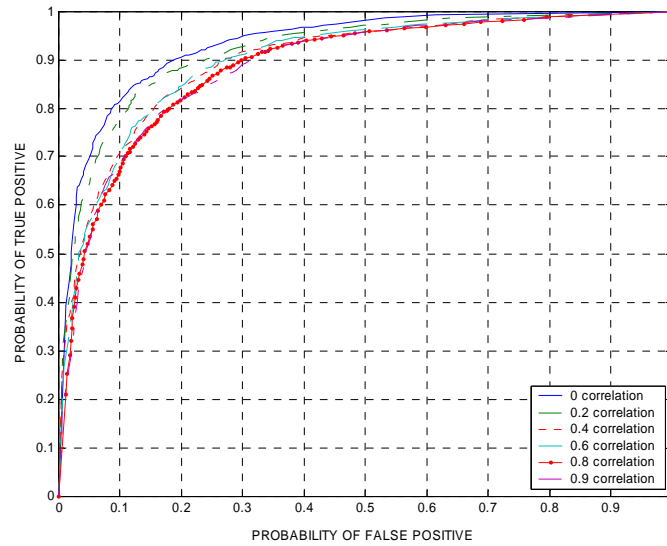


Figure 32: PNN Fusion of Linear, Quadratic & Radial Basis Classifiers (XOR)

The two-classifier ROC ensembles were similar to the two-classifier PNN

ensembles, in that, the linear classifier always has an equal probability of true positive

and false positive and its thresholds are largely of no affect. The quadratic-radial basis

ensemble shows some sensitivity to correlation levels.
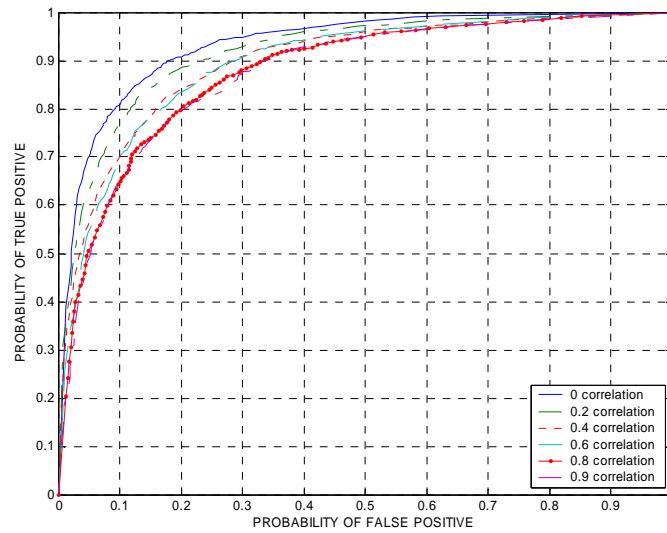


Figure 33: ROC Fusion of Linear & Radial Basis Classifiers (XOR)

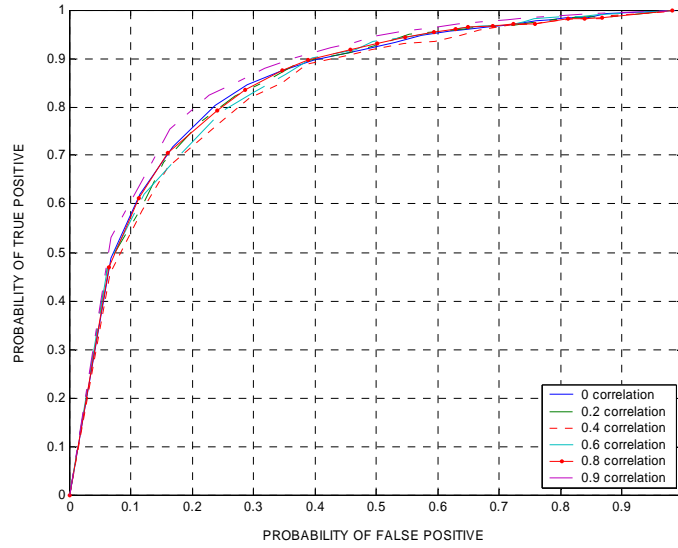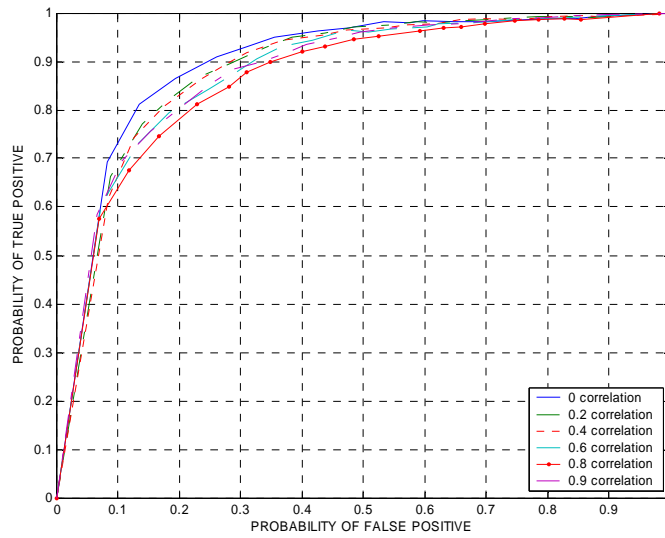

Figure 34: ROC Fusion of Quadratic & Radial Basis Classifiers (XOR)

The three-classifier ROC ensemble is identical to the quadratic-radial basis ensemble.

In the two-classifier ISOC ensembles method ignored the poor linear classifier and classified the exemplars nearly exclusively the same as the better classifiers. There is a very slight difference between the radial basis and quadratic classifiers so that when ISOC fuses the two classifiers there is a slight classification improvement. The three-classifier ISOC ensemble mostly ignores the linear classifier and performs nearly identically to the quadratic/radial basis ISOC ensemble.
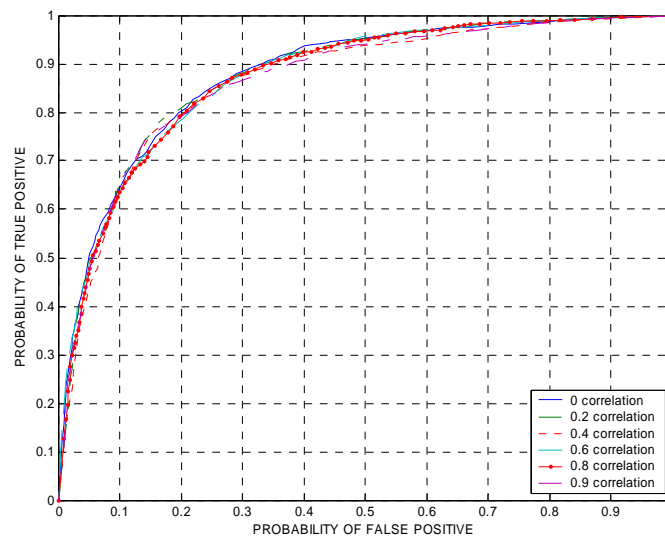


Figure 35: ISOC Fusion of Linear, Quadratic & Radial Basis Classifiers (XOR)

The expected true positive classification rates highlights the ensembles' classification differences more dramatically. It illustrates the linear classifiers poor performance. It also shows that the linear classifier detracted from the classification performance of an ensemble in which it was included. Another trend which is readily evident in Figure 36 is that the ensemble methods are fairly resilient to the poor classifiers affects. For example, the expected linear true positive classification rate is

approximately 10 percent. After fusion with any other classifier, the worst performance

was approximately 57%.
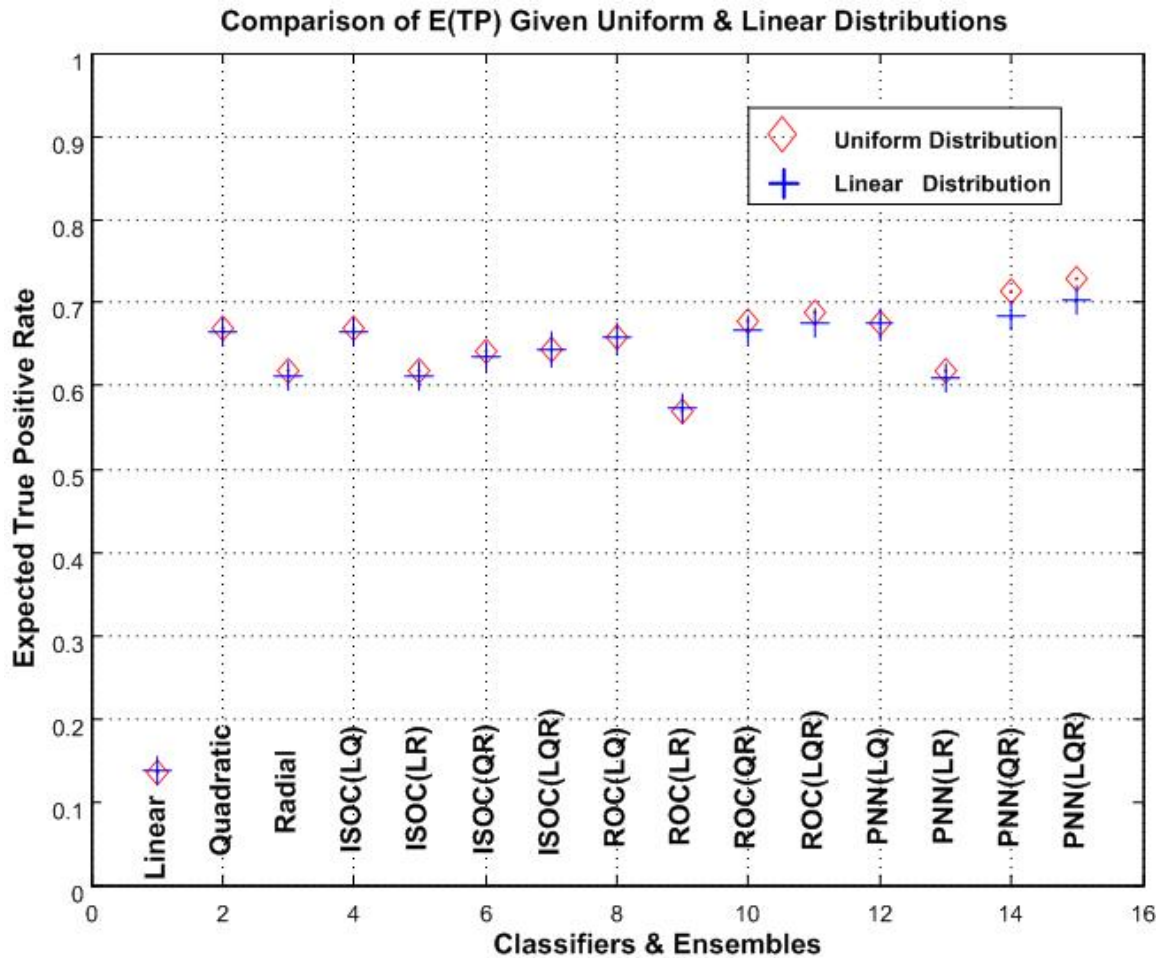


Figure 36: XOR (spread) Gaussian Expected True Positive Classification Rates

The last significant remark regarding the Figure 36 concerns the PNN ensemble

performance. In the linear/radial basis PNN ensemble, its performance is reduced by

fusion with a poor classifier. However, in the case three-classifier PNN ensemble, the

performance is actually increased above the quadratic/radial basis PNN ensemble. The

question may be how the linear classifier can improve classification in one case and

reduce classification in another. In the linear/radial basis PNN ensemble, PNN must use

all the training data given to it.  It discards nothing so it incorporates some of the linear classifiers errors when it fuses.  PNN, by design, can not ignore the linear classifier.  However, in the three-classifier PNN fusion the quadratic and radial basis function correct classifications outweigh the linear classifier's errors.  In addition, the linear classifier and one of the other good classifiers must outweigh some of the errors of the third to improve the three-classifier ensemble's performance.
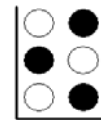
**Result 4: Two-Class "Domino" Gaussian Data Experiment**

This experiment represents the case where two classifiers perform poorly while only one classifier performs well.  The linear and quadratic classifiers perform nearly identically poorly but the modified radial basis classifier performs well.  Figure 37 illustrates the linear and quadratic classifiers' performance while Figure 38 illustrates the modified radial basis classifier's performance.
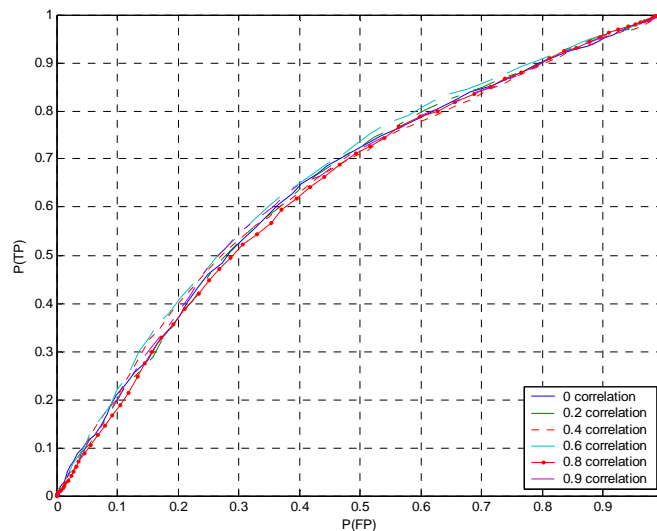


Figure 37:  Linear Classifier ROC Curve
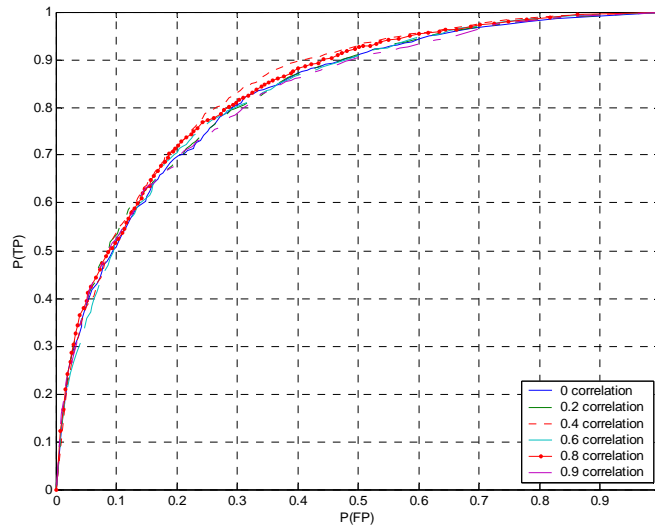
Figure 38: Radial Basis Classifier ROC Curve

The three fusion methods ignore the poor classifiers in the fusion ensembles. All two-classifier ensemble performances are nearly identical. Since the linear and quadratic classifiers perform equally poor and apparently make the same type of errors, all the linear/quadratic fusion ensembles are also nearly identical. The only notable finding is that the three-classifier PNN ensemble classification performance improves as the correlation increases. Figure X shows the three-classifier PNN ensemble performance. This can be attributed to the geometry of the data as the correlation levels increase. At the 0 correlation level, the two-class data appears to be circular in two-space with particular means and equal concentric distributions about the means. However, as the correlation levels increase the distribution of the data becomes oblong about the means. This in turn leads to a better separation of the classes and hence better classification performance. The "domino" case exhibited this phenomenon whereas the other data sets

did not because the means of the classes must be off-center from one another for the elongating data to separate. Otherwise, the data elongates into the other class.
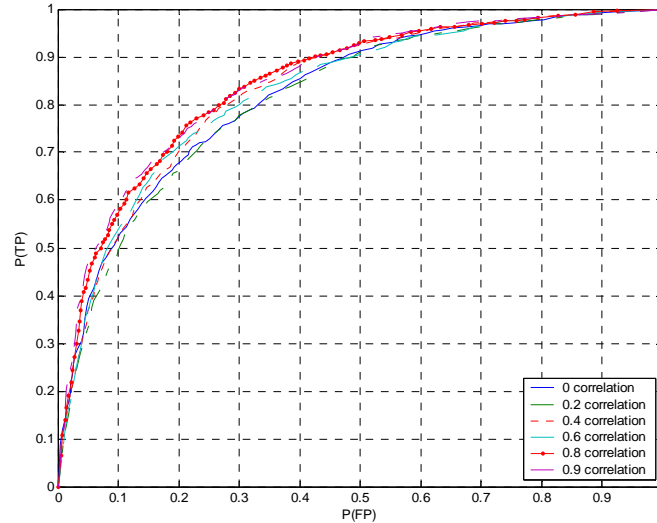


Figure 39: Three-Classifier PNN Ensemble ROC Curve

The expected true positive classification rates showed a variety of results across the ensembles. As is expected, linear / quadratic fusion ensembles performed poorly. The ISOC ensembles which include the modified radial basis classifier, ignore the linear and quadratic classifiers. However, the three-classifier ISOC ensemble does not ignore the linear and quadratic classifiers. This is due to the fact that the modified radial basis performs poorly at the mid-points between the data. Given that the linear and quadratic classifiers perform well at these points, they are included in the optimal ISOC rule. Unfortunately, the two classifiers perform poorly overall and the overall classification decreases. The ROC ensembles perform as expected. The fusion of poor classifiers with a good classifier causes the classification accuracy to decrease. When two poor classifiers are used the classification drops even more.
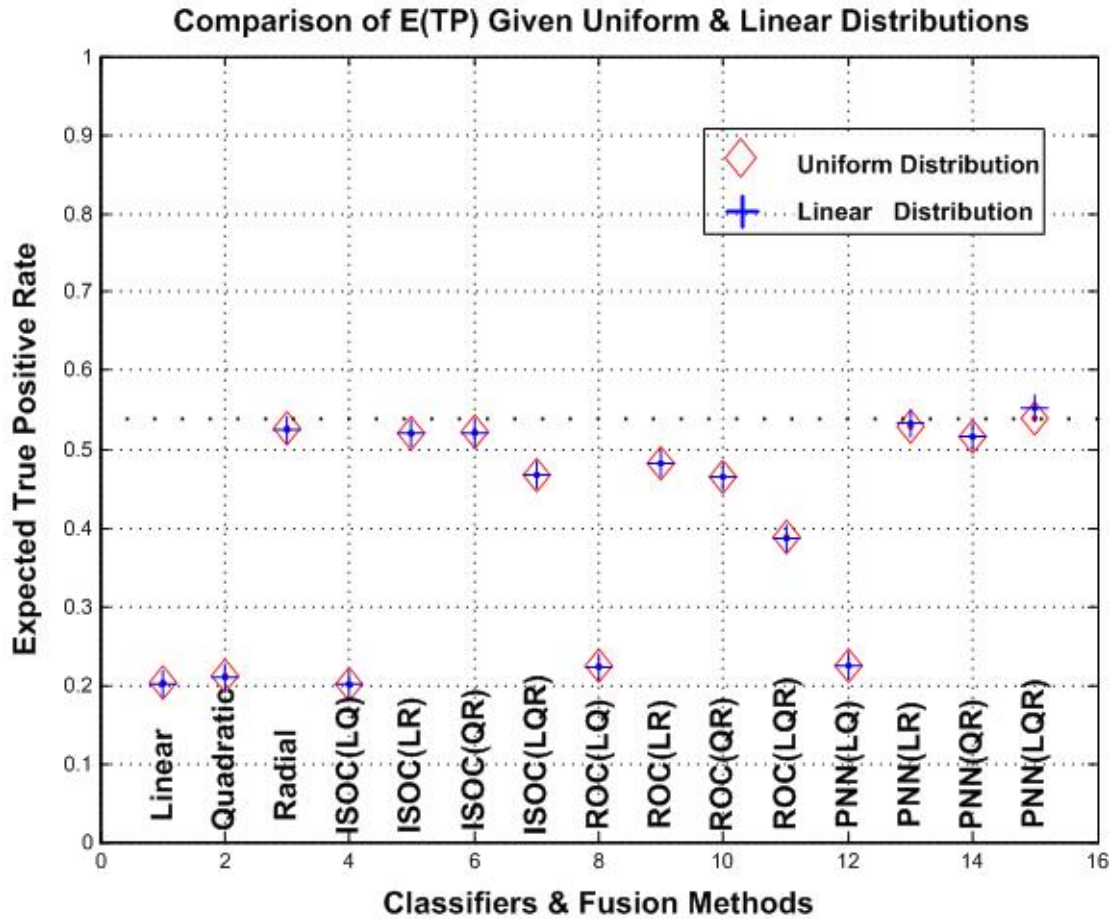
Figure 40: "Domino" Gaussian Expected True Positive Classification Rates

Finally, the PNN ensembles appear to overcome the poor classifiers performance.

This can be attributed to the classifiers making different types of errors in the mid-point

region of the "domino." The PNN ensemble can adjust for these errors using the training

data while the other methods can not.

**Result 5: Radial basis function vs general regression neural network**



An initial conjecture was made that the radial basis function approximates an

optimal Bayes classifier. A general regression neural network does, in fact, approximate

an optimal Bayes classifier. This experiment used scatter plots of both classifiers'

labeled exemplars and their respective ROC curves to compare the classification

performance and determine if the initial conjecture was reasonable.

The scatter plots show the GRNN and modified RBF classifiers classification of

the same exemplars. In this case, the GRNN classifier has very distinct line of

classification between the two classes. The RBF, on the other hand, does not have that

distinct line of classification. There is some "bleeding" into each of the other classes'

space. The most noticeable difference between the two classifiers is that the RBF more

often misclassifies outliers. The differences in classification could be attributed to an

insufficient amount of data. The GRNN ROC curve very slightly outperforms the

modified RBF ROC curve but enough to reject the concept that the modified RBF

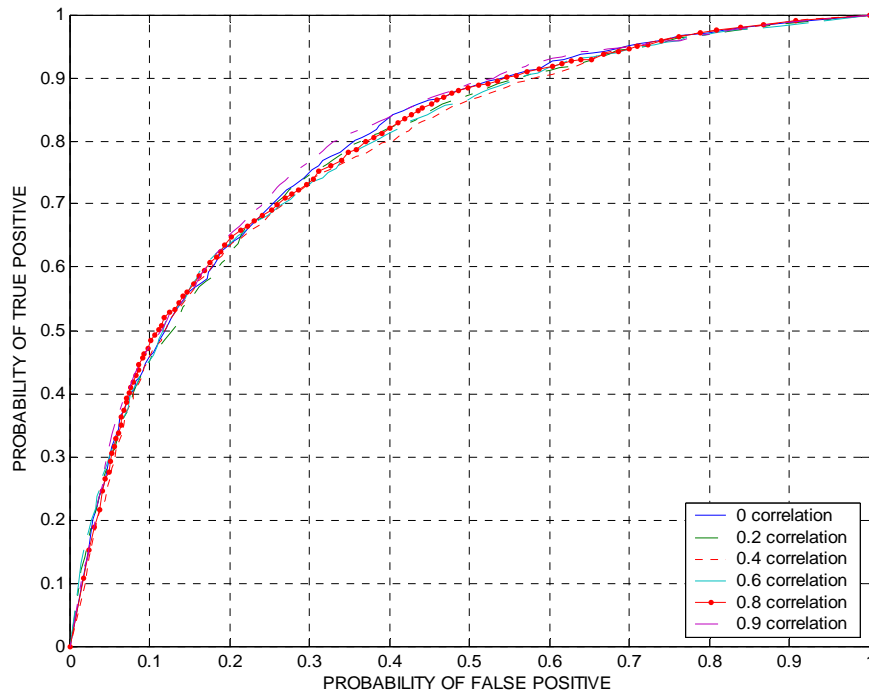classifier's output could be used as posterior probabilities.



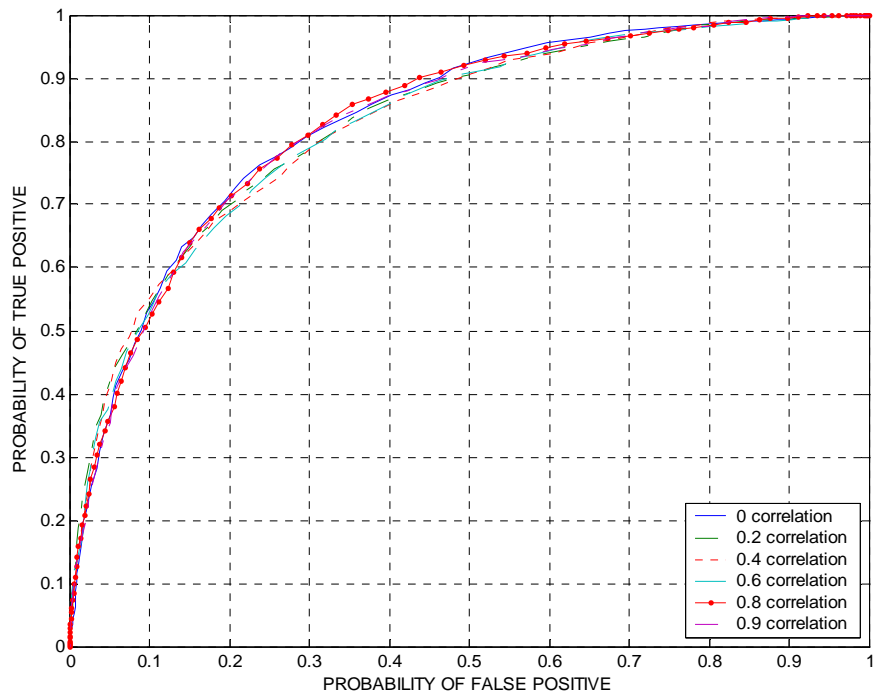Figure 41: Modified Radial Basis Classifier ROC Curve

Figure 42: General Regression Neural Network Classifier ROC Curve

## V.    Conclusions

**Introduction**

The objective of the research was to determine the optimal sensor ensemble and fusion technique combination across differing prior correlation distributions.  To this end, four different feature geometries were generated and classified by all possible ensembles. The possible ensembles consisted of three single classifier ensembles and all possible combinations of the single classifier ensembles using three different fusion techniques. Finally, the ensemble performances were measured using ROC curves and expected true positive classification rates.

**Conclusions**

Several conclusions can be drawn from the findings and analysis.  When only good classifiers were used, the ISOC ensemble was able to reduce the optimal ISOC rule to a majority vote method which successfully eliminated individual classification errors. Only errors common to all three classifiers affected the ISOC ensemble's performance. In this case, ISOC fusion method outperformed both the ROC and PNN fusion methods. When good and bad classifiers are used, then PNN ensembles consistently outperformed the other fusion methods.  Finally, when only poor classifiers are used, none of the three fusion techniques could significantly improve the classification performance.  In addition, the optimal ISOC rule declared a target as "friendly" when the underlying classifiers indicated the was target "hostile."  This caused the ROC curve construction to fail.  The last experiment, affirmed that it was reasonable to use the modified radial basis function neural network outputs as posterior probabilities.

Overall, the fusion techniques exhibited consistent characteristics. ROC fusion always performed no worse than the worst classifier. ISOC and ROC techniques were generally very "robust" to correlation. Lastly, fusion does not yield large increases in classification accuracy above the single best classifier. It is useful, in that, it mitigates the affects of poor classifiers. When a good and poor classifier are fused by any method the resulting classification accuracy is generally close to the better classifier.

**Recommendations for Future Research**

There are several possible areas for future research. This research used feature data generated from code developed in Matlab. However, the next logical step in the investigation of fusion ensembles would use real sensor feature data. This would allow researchers to find any similarities and differences between the artificial environment and the "real" world.

Another area of future research would use sensor classifiers appropriate for the actual sensor data obtained. These classifiers may optimize different regions of the feature space or be sensitive to correlated features or have any number of characteristics that will affect the classification accuracy. In addition, fusion techniques perform best with classifiers possessing complementary error types. In other words, classifiers that make different types of errors.

Lastly, the number of classes used for this experimentation should be expanded to three and/or four classes. In the operational world, there are non-combatants such as civilians and international relief workers that do not participate in combat. A fourth class would be hostile, friendly or non-combatants that can not be identified with any degree of confidence. This class might be called an "unknown" class.

## Appendix A: ISOC Likelihood Ratios and Cost Rules

The ISOC method uses an optimal rule to classify new exemplars. The optimal rule receives the classifiers' output states and if the combination of the classifiers' output states matches an output state in the optimal rule, then the target is declared hostile. Otherwise, the target is declared friendly. However, when the ISOC method fuses a poor classifier and a good classifier, it may exclude the output state of both classifiers indicating hostile from the optimal rule. The effect is that the target is declared friendly when both classifiers indicate hostile. This circumstance requires two conditions in the ISOC process. First, the likelihood ratio of some output state must be higher than the likelihood ratio of an all hostile output state. Secondly, the cost of the ordered set of rules must be less with the rule associated with the all hostile output state excluded.

This research considers only a two class, two classifier problem. The possible number of Combat Identification States (CIS) is defined by the number of classifiers and output states. In this case, there are four CIS states. The probability of each CIS state given a hostile and friendly are found.

Table 10:  ISOC Output States and Conditional Probabilities

| State $(S_i)$ | Output State Classifier$_1$ | Classifier$_2$ | $P(S_i \mid H)$ | $P(S_i \mid F)$ |
|---|---|---|---|---|
| $S_1$ | 1 | 1 | $P(tp)_1 * P(tp)_2$ | $P(fp)_1 * P(fp)_2$ |
| $S_2$ | 1 | 0 | $P(tp)_1 * P(fn)_2$ | $P(fp)_1 * P(tn)_2$ |
| $S_3$ | 0 | 1 | $P(fn)_1 * P(tp)_2$ | $P(tn)_1 * P(fp)_2$ |
| $S_4$ | 0 | 0 | $P(fn)_1 * P(fn)_2$ | $P(tn)_1 * P(tn)_2$ |

Where 1 represents a hostile target and 0 represents a friendly target. The true positive, false positive, true negative and false negative indications are taken from this notation. A confusion matrix would look as follows:

Table 11: Confusion Matrix

|        |       | Truth |      |
|--------|-------|-------|------|
|        |       | 1     | 0    |
| Label  | "1"   | tp    | fp   |
|        | "0"   | fn    | tn   |

The likelihood ratios of the four CIS states are the ratios of the $P(S_i \mid H) / P(S_i \mid F)$.

$$LR_1 = P(tp)_1 * P(tp)_2 / P(fp)_1 * P(fp)_2$$
$$LR_2 = P(tp)_1 * P(fn)_2 / P(fp)_1 * P(tn)_2$$
$$LR_3 = P(fn)_1 * P(tp)_2 / P(tn)_1 * P(fp)_2$$
$$LR_4 = P(fn)_1 * P(fn)_2 / P(tn)_1 * P(tn)_2$$

In this case, first likelihood ratio represents the CIS state where both classifiers indicate that the target is hostile. What is of interest is when one of the other likelihood ratios is

greater than the first.

$$LR_3 > LR_1$$

$$\frac{P(fn)_1 * P(tp)_2}{P(tn)_1 * P(fp)_2} > \frac{P(tp)_1 * P(tp)_2}{P(fp)_1 * P(fp)_2}$$

$$\frac{P(1-tp)_1 * P(tp)_2}{P(1-fp)_1 * P(fp)_2} > \frac{P(tp)_1 * P(tp)_2}{P(fp)_1 * P(fp)_2}$$

$$\frac{P(tp)_2 - P(tp)_1 * P(tp)_2}{P(fp)_2 - P(fp)_1 * P(fp)_2} > \frac{P(tp)_1 * P(tp)_2}{P(fp)_1 * P(fp)_2}$$

$$P(tp)_2 - P(tp)_1 * P(tp)_2 > \frac{P(tp)_1 * P(tp)_2 * (P(fp)_2 - P(fp)_1 * P(fp)_2)}{P(fp)_1 * P(fp)_2}$$

$$P(tp)_2 - P(tp)_1 * P(tp)_2 > \frac{P(tp)_1 * P(tp)_2 * P(fp)_2 - P(fp)_1 * P(fp)_2 * P(tp)_1 * P(tp)_2}{P(fp)_1 * P(fp)_2}$$

$$P(tp)_2 - P(tp)_1 * P(tp)_2 > \frac{P(tp)_1 * P(tp)_2 - P(fp)_1 * P(tp)_1 * P(tp)_2}{P(fp)_1}$$

$$\frac{P(tp)_2 - P(tp)_1 * P(tp)_2}{P(tp)_2} > \frac{P(tp)_1 * P(tp)_2 - P(fp)_1 * P(tp)_1 * P(tp)_2}{P(tp)_2 * P(fp)_1}$$

$$1 - P(tp)_1 > \frac{P(tp)_1 - P(fp)_1 * P(tp)_1}{P(fp)_1}$$

$$P(fp)_1 - P(tp)_1 * P(fp)_1 > P(tp)_1 - P(fp)_1 * P(tp)_1$$

$$P(fp)_1 > P(tp)_1$$

A proof shows that the third likelihood ratio is greater than the first when the probability
of a false positive is greater than the probability of a true positive for first classifier. The
same could be said of the second likelihood ratio and the second classifier. Once it is
established that a likelihood ratio other than the all hostile likelihood ratio is greater, then
the ordered likelihood ratio will have that greater likelihood ratio as the first to enter the
optimal rule.

The second condition requires that the cost of the optimal rule without the all
hostile output state be greater than with the all hostile output state. Given the following

order in which output states enter the optimal rule a comparison can be made between the

first and second possible optimal rules' costs.

Table 12:  Combat Identification States

| | Combat Identification States (CIS) | | | |
|---|---|---|---|---|
| Rule$_i$ | (1 , 1) | (1 , 0) | (0 , 1) | (0 , 0) |
| Rule$_1$ | 0 | 0 | 1 | 0 |
| Rule$_2$ | 1 | 0 | 1 | 0 |
| Rule$_3$ | 1 | 0 | 1 | 1 |
| Rule$_4$ | 1 | 1 | 1 | 1 |

The cost associated with the first rule, which is associated with CIS output state and the

third likelihood ratio, needs to be greater than the cost of the second rule.  The second

rule includes the all hostile CIS output state.  When the first cost is less than the second

then the ISOC method will exclude the all hostile CIS state from the optimal rule.

$$Cost_1 < Cost_2$$
$$C_{fp} * P(F) * P(fp)_1 + C_{fn} * P(H) * P(fn)_1 < C_{fp} * P(F) * P(fp)_2 + C_{fn} * P(H) * P(fn)_2$$
$$P(fp)_1 + P(fn)_1 < P(fp)_2 + P(fn)_2$$

## Bibliography

1. Aircraft Accident Investigation Board Report. "U.S. Army UH-60 Black Hawk Helicopters 87-26000 and 88-26060." 13 July 1994. <http://www.schwabhall.com/opc_report.htm>.

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

3. Air Force Doctrine Document 2-5.2, *Intelligence, Surveillance and Reconnaissance Operations*, 21 April 1999.

4. Air Force Instruction 13-1AOC vol. 3., *Operational Procedures – Aerospace Operations Center*, 1 July 2002.

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

6. Alsing, Stephen. *The Evaluation of Competing Classifiers*. Ph.D. Dissertation AFIT/DS/ENS/00-01, Wright Patterson AFB OH : March 2000.

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

8. Combat Air Forces Concept of Operations for Command and Control against Time Critical Targets, 8 July 1997, Headquarters Air Combat Command.

9. Concept of Operations for Aerospace Operations Center, 9 March 2001, Air Combat Command/Aerospace Command and Control Intelligence Surveillance and Reconnaissance Center.

10. Conway, James. Lieutenant General. "First Marine Expeditionary Force Briefing from Iraq." 2003. 30 May 2003 <http://www.dtic.mil/jcs>.

11. Dillon, W.R., and Goldstein M., *Multivariate Analysis: Methods and Applications*, John Wiley & Sons, Inc.: New York, 1984.

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

13. Fabio Roli, ISIF Tutorial TC1 Fusion of Multiple Classifiers 11 July, 2002, Annapolis MD.

14. Final Report to Congress - Conduct of the Persian Gulf War, April 1992, U.S. Department of Defense.

15. GAO-01-632, Combat Identification Systems – Strengthened Management Efforts Needed to Ensure Required Capabilities, Draft 2001, Government Accounting Office.

16. Garamone, Jim. "Fixes Touted to Combat Friendly Fire Casualties." American Forces Press Service 1999. 21 Feb 1999 <http://defenselink.mil/news/Feb1999/n020211999_9902027.html>.

17. German S., E. Bienenstock, and R. Doursat. 1992. "Neural networks and the bias/variance dilemma." *Neural Computation* (4)1-58.

18. Gilbert, E. S. (1969). "The effect of unequal variance-covariance matrices on Fisher's LDF," *Biometrics*, **25**, 505-516.

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

20. Jain A.K., R.P.W. Duin, J. Mao, "Statistical pattern recognition: a review", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1) (2000) 4-37.

21. Kuncheva, L.I., Bezdek, J.C., Duin, P.W., "Decision templates for multiple classifier fusion: an experimental comparison", *The Journal of the Pattern Recognition Society*, 2001, pp. 299-214.

22. Marks, S. and Dunn, O. J. (1974). "Discriminant functions when covariance matrices are unequal," *Journal of the American Statistical Association*, **69**, 555-559.

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

24. Peters, F. Whitten and General Micheal E. Ryan. Prepared Statement before the Senate Armed Services Committee, 12 Feb 1998, <http://www.senate.gov/~armed_services/statemnt/980212pr.htm>.

25. 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.

26. Report on the Defense Science Board for Combat Identification, 10 Sept 1996, Under Secretary of Defense for Acquisition and Technology.

27. Report of the Defense Science Board Task Force on Combat Identification, May 1996, Under Secretary of Defense for Acquisition and Technology.

28. Richard, M.D. and R.P. Lippman. 1991. "Neural network classifiers estimate Bayesian a posteriori probabilities." *Neural Computation* 4(3):461-483.

29. Ruck, D. W., Rodgers, S. K., Kabrisky, M., Oxley, M. E., and Suter, B. W. "The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function," Letters, IEEE Transactions on Neural Networks, Vol. 1, No. 4, Dec 1990.

30. 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. 135-148.

31. Storm, Susan. *An Investigation of the Effects of Correlation in Sensor Fusion.* MS Thesis AFIT/GOR/ENS/03-22, Wright Patterson AFB OH: March 2003.

32. Wasserman, P.D., *Advanced Methods in Neural Computing*, Van Nostrand Reinhold: New York, 1993.

33. Weisman, Jonathan. "2 Patriot missile mistakes raise questions." Milwaukee Journal Sentinel 26 March 2003.

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

35. Woodley, Robert and Levent Acar. 2003. "Line Detection Using a Biologically – Motivated, Continuous – Time, Feedback System Based on the Visual Cortex." Paper presented at the Artificial Neural Networks In Engineering meeting, St. Louis, November.

**Vita**

Capt Paul P. Clemans graduated from South Haven Christian High School in Portage, Indiana.  He entered undergraduate studies at the United States Merchant Marine Academy (USMMA) where he graduated with a Bachelor of Science Degree in Marine Engineering and a Third Assistant Engineer License for Steam or Diesel Vessels of Unlimited Tonnage in June 1996.  He was commissioned after Field Training at Lackland AFB, TX with a reserve commission.

His first assignment was at Hanscom AFB, MA as a project manager for the Electronic System Center, Distributed Common Ground Stations Program Office.  He then came to Wright-Patterson AFB, OH as a program manager for the Aeronautics System Center, Reconnaissance Program Office.  In August 2003, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology.  Upon graduation, he will be assigned to Randolph AFB, T

# REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From – To)* |
|---|---|---|
| 05-03-2004 | **Master's Thesis** | Jun 2003 - Mar 2004 |

| 4.  TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| AN INVESTIGATION OF THE OPTIMAL SENSOR ENSEMBLE FOR SENSOR FUSION | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |

| 6.    AUTHOR(S) | 5d. PROJECT NUMBER |
|---|---|
| Clemans, Paul, P., Captain, 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<br>Graduate School of Engineering and Management (AFIT/EN)<br>2950 Hobson Street, Building 642<br>WPAFB OH 45433-7765 | AFIT/GOR/ENS/04-03 |

| 9.  SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
|---|---|
| AFOSR<br>Attn:  Major Juan R. Vasquez<br>801 North Randolph Street, Room 933   (703) 696 - 8431<br>Arlington, VA 22203-1977      e-mail:  Juan.Vasquez@afosr.af.mil | 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 continues the research begun by Storm, Bauer and Oxley in 2003 into the fusion of classifiers.  It examines the fusion of up to three correlated classifiers using three different fusion techniques.  The overall objective was to determine the optimal ensemble of classifiers to maximize the expected classification accuracy.  The ISOC fusion method (Haspert, 2000), the ROC "Within" fusion method (Oxley and Bauer, 2002) and a Probabilistic Neural Network were the three fusion techniques employed in these set of experiments.  Performance of the classifiers and the fusion methods is measured via ROC curves. Two possible configurations of feature correlations were examined.  The expected true positive value relative to a prior distribution of correlation levels for each configuration was then used to compare the classifiers' and the fused classifiers' performance and thereby allowing for the selection of an optimal ensemble.

**15. SUBJECT TERMS**
     Sensor, Sensor Fusion, Classifier Fusion, Probabilistic Neural Network, ISOC Fusion, ROC Fusion, Combat Identification, Fratricide

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a.  NAME OF RESPONSIBLE PERSON<br>Kenneth W. Bauer, AFIT/ENS |
|---|---|---|---|---|---|
| a. REPORT | b.  ABSTRACT | c. THIS PAGE | | | |
| U | U | U | UU | 90 | 19b.  TELEPHONE NUMBER *(Include area code)*<br>(937) 255-6565, ext 4328; e-mail: Kenneth.Bauer@afit.edu |