SPECTRAL TEXTILE DETECTION IN THE VNIR/SWIR BAND

THESIS

James A. Arneal, Second Lieutenant, USAF

AFIT-ENG-MS-15-M-049

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-15-M-049

SPECTRAL TEXTILE DETECTION IN THE VNIR/SWIR BAND

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

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 Electrical Engineering

James A. Arneal, B.S.E.E

Second Lieutenant, USAF

MARCH 2015

AFIT-ENG-MS-15-M-049

SPECTRAL TEXTILE DETECTION IN THE VNIR/SWIR BAND

THESIS

James A. Arneal, B.S.E.E
Second Lieutenant, USAF

Committee Membership:

Lt Col Jeffrey D. Clark, PhD
Chair

Richard K. Martin, PhD
Member

Gilbert L. Peterson, PhD
Member

AFIT-ENG-MS-15-M-049

Abstract

Dismount detection, the detection of persons on the ground and outside of a vehicle, has applications in search and rescue, security, and surveillance. Spatial dismount detection methods lose effectiveness at long ranges, and spectral dismount detection currently relies on detecting skin pixels. In scenarios where skin is not exposed, spectral textile detection is a more effective means of detecting dismounts.

This thesis demonstrates the effectiveness of spectral textile detectors on both real and simulated hyperspectral remotely sensed data. Feature selection methods determine sets of wavebands relevant to spectral textile detection. Classifiers are trained on hyperspectral contact data with the selected wavebands, and classifier parameters are optimized to improve performance on a training set. Classifiers with optimized parameters are used to classify contact data with artificially added noise and remotely-sensed hyperspectral data.

The performance of optimized classifiers on hyperspectral data is measured with Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. The best performances on the contact data are 0.892 and 0.872 for Multilayer Perceptrons (MLPs) and Support Vector Machines (SVMs), respectively. The best performances on the remotely-sensed data are AUC = 0.947 and AUC = 0.970 for MLPs and SVMs, respectively. The difference in classifier performance between the contact and remotely-sensed data is due to the greater variety of textiles represented in the contact data. Spectral textile detection is more reliable in scenarios with a small variety of textiles.

Acknowledgements

I thank my advisor, Lt Col Jeffrey Clark, for his help at every stage of this thesis' development. His commitment to my work was the best, last, and only line of defense against thesis deadlines. Have a fantastic retirement, Sir.

Many thanks to my committee members, Dr. Richard Martin and Dr. Gilbert Peterson, for giving me the feedback I needed to make my thesis the best it could be.

To my fellow Sensors Exploitation Research Group (SERG) lab inhabitants, thanks for keeping me sane. Shane Fernandes and Stephen Sweetnich, your support on the technical aspects of remote sensing proved invaluable. And to Capt Khoa Tang: it has been an honor suffering through it all with you.

Thanks to my family for their support through nineteen straight years of school. It hasn't been easy, and I couldn't have done it without you.

James A. Arneal

Table of Contents

Page

List of Figures

List of Tables

SPECTRAL TEXTILE DETECTION IN THE VNIR/SWIR BAND

I. Introduction

Dismount detection, the process of detecting human beings located on the ground and outside of a vehicle, has applications in both civilian and military domains [31, 43]. The need for a reliable dismount detection system has prompted research into various methods of dismount detection. One approach that has been investigated is spectral detection [80], which searches for a spectral signature consistent with the presence of a dismount. The efforts by Nunez [62] capitalize on the spectral domain to detect skin as part of a dismount detection system. However, relying on skin detection for dismount detection poses problems in scenarios where a dismount's skin is not exposed. A spectral dismount detector is more robust if it can detect other spectral signatures that are highly correlated with dismounts. This thesis advances spectral detection of dismounts by investigating the performance of spectral textile detectors on remotely-sensed hyperspectral data.

1.1 Problem Statement

The necessity of dismount detection has inspired numerous efforts to reliably detect dismounts [31, 43, 80]. Spectral dismount detection exploits a spectral signature unique to dismounts. Types of spectral signatures employed to detect dismounts consist of hair and skin, which are closely associated with the presence of a human body. A spectral dismount detector locates dismounts by searching for these unique spectral signatures.

While the spectra of hair and skin are typically consistent with the presence

of a dismount, detecting these spectra may prove difficult or impossible in certain conditions. For instance, in a search and rescue operation in a cold climate, it is likely that the dismounts have a significant portion of their body's surface area covered by clothing. A spectral detector searching for human skin and hair in a cold climate has limited capability due to the high probability that very little or no hair or skin is exposed for detection. In such a scenario, a spectral textile detector will detect the clothing that the dismounts are wearing and provide valuable assistance to rescuers.

The effectiveness of any spectral detection system depends on the set of wavelengths used in the detection algorithm. Hyperspectral Imagers (HSIs) are sensors that collect the radiance over hundreds of wavebands throughout the Visible/Near-Infrared (VNIR) and Short-Wave Infrared (SWIR) ranges for each pixel in an image [88]. Unlike a standard color camera, which only collects radiance at three distinct wavebands: red (620-720 nm), green (495-570 nm), and blue (450-495 nm). HSIs measure the VNIR and SWIR spectral signatures of a subject with high spectral resolution. This abundance of information creates a multitude of characteristics for textile detection capabilities.

The abundance of information from a hyperspectral image is useful in detecting textiles. However, using the entire spectrum of HSI information could be costly and possibly degrade detection capabilities. Depending on the type of detection algorithm used, it may be overly time-consuming to process hundreds of spectral bands for each of the thousands of pixels in a hyperspectral image. Hyperspectral data is generally highly redundant, so many bands of a hyperspectral image may be removed without significantly hindering classification accuracy [87]. It is therefore desirable to reduce the dimensionality of hyperspectral data. Feature selection methods identify the features in a data set that are most relevant to a machine learning problem. Feature selection can be used to identify the wavebands in hyperspectral data that are best-

suited for textile detection purposes.

Textile detection is a valuable method for detecting dismounts independently, or as an extension of an existing dismount detection system. This thesis determines a feature set of HSI wavebands, and a detection method than can detect textiles with high accuracy. The feature sets and detection methods are applied to remotely-sensed data representative of a dismount detection scenario.

1.2 Justification

An accurate spectral textile detector utilizing the VNIR/SWIR wavebands is a valuable asset to dismount detection systems. A spectral textile detector would increase the effectiveness of existing spectral dismount detection systems, which currently rely on skin detection for cueing. Estimates of body surface area measurements reveal that shoes, long pants, and a long-sleeved shirt cover approximately 85% of a dismount's body [57]. Detecting the small fraction of skin exposed to the sensor becomes a difficult subpixel detection problem when the imager resolution is not sufficient to yield full textile pixels. However, with textiles covering 85% of a dismount, it is more probable to encounter pixels with only textile endmembers. Thus a textile detector allows for a more simple detection methodology.

Spectral detection methods are used in the development of subpixel detection, producing detectable targets for pixels encompassing multiple endmembers [13]. Subpixel detection methods become more effective as the abundance of the target endmember within the pixel increases [88]. In scenarios where a dismount does not occupy a full pixel, detecting the dismount's textile signature may be easier than detecting its skin signature as there will be a greater abundance of textile for the subpixel detection process.

1.3 Assumptions

Using a textile detector as part of a dismount detection system assumes that the presence of textiles is an indicator of the presence of a dismount. This assumption is based on the following: first, the dismounts are presumed to be wearing clothing composed of textiles that are exposed to the sensor's field of view; second, the majority of objects in the scene, other than clothing worn by dismounts, are composed of non-textiles. While this assumption may be suspect in certain cases, considering the variety of applications in which textiles are used, this research considers this assumption valid.

The hyperspectral signatures used in this thesis are processed in the reflectance domain. Reflectance is the ratio of electromagnetic power reflected by an object to the electromagnetic power incident on the object, inclusively bounded from 0 to 1 [72]. The electromagnetic power reflected by an object is measured directly by the sensor. To calculate reflectance, an accurate measurement of the radiance incident on the objects in a scene must be determined. In the hyperspectral images used in this thesis, a measurement of incident radiance is provided by pixels fully occupied by a Spectralon® white reflectance panel. Spectralon® panels are commonly used to approximate a surface with reflectance equal to 1 at all wavelengths [71].

The hyperspectral data used in this thesis consists of both contact and remotely-sensed data. Contact data was collected using a contact probe with a built-in lamp that produced electromagnetic energy in the VNIR/SWIR range. Remotely-sensed data was collected with VNIR and SWIR line scan imagers outdoors on a sunny day. Thus, the results presented in this thesis assume that the incident electromagnetic energy in the VNIR/SWIR range is sufficient to produce meaningful reflectance measurements.

1.4 Standards

The performance of spectral textile detectors is presented using probability of detection (P_D), probability of false alarm (P_{FA}), and Equal Weighted Accuracy (EWA). For this thesis, P_D is defined as the number of instances of correctly idenfiied textiles divided by the total number of instances of textiles, and P_{FA} is defined as the number of instances of non-textile spectra incorrectly identified as textiles divided by the total number of instances of non-textile spectra. EWA is a measure of accuracy for both textiles and non-textiles, defined as [16]:

$$\text{EWA} = \frac{P_D + (1 - P_{FA})}{2}, \tag{1.1}$$

bounded inclusively from 0 to 1. P_D, P_{FA}, and EWA will be calculated for multiple spectral textile detectors. It is desired to have a spectral textile detector with a high P_D and a low P_{FA}, resulting in a high EWA.

This thesis utilizes the Wilcoxon Rank Sum Test (WRST) [30] to determine if a classifier's median performance is superior to that of another classifier. The threshold of significance used in this thesis is 95% ($\alpha = 0.05$). The WRST results that meet or exceed this threshold are considered statistically significant.

The performance of selected classifiers is analyzed in depth with Receiver Operating Characteristic (ROC) curves, which evaluate the tradeoff between P_D and P_{FA} for a classifier's threshold settings [4]. The Area Under the Curve (AUC) statistic [4] is used as a measure of a classifier's performance for all threshold settings.

1.5 Approach

To create a spectral textile detector, a subset of the HSI wavebands that will produce accurate classification must be determined. Feature selection methods are

used to find wavebands that represent intrinsic spectral properties of textiles. Feature selection methods use labeled training data to determine a subset of wavebands that best differentiate between the classes [42]. Feature selection methods will be used on a set of pristine training data, and on the same set of data with noise added, to determine the effect of noise on feature selection.

To determine a feature set's differentiation ability, a detector will be trained on a set of training data containing only the selected features. The trained detector's accuracy (in terms of P_D, P_{FA}, and EWA) will be evaluated using a separate testing data set. This will be performed for multiple feature selection algorithms, and for multiple detectors.

1.6 Materials and Equipment

To collect data on background and textile materials without the atmospheric distortion associated with remote sensing data, an Analytical Spectral Devices (ASD) Fieldspec® 3 spectroradiometer with a contact probe is used. Remotely-sensed HSI data is collected using SpecTIR® VNIR and SWIR scanner imagers. MATLAB® is used for data processing, feature selection, classification, and displaying results.

II. Background

Spectral textile detection with hyperspectral data involves a broad range of concepts from multiple fields of study. A basic understanding of dismount detection is necessary for understanding a textile detection approach. In addition, classifiers and feature selection methods are critical in the use of hyperspectral data for detection. Physics and chemistry play an important role in determining the reflected electromagnetic energy of textiles.

This chapter explores the relevant concepts and works accomplished in hyperspectral dismount detection. Section 2.1 explains the utility of hyperspectral imaging as a tool for detecting dismounts. Methods of feature selection implemented on hyperspectral data are summarized in Section 2.2. Section 2.3 elaborates on techniques for detecting and classifying target spectra. Finally, Section 2.4 focuses on the unique spectral properties of textiles.

2.1 Hyperspectral Imaging for Dismount Detection

A "dismount" is defined as a person located on the ground, outside of a vehicle [43]. There are a number of applications for dismount detection in both civilian and military operations. However, there are significant practical problems with dismount detection. The relative size of dismounts in a traditional remote sensing scenario creates a subpixel detection problem due to the the low ratio of target size to ground sampling distance [13]. This has led to the application of Synthetic Aperture Radar (SAR) to detect dismounts, since the resolution with SAR is not affected by distance between the sensor and the target [31]. Unfortunately, SAR relies on a temporal data collection scheme to capture the motion of the target relative to the background [43]. Therefore, SAR is not desirable when detecting stationary targets.

Spectral detection utilizes the spectral information present in each pixel of an image to determine the presence of a target. Spectral detection presents an alternative method of dismount detection that capitalizes on known spectral signatures unique to a dismount (e.g. skin, hair, or clothing). For instance, in an electro-optical image, skin detection can be implemented by locating pixels with RGB values similar to those of skin [55]. However, methods that are limited to electro-optical spectral features are prone to producing false alarms for pixels that have similar RGB characteristics to the target [73].

Hyperspectral cameras, which collect data from hundreds of spectral bands in the visible through short-wave infrared (SWIR) range, provide additional information for each pixel that can be used to more accurately distinguish target spectra from background spectra. However, this large amount of information can be problematic. Utilizing all wavebands in a hyperspectral image is time-consuming and computationally costly. In addition, some spectral bands can be heavily influenced by atmospheric effects, rendering them irrelevant for detection purposes [76]. Feature selection methods aim to identify relevant spectral features that preserve the "target concept" and exclude spectral features that are irrelevant.

2.2 Methods of Feature Selection

The high-dimensional data that hyperspectral images contain must be condensed in such a way as to preserve the relevant spectral characteristics of each pixel while minimizing the amount of information. Data can be decomposed into (or used to generate) features. Feature selection methods identify the relevant features from a larger set of available features [42].

Blum and Langley [8] present a number of definitions of a "relevant" feature, e.g. "strongly relevant to the distribution," meaning that a feature is relevant if it is the

only one that differentiates between classes. Another defintion by Blum and Langley is "incremental usefulness": a feature is relevant if it improves the classification ability of the feature set [8].

There are many feature selection methods available. The choice of a feature selection method is dependent on the specific application and data type. Dash and Liu [21] group feature selection methods according to their feature set generation and evaluation algorithms. They define three ways of generating feature sets: complete, heuristic, and random. Complete generation algorithms search the entire space of possible sets. For heuristic generation algorithms, a measure of success is used to determine which sets should be generated. Random generation uses an element of stochasticity to assist in finding a proper feature set. The authors also define five ways to evaluate the generated feature sets: distance measures, information measures, dependence measures, consistency measures, and classifier error rate measures [21].

Blum and Langley [8] present the three broad categories of feature selection: embedded, filter, and wrapper. Embedded methods embed their feature selection within a classifier algorithm. Filters operate by filtering out irrelevant or redundant features prior to passing a set of features to a classifier. Wrapper methods use a classifier as a subroutine to generate feature sets that are evaluated by determining the classifier error rate [8].

Genetic Algorithms.

Genetic algorithms (GA) are wrapper methods that generate new feature sets based on the most successful feature sets of a previous generation. Each spectral feature is assigned a symbol that represents the feature in the gene space. For a set of n features, a genome may be a vector of length n, consisting of zeros and ones, where ones represent the selected features [28]. When the number of features to be

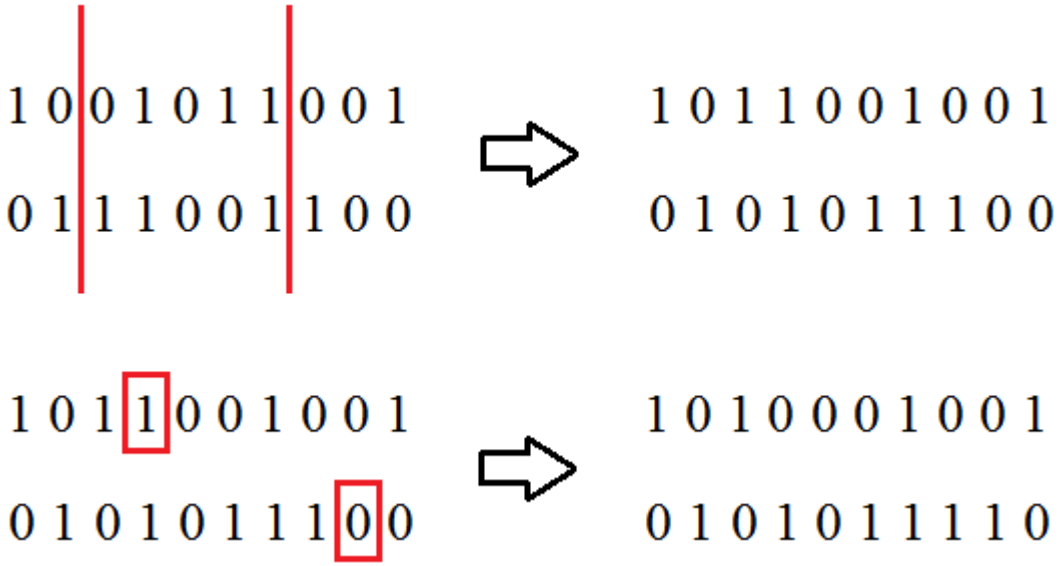


Figure 2.1. Top: A 2-point crossover operation. The digits from the top genome and the bottom genome between the two lines are swapped, while the digits outside the lines remain the same. Bottom: A single-point mutation operation. The top and bottom genomes remain the same, except for the boxed digits, which are logically negated [53].

selected is known to be k, the genome may be a vector in $\mathbb{N}^k$ where each element is the number of the feature [28]. The algorithm begins by generating an initial population of feature sets. These feature sets are all evaluated using a fitness function, and reproduce if they are sufficiently fit. Reproduction entails two operations: crossover and mutation. Crossover takes the parent genomes and crosses them over in one or more places, producing two children genomes. Mutation takes the resulting child genomes and randomly changes one or more of the genes (elements) in each [84]. Crossover and mutation are illustrated in Figure 2.1. The result of each reproduction instance is a pair of unique child genomes. The new generation is comprised of all children resulting from the previous generation. This new generation is in turn evaluated and allowed to reproduce. The algorithm loops in this manner until a stopping criterion is reached [84].

Local Search Methods.

There are a number of feature selection algorithms that use local search methods in conjunction with a heuristic to iteratively add or remove features to generate a new feature set with a better heuristic value. Local search is a type of search algorithm that begins with a candidate solution, and iteratively moves to better solutions adjacent to the candidate solution in the search space [7].

Sequential Forward Selection (SFS) begins with an empty feature set and adds features until it is halted. Sequential Backward Selection (SBS) is the opposite: it begins with a full set and removes features until it is halted. In both algorithms, the feature that is added or removed produces the best resulting feature set [74]. Both of these methods are greedy: they traverse a small subset of the feature space. Sequential Floating Forward Selection (SFFS) and Sequential Floating Backward Selection (SFBS) are modified versions of SFS and SBS, respectively. SFBS allows the removal of a feature once it has been added, while SFBS allows the addition of a feature once it has been removed [67]. The steepest-ascent method greedily traverses the feature set space by iteratively moving to the adjacent feature set with the highest heuristic value [74].

Information Theory Methods.

Information theory is often used to determine the relevance of a feature to a target class. Feature selection methods that use information theory measures rely on the "relevant to the distribution" definition of feature relevance as it pertains to the correlation and redundancy to the target class [26].

The fundamental useful measure in information theory is the entropy of a variable, X, defined as [39]:

$$H(X) = -\sum_i P(x_i) log_2(P(x_i)), \tag{2.1}$$

where $P(x_i)$ is the probability of the event $X = x_i$. Entropy is a measure of the unpredictability of a variable. Lower values of $H(X)$ indicate that X is more easily predictable. Another measure used in information theory is conditional entropy. The conditional entropy of X given Y is [39]:

$$H(X|Y) = -\sum_{ij} P(x_i|y_j) log_2(P(x_i|y_j)), \tag{2.2}$$

where $P(x_i|y_j)$ is the conditional probability of an event $X = x_i$ given an event $Y = y_j$. Conditional entropy is the measure of the entropy of X given that it is conditioned on Y. Using entropy and conditional entropy, it is possible to define a measure of how well a variable predicts another variable. This measure is called the Information Gain (IG), or mutual information. The IG of X given Y is [39]:

$$\text{IG}(X|Y) = H(X) - H(X|Y). \tag{2.3}$$

Thus, IG is the difference between the entropy of a variable and the entropy of that same variable with the added knowledge of a second variable. It is intuitive that a feature, Y, with a high IG on a class, X, would be an ideal candidate for selection in a feature set. Thus IG can be used in feature selection to determine which features are most relevant to a class distribution.

Fast Correlation-Based Filter (FCBF).

Fast Correlation-Based Filter (FCBF) method uses a measure of correlation called Symmetrical Uncertainty (SU), which is defined as twice the ratio of the IG to the sum of the individual entropies [86]:

$$\text{SU}(X, Y) = 2 \left[\frac{\text{IG}(X|Y)}{H(X) + H(Y)} \right]. \tag{2.4}$$

The FCBF algorithm determines the SU between each feature and the target class, C. Features with an SU above a set threshold are added to a list, S. The list, S, is ranked from highest to lowest according to the SU value. The SU between the first feature in the list, f_1, and all of the other features, $f_2 \cdots f_n$ (where n is the number of features in S), is determined. Every f_k, $2 \geq j \geq n$, such that $\mathrm{SU}(f_1, f_k) \geq \mathrm{SU}(f_k, C)$ is considered redundant and is removed from the list. This process is repeated with the next feature, f_2, in S and continues until there are no more redundant features to be eliminated. The features that remain in S after all redundant features are eliminated are returned as the final feature set [86]. The psuedocode for FCBF is presented in Algorithm 1.

Algorithm 1. Fast Correlation-Based Filter [86]

Input:

$S(f_1, f_2, \cdots, f_N, C)$: Labeled Training Samples
δ: user defined threshold

Output:

S_{best}: selected feature set

```
for i = 1 to N do
    SU_temp = SU(f_i, C) for f_i
    if SU_temp ≥ δ then
        add f_i to S_list
Sort S_list in descending order of SU(f_i, C) value
f_j = firstElement(S_list)
while f_j ≠ NULL do
    f_k = nextElement(S_list)
    while f_k ≠ NULL do
        if SU_{j,k} ≥ SU_{k,c} then
            Remove f_k from S_list
        f_k = nextElement(S_list)
    f_j = nextElement(S_list)
S_best = S_list
return S_best
```

Minimal-Redundancy-Maximal-Relevance (MRMR).

The MRMR feature selection method incrementally selects features that have low redundancy with other features and high relevance with the target class [22]. The relevance of a feature set, S, to the target class is defined by the following equation [65]:

$$D(S,c) = \frac{1}{|S|} \sum_{x_i \in S} \mathrm{IG}(x_i|c), \tag{2.5}$$

where $|S|$ is the cardinality of features in S, x_i are individual features in S, and c is the target class. The redundancy of a feature set S is [65]:

$$R(S) = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} \mathrm{IG}(x_i|x_j), \tag{2.6}$$

where $|S|$ is the cardinality of features in S, and x_i and x_j are individual features in S. In general, it is difficult to find the ideal feature set that maximizes

$$\Phi = D(S,c) - R(S), \tag{2.7}$$

but a good feature set may be acquired by incrementally adding features that maximize $D(S,c) - R(S)$. Starting with an empty set, a feature x_j is added to S according to the following criterion:

$$\max_{x_j \in X - S} \left[\mathrm{IG}(x_j|c) - \frac{1}{m-1} \sum_{x_i \in S} \mathrm{IG}(x_j|x_i) \right], \tag{2.8}$$

where $X - S$ is the set of features not currently in S, x_j is a feature not in S, x_i is a feature in S, and $m - 1$ is the number of features in the current feature set [65].

Bhattacharyya Methods.

Many feature selection methods use the Bhattacharyya coefficient to determine the best feature set. The Bhattacharyya coefficient (defined on the range $[0, 1]$) is a measure of the similarity between two probability distributions p and q, and is defined as [32]:

$$B = \sum_{i}^{k} \sqrt{p_i q_i}, \tag{2.9}$$

where p and q are defined to be the probability distributions of a feature, f, over the classes a and b. The Bhattacharyya coefficient measures how effectively f differentiates class a from class b. A lower Bhattacharyya value indicates better separability [32].

The Bhattacharyya coefficient and the related Bhattacharyya distance have been used effectively in a number of feature selection methods [32, 36, 69, 78]. These approaches differ on their use of their Bhattacharyya measure. For instance, the method in [32] returns the set of features that have minimum Bhattacharyya values for any pair of classes. The approach in [69] returns n features that have the lowest sum of all pairwise Bhattacharyya values. However, it has been noted that Bhattacharayya methods do not perform well with highly correlated data [69].

Principal Component Analysis.

Principal component analysis is a method for identifying the vectors of highest variance in the sample space. Identifying these vectors determines the dimensionality that the data can be reduced by eliminating bands that do not correspond to variance in the dataset [60]. Determiming the principal component vectors is accomplished by calculating the sample covariance matrix, C, of the data, X [81]:

$$C = X^T X. \tag{2.10}$$

Then eigenvectors and eigenvalues of C are determined using eigenvalue decomposition. The eigenvectors of C are rearranged according to the magnitude of their eigenvalues. The k eigenvectors that correspond to the k highest eigenvalues $(\lambda_1 \cdots \lambda_k)$ are the columns of a matrix A [81]. The principal component matrix S is calculated as:

$$S = A^T X, \tag{2.11}$$

where the columns of S are called the principal component vectors. To evaluate whether a feature is relevant to the distribution of the class represented by X, the sum

$$b_i = \sum_{j=0}^{k} v_{ji} \tag{2.12}$$

is computed for each component, i, where v_{ji} is the ith component of the jth eigenvector. The highest b_i values correspond to the features that are most relevant to the distribution of X [75].

Support Vector Machine - Recursive Feature Elimination.

Support Vector Machine - Recursive Feature Elimination (SVM-RFE) is an embedded method that uses the Support Vector Machine (SVM) classifier (see Section 2.3) to identify features that are highly weighted in the SVM [24]. The algorithm initializes with all available features, and trains an SVM on those features. The feature with the lowest weight is eliminated from the feature set. This process repeats with the reduced feature set. This elimination process continues until only the desired number of features remain.

Relief/Relief-F.

Relief is a widely-used feature selection method that incorporates Euclidean distance to determine the features that separate near samples of different classes in the feature space. Relief measures the distance between a sample and its near hit (the sample closest to it that shares its class), and its near miss (the sample closest to it that is of a different class). If a feature distinguishes between the sample and its near hit, it does not aid in the separability of the classes, and is given a lower weight. However, if a feature distingushes between the sample and its near miss, it is given a higher weight because it can be used to separate the classes [21] . Relief-F is a variation on standard Relief that determines the k nearest misses and k nearest hits. This allows Relief-F to determine features for multi-class problems [49].

2.3 Techniques for Detection or Classification

Classifiers use a feature set to determine the class of a sample. For binary classification, it is sufficient to determine if the sample belongs to the class of interest or not. For multi-class problems, the feature sets must distinguish between more than two classes. Some approaches for detecting and/or classifiying targets are explained in this section.

Spectral Matching.

Spectral matching is performed on a target spectrum $\boldsymbol{x}$, and a hyperspectral image pixel $\boldsymbol{y}$. Multiple metrics can be used to compute the similarity of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Spectral Angle (SA) is a commonly-used metric that is defined as [70]:

$$\mathrm{SA}(\boldsymbol{x}, \boldsymbol{y}) = arccos\left(\frac{\boldsymbol{x} \cdot \boldsymbol{y}}{||\boldsymbol{x}||\,||\boldsymbol{y}||}\right) \tag{2.13}$$

where $||\boldsymbol{x}||$ and $||\boldsymbol{y}||$ are the L^2 norms of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, and $\boldsymbol{x} \cdot \boldsymbol{y}$ is the dot product of $\boldsymbol{x}$ and $\boldsymbol{y}$. Spectral Information Divergence (SID) is a measure of the difference between the probabilistic distributions defined by the input vectors that is calculated as [70]:

$$\text{SID}(p, q) = \sum_{l=1}^{L} p_l log\left(\frac{p_l}{q_l}\right) + \sum_{l=1}^{L} q_l log\left(\frac{q_l}{p_l}\right), \tag{2.14}$$

where p_l and q_l are the l^{th} elements of spectral vectors normalized to the range $[0, 1]$ [12]. Spectral Gradient Angle (SGA) is determined by finding the SA of the spectral gradient vector of $\boldsymbol{x}$ and of $\boldsymbol{y}$ [70].

Spectral Matched Filter.

A Spectral Matched Filter (SMF) uses the background covariance and the target signature to determine an ideal filter, which maximizes the ratio of the target signature to the background [56]. A linear SMF assumes that every pixel can be modeled as a linear combination of a target signature, s, and background noise, n. Thus the spectral vector of a pixel, x, can be modeled as [56]:

$$x = as + n, \tag{2.15}$$

where a is a scalar attenuation constant associated with the presence of the target signature [56]. The ideal matched filter for a target signature (s) is [61]:

$$h = \frac{C^{-1}s}{s^T C^{-1} s}, \tag{2.16}$$

where C is the covariance matrix of the background clutter.

Support Vector Machines.

Support Vector Machines (SVMs) operate by determining a hyperplane that gives the greatest margin between two classes in feature space. This hyperplane is used to classify a vector according to the location of the vector in relation to the hyperplane [86]. A hyperplane is defined by the set of all input vectors $\boldsymbol{x}$, that satisfy,

$$\boldsymbol{w} \cdot \boldsymbol{x} - b = 0, \tag{2.17}$$

where $\boldsymbol{w}$ is the weight vector that is normal to the hyperplane, and b is the hyperplane's offset from the origin [66]. For a set of n-dimensional data to be fully separable by the parameters $\boldsymbol{w}$ and b, the data samples $\boldsymbol{x}_i \in \mathbb{R}^n$ and their respective class labels $y_i \in \{-1, 1\}$ must be such that:

$$\begin{aligned} \boldsymbol{w} \cdot \boldsymbol{x}_i + b \geq 1 \quad & y_i = 1, \\ \boldsymbol{w} \cdot \boldsymbol{x}_i + b \leq -1 \quad & y_i = -1, \end{aligned} \tag{2.18}$$

where i is the number of the sample [66]. The optimal hyperplane is the hyperplane that has the greatest margin m given by [66]:

$$m = \frac{2}{||\boldsymbol{w}||}. \tag{2.19}$$

Thus, the object of SVM is to find the hyperplane parameters $\boldsymbol{w}$ and b that maximize Equation 2.19 subject to Equation 2.18 [66]. Figure 2.2 shows the concept of hyperplane classification in two dimensions. Line a in Figure 2.2 is the optimal hyperplane because it has the widest margin between members of different classes. The optimal

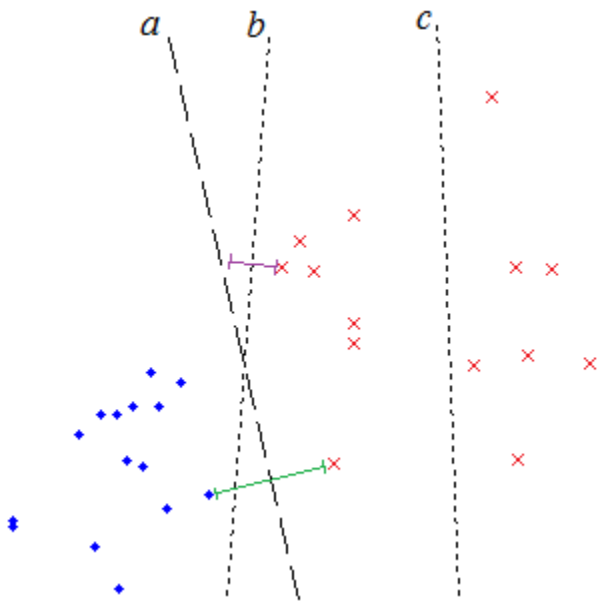


Figure 2.2. Selection of an optimal hyperplane in SVM. Blue diamonds denote members of class 1 and red "x"s denote members of class 2. Line c does not divide the classes. Line b divides the classes, but has a small margin (shown with the purple line). Line a divides the classes with a large margin (shown with the green line).

hyperplane is calculated by minimizing the cost J, defined as [37]:

$$J = (1/2) \sum_{h,k} y_h y_k \alpha_h \alpha_k K(\boldsymbol{x}_h, \boldsymbol{x}_k) - \sum_k \alpha_k, \tag{2.20}$$

subject to [37]:

$$0 \leq \alpha_k \leq C \qquad \text{and} \qquad \sum_k \alpha_k y_k = 0 \tag{2.21}$$

where $\boldsymbol{x}_h$ and $\boldsymbol{x}_k$ are data samples, y_h and y_k are corresponding class labels, α_h and α_k are corresponding Lagrange multipliers, and K is called a "kernel function." The kernel function is used to transform the data space into a higher dimensional space in which the classification problem is better solved [66].

Bayesian Classifiers.

Bayesian classifiers use Bayes' Theorem to determine the posterior probability of a particular class' presence given a measured spectrum. Bayes' Theorem can be represented as [82]

$$P(c_j|x_1, x_2, \cdots, x_n) = \alpha P(c_j) \cdot \prod_{i=1}^{n} P(x_i|x_1, x_2, \cdots, x_{i-1}, c_j), \tag{2.22}$$

where c_j is a classification, x_i are the attributes of a signal, and α is a normalization constant [82]. Bayesian Classifiers differ based on the methods used to estimate the conditional probability shown on the right side of Equation 2.22. The Naive Bayes Classifier assumes that all attributes are independent. When this assumption is applied to Equation 2.22, it yields [27]:

$$P(c_j|x_1, x_2, \cdots, x_n) = \alpha P(c_j) \cdot \prod_{i=1}^{n} P(x_i|c_j). \tag{2.23}$$

However, the assumption that the attributes are independent of each other is not necessarily an accurate model, and can lead to classifier inaccuracy [27]. As a result, alternative Bayesian classifiers make more conservative assumptions.

Multilayer Perceptrons.

Multi-Layer Perceptrons (MLPs) are classifiers that have been used on a variety of classification problems [6, 9, 33]. MLPs are a type of neural network that use only feed-forward connections between layers of the network [35]. A MLP has the basic structure shown in Figure 2.3.

At each node in a MLP, the outputs of the previous layer nodes are multiplied by their corresponding weights, and summed at the nodes of the next layer. The result of this sum of products is the Induced Local Field (ILF). The weights are denoted as

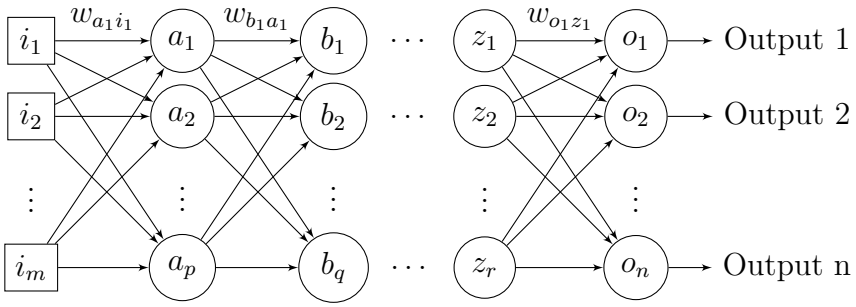


Figure 2.3. A MLP network with a m-dimensional input and n-dimensional output.

w_{ij} where i and j represent the next node and previous node in the directed graph respectively. Not shown in Figure 2.3 is the input bias i_0 and the bias weight of each node, which are also included in ILF calculation. For example, the ILF of node a_1 in Figure 2.3 is calculated as [41]:

$$v_{a_1} = i_0 w_{a_1 bias} + i_1 w_{a_1 i_1} + i_2 w_{a_1 i_2} + \cdots + i_m w_{a_1 i_m}, \tag{2.24}$$

where $w_{a_1 bias}$ is the weight of the bias at node a_1, and $w_{a_1 i_1} \cdots w_{a_1 i_m}$ follow the same naming convention [41]. The output of node a_1 is $\phi(v_{a_1})$, where ϕ is the activation function or transfer function of the node.

The outputs of all other nodes are calculated similarly. A calculation of the outputs of an MLP is called a forward pass.

To train a MLP, an algorithm called back-propagation is used to iteratively update all of the weights of the network. The backpropagation used in this thesis is Levenberg-Marquardt (LM) backpropagation. LM backpropagation is an adaptation of the LM method of finding solutions to least-squares problems. The weight update equation for LM backpropagation is [38]:

$$\mathbf{w} = \mathbf{w} + [J^T(\mathbf{w})J(\mathbf{w}) + \mu I]^{-1} J^T(\mathbf{w})E(\mathbf{w}), \tag{2.25}$$

where $\mathbf{w}$ is the vector of weights, $J(\mathbf{w})$ is the Jacobian matrix, μ is called the damping factor, and $E(\mathbf{w})$ is a matrix of output errors associated with the weights $\mathbf{w}$.

The elements of the Jacobian matrix are [38],

$$J = \begin{bmatrix} \frac{\partial \mathbf{e}_1(\mathbf{w})}{\partial w_1} & \frac{\partial \mathbf{e}_1(\mathbf{w})}{\partial w_2} & \cdots & \frac{\partial \mathbf{e}_1(\mathbf{w})}{\partial w_m} \\ \frac{\partial \mathbf{e}_2(\mathbf{w})}{\partial w_1} & \frac{\partial \mathbf{e}_2(\mathbf{w})}{\partial w_2} & \cdots & \frac{\partial \mathbf{e}_2(\mathbf{w})}{\partial w_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{e}_N(\mathbf{w})}{\partial w_1} & \frac{\partial \mathbf{e}_N(\mathbf{w})}{\partial w_2} & \cdots & \frac{\partial \mathbf{e}_N(\mathbf{w})}{\partial w_m} \end{bmatrix}, \tag{2.26}$$

where $w_1 \cdots w_m$ are the elements of the vector of weights $\mathbf{w}$, and the vectors $\mathbf{e}_1 \cdots \mathbf{e}_N$ are rows of the error matrix,

$$E = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_N \end{bmatrix} = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,n} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N,1} & e_{N,2} & \cdots & e_{N,n} \end{bmatrix}. \tag{2.27}$$

The element $e_{a,b}$ of E is the difference between the desired and actual values of the b^{th} output of the network with the a^{th} training sample [38].

2.4 Spectral Properties of Textiles

A textile is a woven material consisting of strands of natural or artificial fibers [17]. Textiles assume many appearances, differing in density, fiber composition, and other factors [17, 59]. These factors affect the way that a textile reflects electromagnetic

energy, leading to unique spectral properties [29, 59]. The uniqueness of a textile sample allows it to be identified among other textiles. It has been shown that, given a constant signal-to-noise ratio, a particular clothing sample spectrum is more identifiable among other clothing samples than a particular skin sample spectrum among other skin samples [44]. As such, the spectral properties of textiles can be used to detect dismounts.

Composition.

Commonly used plant fibers are cotton, rayon, flax, and hemp. Cotton and rayon are composed of cellulose, a natural polymer that composes about 30% of bushes and 40-50% of woods [1]. Flax and hemp are bast fibers, which are made up of plant material surrounding the plant stem [48]. Methods of natural textile processing such as mercerization, which enhances luster and strength of cotton fiber, influence target spectra depending on their abundance [29].

Animal fibers, including wool, fur, and silk, are also common in the composition of textiles. Each is composed of protein fibers that are in turn composed of amino acids. The protein structures of animal fibers are unique to the animal that produced them, however all are built upon the same selection of amino acids [2].

Some of the most commonly used textiles in the world are comprised of synthetic fibers. These include polyester, acrylic, nylon, and spandex. Artificial textile spectra are influenced by the chemical properites such as the polymer type and the processing type [29].

Even among textiles of the same material composition, such as 100% polyester, there is a significant amount of variance between spectral signatures [40]. This variance can be attributed to the various patterns and colors in which textiles are manufactured.

Chemicals used in the production of textiles may also impact textile spectra. Dyes, which have wide use in the textile industry, significantly affect textile spectra. However, this effect is largely limited to the visible spectrum, and does not expand into the NIR/SWIR spectrum [19]. Synthetic fibers often have a finish applied to them during manufacturing [29]. The spectral characteristics of fire retardants and antibacterial treatments used in textile production have also been investigated [25, 46].

The spectral characteristics of a textile may be used to determine the ratio of fiber compositions used in textile production. This has been shown for blends of plant and animal fibers [83], and blends of plant and synthetic fibers [29, 58].

Environment.

Different types of textiles may be more difficult to detect as a result of background spectra. Textiles composed of cotton and rayon, which have spectral similarities to background vegetation, are generally more difficult to detect than animal fibers such as wool and artificial fibers such as polyester [77].

Chemicals used to maintain clothing such as detergents and fabric softeners have been shown to alter the color characteristics of textiles. Some softeners tend to cause yellowing in white textiles when they are heated [64].

Identical textile materials may have different spectral properties due to their surrounding environment. A textile swath can allow light to transmit to layers beneath it, making the resulting spectrum a combination of the textile and the lower layer [45]. The transmittance of textiles has led to the investigation of the possible use of hyperspectral imaging to detect improvised explosive devices (IEDs) underneath layers of clothing [18]. The effect of moisture in textile material has also been investigated, and has been shown to cause a uniform reduction in reflectance throughout the visible range [20].

Atmospheric chemistry can alter the spectral characteristics of textiles. In high-pollution areas, high concentrations of nitrogen oxides in the air can cause yellowing in clothing [64].

III. Methodology

A spectral textile detection method would increase the effectiveness of dismount detection systems. Feature selection and classification methods capitalize on the abundance of information generated by Hyperspectral Imagers (HSIs) to reliably spectrally detect textiles. This chapter explains the methodology used to investigate the performance characteristics of feature selectors and classifiers in detecting textiles using hyperspectral data.

3.1 Data Sets

The hyperspectral data used in this thesis consists of both contact data and remotely-sensed data. Contact data is collected using a sensor that has physical contact with the target, while remotely-sensed data is collected at an unspecified standoff range from the target. Contact data negates the atmospheric and scattering effects associated with remotely-sensed data. Therefore, contact data is considered a true measurement of an object's spectral signature. However, a spectral detector's ability to classify contact data is not an accurate representation of its performance with remotely-sensed data. An accurate spectral textile detector must be capable of detecting textiles even with the atmospheric effects inherent in remotely-sensed data. Figure 3.1 shows the significant differences between contact and remotely-sensed spectra of the same material, which is attributable to the unique illumination, noise, and atmospheric effects present in the scene [14].

It is desirable to have a classification methodology in which a set of contact textile reflectance samples are used to train the classifier, as it avoids the time-consuming and impractical process of locating and extracting data from full textile pixels in a hyperspectral image. Once trained on the contact samples, a classifier can identify the

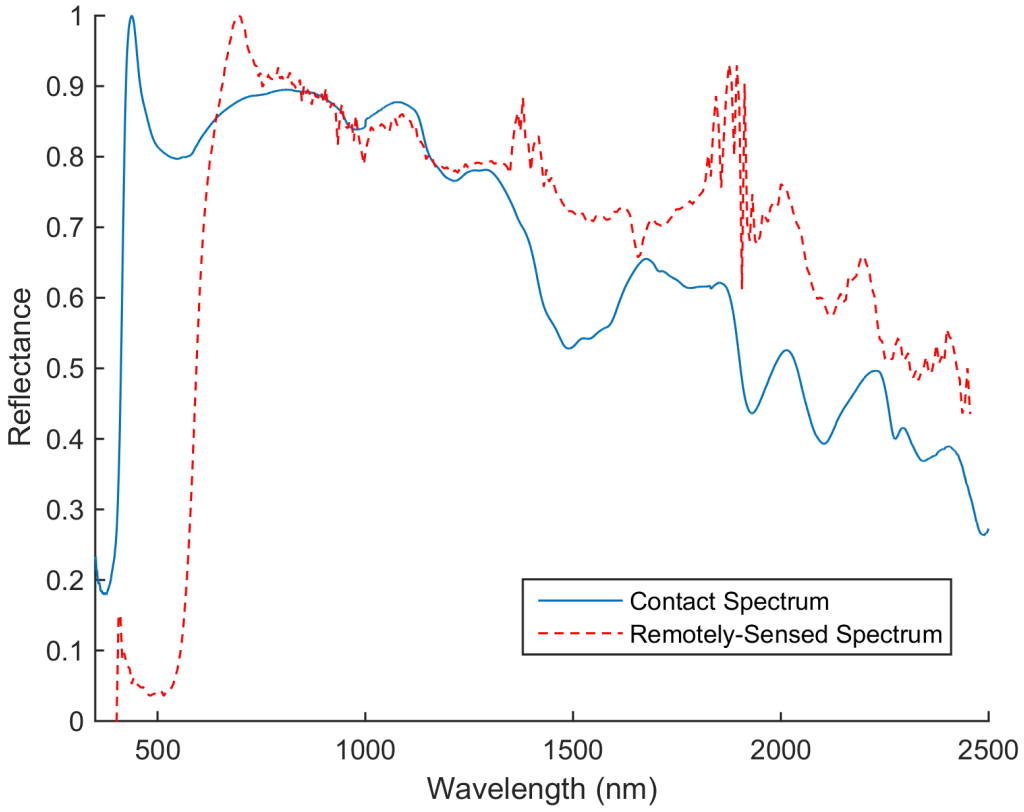


Figure 3.1. Comparison of contact and remotely-sensed normalized reflectance data of the same textile swath (a red cotton shirt). The spectrum collected using a contact probe is shown in blue (solid line), while the spectrum collected with a remote sensor is in red (dashed line). The jagged remotely-sensed curve is the result of illumination and atmospheric effects that are not significant in the contact data.

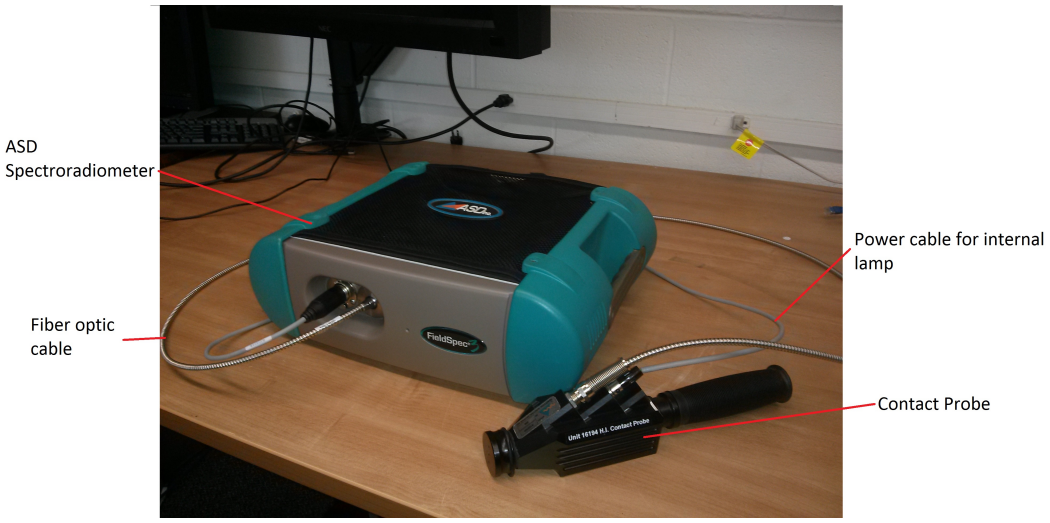


Figure 3.2. The ASD Fieldspec® 3 spectroradiometer and contact probe. The power cable for the halogen light source and the fiber optic cable are shown connected from the spectroradiometer to the contact probe.

pixels in a hyperspectral image that contain textiles, provided that the classifier has sufficient generalization ability to accomodate illumination and atmospheric effects.

Contact Data Collection.

Contact data is collected using an Analytical Spectral Devices (ASD) Fieldspec® 3 spectroradiometer [51]. The ASD Fieldspec® 3 (shown in Figure 3.2) measures radiance from 350nm to 2500nm, with a sampling interval of 1.4nm in the 350-1000nm range and a sampling interval of 2nm in the 1000nm-2500nm range. The full width at half maximum (FWHM) spectral resolution is 3nm at 700nm, 10nm at 1400nm, and 10nm at 2100nm.

The ASD spectroradiometer contact probe Model A122300 (shown in Figure 3.2) is a special foreoptic that collects data from surfaces of a swath via direct contact. The contact probe is a handheld device with a handle, internal lamp, and aperture. During data collection, electromagnetic energy from the lamp passes through the transparent aperture, and is reflected by the swath's surface into the fiber optic

cable. The energy passes through the fiber optic cable into the spectroradiometer, where it is processed into spectral reflectance data. ASD RS3^{TM}[50], a proprietary data processing software, is used to execute data collection. RS3^{TM} allows the user to specify a number of samples to be collected consecutively. For this research, 10 samples were collected consecutively from each of 79 textile swaths and 80 non-textile swaths.

The method of data collection from textile materials differs slightly depending on the thickness of the textile materials. Thicker materials are folded 1-2 times and laid flat on a table before data collection. Thinner materials had an increased risk of allowing electromagnetic radiation to pass through the material and reflect off of a background surface. Therefore, thinner materials were folded 3-5 times and laid flat onto a Spectralon black reflectance panel to minimize background reflectance.

Most non-textile spectra in the data set are collected using the ASD FieldSpec® 3's Ergonomic Pro-Pack, allowing the contact probe to be used on objects such as trees and external building surfaces. Some non-textile swaths had nonuniform contours that rendered consistent orientation of the contact probe in relation to the swath surface impractical. The ASD contact probe is pressed onto the swath surface such that the probe's aperture lay parallel to the surface.

Remotely-Sensed Data Collection.

An AisaDUAL hyperspectral sensor array is used to collect remotely-sensed hyperspectral data. AisaDUAL contains two sensors: an AisaHAWK sensor, which collects radiance in the range 400nm-970nm, and an AisaEAGLE sensor, which collects radiance in the range 970nm-2450nm. Each sensor is a line scan camera that produces images by panning across a scene. The AisaDUAL sensors are set in a rotating enclosure that allows the sensor apertures to be panned, thereby allowing the

Figure 3.3. A sensor similiar to the AisaDUAL hyperspectral sensor. A SWIR line scan camera (left) and a VNIR line scan camera (right) are contained in the rotating enclosure.

sensors to create image data of a scene. A sensor similar to the AisaDUAL in its rotating enclosure is shown in Figure 3.3.

The slight overlap in the spectral range between the sensors allows a set of wavebands (950nm-1050nm) in which the processed data cube contains reflectance information from both the AisaHAWK and AisaEAGLE. Due to the horizontal offset of the sensor apertures, the image cube in the range 950nm-1050nm contains offset copies of a scene, rendering those wavebands impractical for detection purposes.

3.2 Contact Data Pre-Processing

The contact spectral samples are processed and converted to reflectance using ASD ViewSpec^{TM} Pro [52], a proprietary post-processing software. ViewSpec^{TM} Pro performs cubic spline interpolation to produce a reflectance curve with a data point at every 1nm wavelength (350nm, 351nm, $\cdots$, 2500nm). The interpolated reflectance samples are imported into MATLAB®.

Not all wavebands in the 350nm-2500nm are used for spectral textile detection. The wavebands from 350nm - 800nm are associated with the visible spectrum, i.e. color, which is not relevant to the detection of textiles, as dyes can be used to make textiles any color. Atmospheric attenuation also prevents electromagnetic energy from reaching a remote sensor. Wavebands in the ranges 1350nm-1430nm and 1800nm-1950nm have significant atmospheric attenuation characteristics [5]. Although atmospheric attenuation has little effect on data collected with the contact probe, it renders the wavebands unusable in a practical remote sensing enviromnment. The wavebands 350nm-800nm, 1350nm-1430nm and 1800nm-1950nm are removed from each sample in the data to decrease computation time and allow only practically useful features to be selected in feature selection.

It is desired to produce classifiers that can classify the remotely-sensed data collected for this thesis. The wavebands 950nm-1050nm are unusable in the remotely-sensed data due to the sensor offset problems described in Section 3.1. In addition, bands in the range 2455nm-2500nm cannot be collected by the AisaDUAL, as these bands lie outside its operating range. Thus the wavebands 950nm-1050nm and 2455nm-2500nm are removed from the contact data set. The removal of these wavebands prevents the feature selection methods from selecting one or more wavebands that are unusable with the remotely-sensed data.

Most commerically available HSIs have high spectral resolution, but they do not yet yield spectral data with a spectral resolution of 1nm. For example, the AisaHAWK and AisaEAGLE imagers used in this research produce hyperspectral images with a resolution of 2.9 nm to 8.5 nm. Because HSIs cannot take advantage of the high sampling rate of the contact data set, the contact data set is downsampled by a factor of five. Downsampling is accomplished by retaining only the reflectance measurements corresponding to wavelenths that are multiples of 5nm. Thus the first 3 wavebands in

the set are the bands centered on 800nm, 805nm, and 810nm. The downsampling has the additional effect of dimensionality reduction, which reduces computation time for feature selection processes.

Each reflectance sample $\mathbf{r}$ is normalized, producing a normalized reflectance sample $\mathbf{r}_n$. Two normalization methods are applied in this thesis. The first normalization method is division by the maximum, where $\mathbf{r}_n$ is calculated through the relation [79]

$$\mathbf{r}_n = \frac{\mathbf{r}}{r_{max}}, \tag{3.1}$$

where r_{max} is the maximum value in $\mathbf{r}$. The second normalization method is division by the L^2 norm, in which $\mathbf{r}_n$ is calculated using [54]

$$\mathbf{r}_n = \frac{\mathbf{r}}{\sqrt{\sum_{k=1}^{K} r_k^2}}, \tag{3.2}$$

where K is the number of elements in $\mathbf{r}$ and r_k is the k^{th} element in $\mathbf{r}$. The methods in Equation 3.1 and Equation 3.2 are hereafter referred to as "max-normalization" and "L^2-normalization" respectively.

The contact data set is separated into two subsets: a training/testing data set and a generalization data set. All 10 samples of each swath in the data set were placed together in either the training/testing data set or the generalization data set. Both textile and non-textile swaths are distributed between the training/testing and generalization data sets such that each set contains a wide variety of materials. However, none of the swaths represented in the training/testing data set are represented in the generalization data set, and vice versa. A list of swaths represented in the training/testing data set and generalization data set is provided in Appendix A.

The generalization data set is left out of the feature selection and classifier training process. This allows detector accuracy on the generalization data set to be a measure

of generalization accuracy.

Some swaths of textiles have identical material compositions to others in the data set. For example, 13 textile swaths in the contact data set were composed of 100% cotton. It is desirable to measure the performance of textile detectors on spectral samples of textile materials with material compositions that the detectors are trained with. Thus the 13 100% cotton swaths were distributed with a rough 2:1 ratio in the training/testing set and the generalization set, respectively. Distribution between the training/testing set and the generalization set is performed for other abundant material compositions such as 100% nylon and 100% polyester. It is also beneficial to determine textile detector performance on material compositions that the detectors are not trained on. To this end, the generalization set contains some material compositions that are not represented in the training/testing set, such as 100% wool and 100% acrylic. The generalization set therefore contains samples from textile swaths of material compositions that are present in the training/testing set, and samples from textile swaths of material compositions that are absent in the training/testing set.

3.3 Noise Addition

The data collected by the ASD contact probe lacks the noise present in remotely-sensed hyperspectral data. To simulate data representative of remotely-sensed hyperspectral data, noise is artificially added to the contact data. To create noise representative of a hyperspectral image, a model for noise as a function of wavelength is developed. All noise in each waveband is assumed to be Gaussian with a mean of 0 and a variance σ dependent on the wavelengths of electromagnetic energy unique to the waveband. Thus, to create a noise model, it is sufficient to find the noise variance in each waveband.

Figure 3.4. A color representation of the hyperspectral image used to determine noise variance. The Spectralon white reflectance panel (indicated by a red arrow) is on the left.

Hyperspectral image data of objects with known reflectance is used to accurately determine the noise variance in each waveband. In general, the true reflectance of an object in a hyperspectral image is not known, as the reflectance signature of the object(s) occupying the pixel is influenced by electromagnetic noise. However, if the reflectance of an object in a hyperspectral image is known, then the standard deviation of reflectance measurements across multiple pixels of the object emulate noise standard deviation. In the hyperspectral image used to calculate the noise variance (shown in Figure 3.4), a NIST-certified Spectralon white reflectance panel is present. The white reflectance panel is chosen as the object for noise standard deviation calculation due to its uniform lighting conditions and reflectance. Spectralon has a known reflectance of 0.99 - 1.00 throughout all wavelengths in the VNIR-SWIR range. To calculate an estimate of the noise variance at a given wavelength, the sample variance of reflectance measurements at that wavelength for all pixels fully occupied by the panel is calculated. The result is a function $\sigma(\lambda)$, the noise standard deviation as a function of wavelength.

The noise vector $\boldsymbol{n}$ is modeled as a vector of independent normal random variables with mean zero and varying standard deviations,

$$\boldsymbol{n} = [\mathcal{N}(0, \sigma(\lambda_1)) \quad \mathcal{N}(0, \sigma(\lambda_2)) \quad \cdots \quad \mathcal{N}(0, \sigma(\lambda_M))], \tag{3.3}$$

where $\mathcal{N}(0, \sigma)$ is the normal random variable with mean 0 and standard deviation σ, and $\sigma(\lambda_1) \cdots \sigma(\lambda_M)$ are the standard deviations of noise at each waveband in the contact data.

A noisy sample is generated by summing the sample vector with a randomly generated noise vector. Noise vectors are generated independently for each sample.

3.4 Classification Algorithms

Textiles vary widely in their spectral characteristics (see Section 2.4). It is desired to detect all textiles regardless of their chemical composition or production method. In the case of textiles, a single "target signature" cannot be identified, rendering spectral matching classifier impractical. It is also impractical to use multiple binary classifiers to search a scene to detect different textile materials, e.g. cotton, polyester, and nylon independently. This thesis is concerned with identifying all textiles, which renders such a methodology unnecessary. Instead of relying on a single target signature to perform classification, classifiers investigated in this thesis perform supervised learning on a set of training data to determine the characteristics of textiles.

The classifiers used in this research are Support Vector Machines (SVMs) and Multi-Layer Perceptrons (MLPs). SVMs have been successfully applied to a number of hyperspectral classification problems [10, 34, 47], as have MLPs [10, 15, 23]. Each classifier is implemented using proprietary MATLAB® functions.

The SVM classifier is implemented using the "svmtrain" function. The "svmtrain"

function allows user selection of the type of kernel function implemented to map to the feature space. The Gaussian kernel (see Table 3.3), also called the radial basis function kernel, is considered the baseline kernel function in this thesis. It is used in the SVMs implemented for Sequential Forward Selection (SFS) feature selection in Section 3.5. Parameter settings for the kernels investigated in this thesis are provided in Table 3.3. Classification decisions with the SVM are decided using the scalar "soft score," which is calculated as [66]:

$$O(s) = \sum_i \alpha_i y_i K(x_i, s) + b \tag{3.4}$$

where $O(s)$ is the soft score of the sample vector s, α_i is the Lagrange multiplier of the i^{th} support vector, y_i is the class of the i^{th} support vector, K is the kernel function, x_i is the i^{th} support vector, s is the sample input vector, and b is the bias (see Section 2.3 for an explanation of these values). $O(s)$ is used to make a classification decision by comparing it to a classification threshold, which is by default set to 0. Therefore, the default rule for deciding the class C of a sample s is

$$C(s) = \begin{cases} 1, & O(s) > 0 \\ 0, & O(s) \leq 0, \end{cases} \tag{3.5}$$

where $O(s)$ is the soft score of the sample vector s.

MLPs have many operating parameters that are not explored in this thesis. The activation function, $\phi(v)$, used by all neurons in the MLPs in this thesis is the hyperbolic tangent function, defined as [68]:

$$\phi(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}, \tag{3.6}$$

where v is the Induced Local Field (ILF) of a node. Unless otherwise stated, the MLP classifiers used in this thesis have five neurons in the first hidden layer, and three neurons in the second hidden layer. This topology is chosen for its compromise between complexity and simplicity, and will be considered the baseline MLP topology for this thesis. All MLP classifiers contain one output neuron, with a single scalar output. This scalar output is the soft score, which is used to make classification decisions on samples. The ideal value of the soft score is "1" for inputs corresponding to textile materials, and "0" for inputs corresponding to non-textile materials. A threshold of 0.5, which lies between 0 and 1, is chosen to be the classification boundary. Therefore, the default rule used to decide the class C of a sample s is

$$C(s) = \begin{cases} 1, & O(s) > 0.5 \\ 0, & O(s) \leq 0.5, \end{cases} \tag{3.7}$$

where $O(s)$ is the soft score of the sample s. The ideal outputs of 0 and 1 for non-textiles and textiles respectively as well as the classification threshold of 0.5 are not standard for the hyperbolic tangent activation function, which has a range of -1 to 1. Performance of the MLPs may be improved by instead having ideal outputs of -1 and 1 for non-textiles and textiles respectively, and setting a classification threshold of 0. However these latter settings were not used in this research. All MLPs are trained with the Levenberg-Marquardt (LM) method (see Section 2.3). In MLP training, there is a danger of "overtraining." Overtraining produces a classifier that is too specialized to its training set, preventing it from performing well on new data. To prevent overtraining, the mean squared error (MSE) on a separate testing set is calculated after each training iteration. The training is stopped when MSE on the testing set fails to improve for six consecutive training iterations. The MATLAB® documentation refers to this procedure as a "validation check" stopping condition

with a maximum validation check value of six.

3.5 Feature Selection

Feature selection methods find subsets of spectral features that accurately encapsulate the unique properties of textiles. Utilizing a reduced feature set that maintains the information relevant to textile classification has two benefits. First, the computation time associated with data manipulation and classification is decreased. Second, a specialized spectral textile detector is simpler and less expensive if less wavebands are required to be sensed. In this research, feature selection is accomplished in MATLAB®. The feature selection methods investigated in this research are FCBF (Section 2.2) and SFS (Section 2.2). Feature selection is performed on both noiseless and noisy versions of the training set.

FCBF Implementation.

FCBF (see Section 2.2) is implemented in MATLAB® using the Arizona State University Feature Selection Repository's fsFCBF script, which in turn uses the WEKA FCBF algorithm. In the FCBF algorithm, the full training set is used for the feature selection process.

SFS Implementation.

The SFS implementation is an original work in MATLAB®. SFS operates by training classifiers with a prospective feature set. It is not sufficient for a feature set to be useful in correctly classifying samples in a training set. Instead, it is necessary to determine a prospective feature set's ability to generalize the spectral properties of all textiles. Thus the training/testing set used for SFS feature selection must be subdivided into a training set and a testing set. The training and testing sets are

generated by randomly distributing the training data, with 80% of the data in the training set, and 20% of the data in the testing set. By calculating the MSE of a classifier on the testing set, a feature set's generalization ability is more accurately estimated.

Because the data are randomly distributed among the training and testing sets, it is possible for a training/testing set pair to be abnormally well-suited or ill-suited for a feature set. If a training set adequately prepares the classifier for a testing set, it can be indicative that the features used in that classifier have good generalization ability. However, it is also possible that the training and testing sets were by chance particularly ideal for that feature set. The latter conditions produce testing accuracy results not typical in the space of possible training and testing sets, causing a feature set's performance to be overestimated. This is not desirable, as it could cause the selection of an arbitrary feature that happens to be compatible with the training and testing sets, rather than a feature with a generally higher expected performance. Generally, this problem is avoided by accomplishing K-fold cross validation. However, it is desired to have a large number of folds so that a feature with the highest average performance is more likely to be the best feature in actuality. Because the training/testing set is so small, performing K-fold cross validation with a high number K makes accuracy on the holdout testing set highly dependent on a small number of samples. Instead, each feature set explored by SFS is evaluated 50 times, each time with a different randomly generated training and testing set with 80% and 20% of the samples, respectively. Multiple calculations of a classifier's performance on the feature set under slightly different conditions produce a better estimate of a feature's value.

SFS is a wrapper method that generates feature sets based on the classifier its feature set is intended to operate with. Therefore, separate feature sets, SFS-SVM

Table 3.1. Parameters for the SVMs used in SFS feature selection.

Parameter	Value
Kernel Function	Gaussian
autoscale	true
boxconstraint	1
kernelcachelimit	5000
kktviolationlevel	0
method	SMO
maxiter	400000
tolkkt	1e-3

and SFS-MLP, are produced. The manner in which the training and testing sets are used within SFS depends on the classifier. When the SVM is trained, it is trained using only the training data, then evaluated using the validation data. The classifier MSE on the validation data is recorded for each of the 10 iterations. When the MLP is trained, it is trained on the training data, and evaluated with the validation data after each training iteration. The continuous evaluation against the validation set allows the stopping condition described in Section 3.4, which prevents overtraining. Tables 3.1 and 3.2 show the operating parameters of the SVMs and MLPs used in SFS feature selection, respectively.

The SFS algorithm adds features to the feature set until degredation of classification accuracy occurs. The pseudocode for SFS used in this thesis is shown in Algorithm 2.

Generation of Varied Feature Sets.

It is desired to determine whether normalization and the presence of noise in a data set influence the effectiveness of the feature set produced by feature selection

Algorithm 2. Sequential Forward Selection Implementation

Input:

$\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}$: Training Samples

$y_1, y_2, \cdots, y_n$: Training Class Labels

Output:

feature set

available $\leftarrow 1, 2, \cdots, m$
current best $\leftarrow 10000$
feature set $\leftarrow []$
while 1 **do**
    for $k = 1$ *to* $length(available)$ **do**
        current feature $\leftarrow$ *available*(k)
        for $t = 1$ *to* 10 **do**
            Generate random training and validation sets
            Train classifier using *feature set* and *current feature*
            Calculate validation MSE
            featureMSE$(t) \leftarrow$ validation MSE
        E(k) $\leftarrow mean($*featureMSE*$)$
    $M = max(E)$
    $I = argmax(E)$
    if $M \leq$ *current best* **then**
        Append *available*(I) to *feature set*
        Remove *available*(I) from *available*
        current best $\leftarrow M$
    else
        break
return *feature set*

Table 3.2. Parameters for the MLPs used in SFS feature selection.

Parameter	Value
Activation Function	Gaussian
Number of hidden layers	2
Number of neurons in first hidden layer	5
Number of neurons in second hidden layer	3
Maximum Epochs	1000
Maximum Validation Checks	6
Training Method	Levenberg-Marquardt
Levenberg-Marquardt μ	0.001
μ Decrease Ratio	0.1
μ Increase Ratio	10

methods. Feature sets are generated for the max and L^2 normalization methods, and for noiseless and noisy training/testing contact data sets. Thus four feature sets are produced with the Fast Correlation-Based Filter (FCBF) feature selection method. Because SFS has the additional two-level parameter of the classifier type (MLP or SVM), eight feature sets result from SFS computation.

3.6 Classifier Optimization

SVMs and MLPs are complex classifiers with numerous operating parameters. The performance of an SVM or MLP can be improved by varying these parameters. In this thesis, the kernel used in the SVM is varied to determine the kernel that produces the best classifier performance. Similarly, the MLP topology is varied to improve performance. Optimization of the classifiers is carried out by maximizing Equal Weighted Accuracy (EWA) for a given operating parameter on the training/testing contact data set.

Table 3.3. Kernel Functions used for optimization [3]. The symbol $\cdot$ indicates a dot product, the $||$ symbols denote an L^2 norm, and *exp* indicates an exponent. The constants p and σ are set by the user. The default values $p = 3$ and $\sigma = 1$ were used in this research.

Kernel Name	Function
Linear	$K(\boldsymbol{x}_h, \boldsymbol{x}_k) = \boldsymbol{x}_h \cdot \boldsymbol{x}_k$
Polynomial	$K(\boldsymbol{x}_h, \boldsymbol{x}_k) = (\boldsymbol{x}_h \cdot \boldsymbol{x}_k + 1)^p$
Gaussian	$K(\boldsymbol{x}_h, \boldsymbol{x}_k) = exp\left(-\lvert\boldsymbol{x}_h - \boldsymbol{x}_k\rvert^2/\sigma^2\right)$

The MATLAB® "svmtrain" function has three options for kernel functions: the "Gaussian" (or "radial basis function") kernel, "linear kernel," and "polynomial kernel." When a kernel function is implemented in an SVM, the equation for that kernel function (shown in Table 3.3) is substituted into Equation 2.20. For each kernel function, the contact training/testing data set is partioned into 5 bins for a 5-fold cross-validation. The 5-fold cross validation process produces 5 SVMs, each with its performance measured in testing EWA. The highest testing EWA score out of the 5 is recorded. This process is repeated 25 times for each kernel so that the Wilcoxon Rank Sum Test (WRST) can be used to show the certainty that one kernel is superior to another in terms of resulting EWA. With the exception of the kernel function, the parameters of the SVMs remain the same as in Table 3.1.

The process for optimizing the MLP is similar to optimization for SVM, with the key difference being the parameter that is varied. In the MLP, the topology (the number of layers of hidden nodes and the number of nodes in each layer) is varied. The space of possible topologies for MLPs is infinitely large, so the highest number of hidden layers explored is 3, and the highest number of nodes in a layer is limited to 6. Every hidden hidden node topology within these maximum constraints is explored. Thus there are six one-hidden-layer topologies explored, 6 * 6 = 36 two-hidden-layer topologies explored, and 6 * 6 * 6 = 218 three-hidden-layer topologies explored, for a

total of 258 topologies explored. With the exception of the topology, the parameters of the MLPs remain the same as in Table 3.2.

The MLP is 5-fold cross-validated 50 times on the contact training/testing data set, each time having the highest validation EWA of the 5 folds recorded. This is performed for all 258 topologies in the explored space. The 50 trials for each topology are used to calculate the mean accuracy of each topology. More repetitions of 5-fold cross validation are required for the MLP because the number of explored network topologies (258) is much larger than the number of explored kernel functions (3). The larger number of explored topologies requires more repetitions to be performed before the best parameter setting becomes obvious. The best topologies are compared using WRST.

A classifier parameter is considered "optimized" when it produces a higher EWA than all other levels of that parameter to a statisically significant margin under the WRST. In some cases, such an optimization does not exist because the EWA produced by two or more levels of the same parameter are statistically identical. In this case, the most simple classifier in the set of statistically identical classifiers will be considered the "optimized" classifier. For example, a single hidden layer MLP with four hidden nodes is selected over a single hidden layer MLP with six hidden nodes, because the former has a less complex topology than the latter. For this thesis, the Gaussian kernel is considered to be the most complex, the polynomial of middling complexity, and the linear kernel the least complex. Thus given statistically identical SVMs, the one implementing a linear kernel is chosen over one with a polynomial kernel, and one with a polynomial kernel is chosen over one using the Gaussian kernel.

For each of the four FCBF feature sets, an optimized SVM and an optimized MLP are produced. For each of the four SFS-MLP feature sets, an optimized MLP is produced. Finally, for each of the four SFS-SVM feature sets, an optimized SVM is

produced. The peformance of all 16 of these optimized classifiers is measured using the generalization data set. Because the generalization data set contains samples from fabric swaths it has not trained on, the EWA from the generalization set will provide a measure of generalization error for each optimized classifier.

3.7 Remotely-Sensed Data Pre-Processing

The hyperspectral image cubes collected by the AisaDUAL sensors are not registered by default, and must be registered before they can be used for detection purposes.

Because the AisaHAWK and AisaEAGLE collect radiance in different wavebands, both are used to create a single hyperspectral data cube. The horizontal offset between the sensor apertures creates a horizontal spatial disparity between the portion of the data cube provided by AisaHAWK and that provided by AisaEAGLE. Thus these portions of the data cube must be registered to provide accurate spectral information. However, the size of the spatial disparity, called parallax, is dependent upon the distance of a subject from the sensor apertures [63]. Figure 3.5 shows the varying effects of parallax on objects of different distances. Because objects in the hyperspectral imagery in this thesis have varying distances from the sensors, it is not possible to register the image data of the entire scene at once. Instead, individual subjects in the scene are selected so that the pixels of those subjects can be registered independently of each other.

Once the data cubes registered, they must be processed so that they are usable for the classifiers produced by Section 3.6. The spectral data in each pixel is cubic spline interpolated to 1nm resolution (the same resolution as the contact data). The remaining processing steps are the same as for the contact data: the bands 350nm-800nm, 950-1050nm, 1350nm-1430nm, and 1800nm-1950nm are removed from each

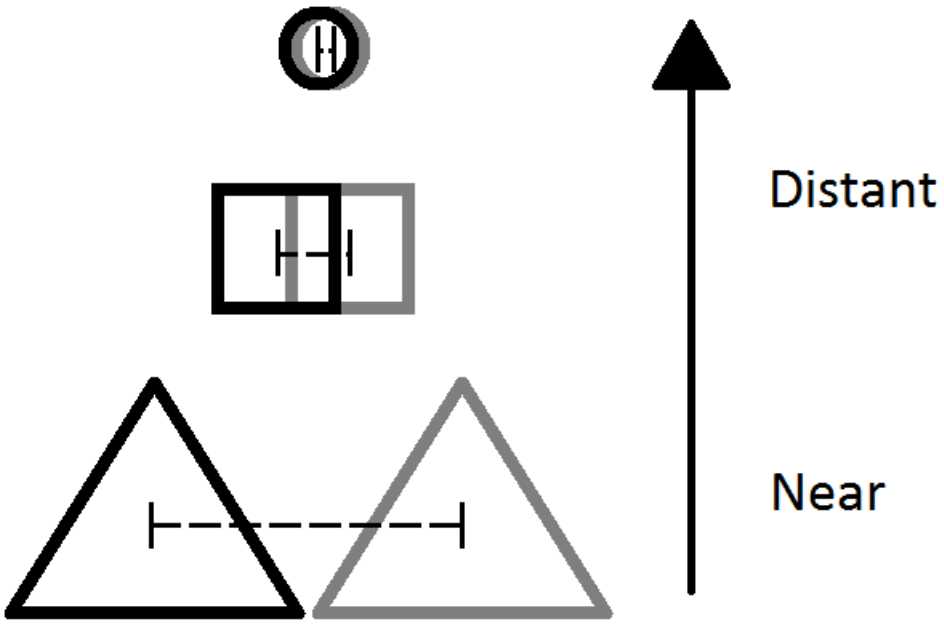


Figure 3.5. The parallax between objects in an image with horizontally displaced sensors. Shapes in black are the apparent positions of objects in the right sensor's image. Shapes in grey are the apparent positions of objects in the left sensor's image. The parallax between objects (indicated by the dashed lines) is larger for closer objects (the triangles) than for farther objects (the circles).

pixel, the data is downsampled by a factor of five, and the data is normalized using either max-normalization (Equation 3.1) or L^2-normalization (Equation 3.2).

3.8 Analysis of Optimized Classifiers

It is desired to analyze the performance of the optimized classifiers on both the generalization contact data set and the remote sensing data set in depth. Section 3.4 shows that classification with both the SVM and the MLP is performed by comparing a scalar soft score with a classfication threshold in Equation 3.5 and Equation 3.7, respectively. Measures of classifier performance such as EWA assume a set threshold for the classifier. However, a classifier's performance can be analyzed in greater depth by evaluating classification accuracy for a range of possible thresholds. Such analysis is achieved with the Receiver Operating Characteristic (ROC) curve. A ROC curve is a plot of a detector's probability of detection (P_D) versus its probability of false alarm (P_{FA}) [4]. ROC curves are produced by varying the classification threshold from the maximum soft score in a data set to the minimum soft score in a data set, producing results at the extremes $P_D = P_{FA} = 0$ and $P_D = P_{FA} = 1$. An example of a ROC curve is shown in Figure 3.6.

Because it is desired to have a classifier with a simultaneously high P_D and low P_{FA}, a detector is considered better the further up and to the left its ROC curve passes [4]. To compare ROC curves of different shapes, the Area Under the Curve (AUC) of the ROC curve can be calculated [4]. AUC is bounded from 0 to 1, where a higher value indicates superior detection performance. The optimized classifiers in this thesis are compared using their AUC values.

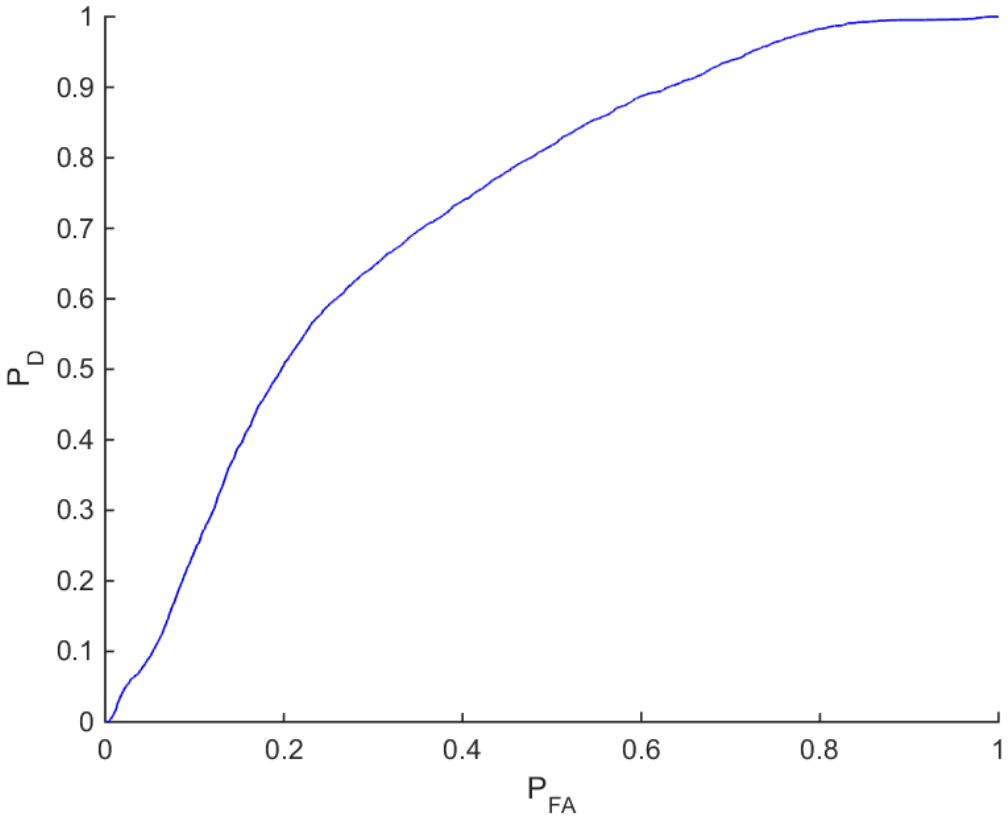


Figure 3.6. A ROC curve. P_D increases as P_{FA} increases.

IV. Results

This chapter presents the feature sets and detection characteristics of textile detectors developed on simulated and real hyperspectral remotely-sensed data. Fast Correlation-Based Filter (FCBF) and Sequential Forward Selection (SFS) feature selection methods are applied to the training/testing set to find suitable wavelengths for accurate classification. Multiple Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) classifiers are tested to determine optimal parameter settings for the classifiers. The performances of the optimized classifiers on a generalization data set and a hyperspectral image are analyzed.

4.1 Data Collection

Both contact and remotely-sensed hyperspectral data of textile and non-textile materials are collected. The ASD Fieldspec® 3 spectroradiometer is used to collect contact spectral measurements of textile and non-textile swaths. The AisaDUAL imager is used to produce a hyperspectral image of an outdoor scene with dismounts present.

Contact Data Collection.

Reflectance measurements of 80 textile and 79 non-textile swaths are collected using the Analytical Spectral Devices (ASD) spectroradiometer. Sample reflectance curves over the range 350nm to 2500nm for selected swaths (both textile and non-textile) are shown in Figure 4.1. Among the 80 textile swaths measured, 45 different textile compositions are represented. Multiple swaths of more common textile compositions are included to characterize the varying spectral signatures produced by different processing and dyeing techniques. Examples of some common types of

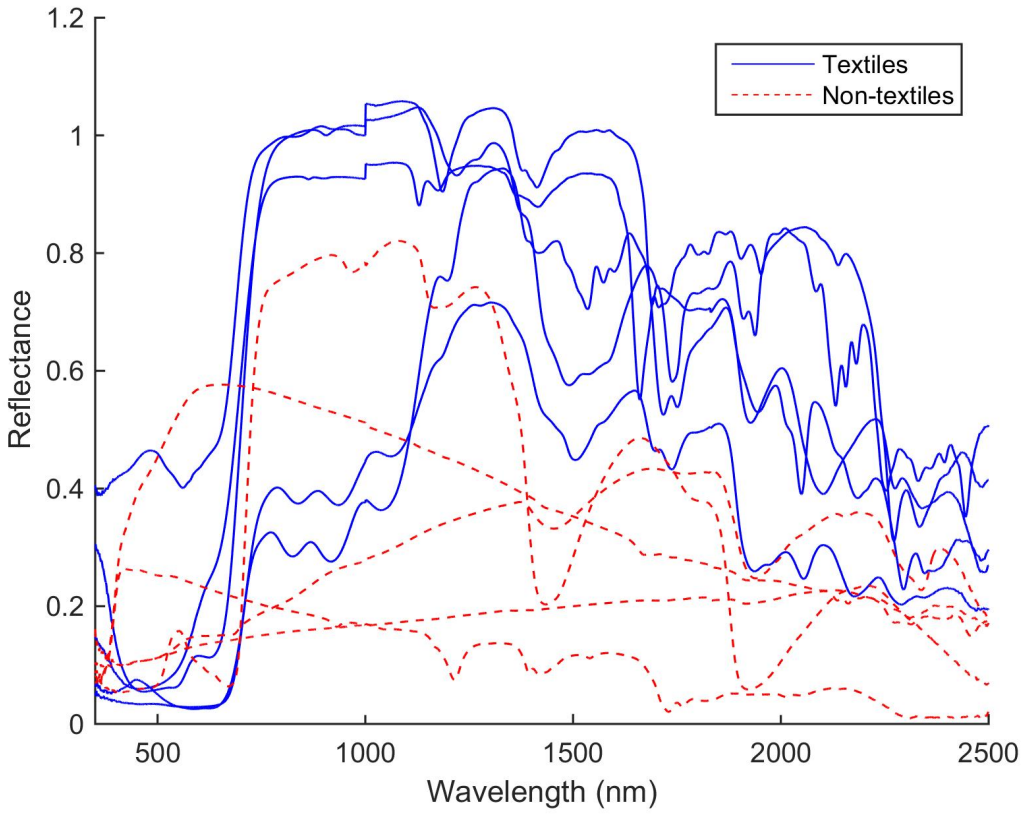


Figure 4.1. Reflectance curves for select swaths measured using a contact probe and the Fieldspec® 3 spectroradiometer. Curves corresponding to textiles (cotton, polyester, nylon, acrylic, and wool) are shown in blue (solid) lines, while curves corresponding to non-textiles (asphalt, grass, plastic, metal, and rock) are shown in red (dashed) lines.

textiles included in the data set were cotton, polyester, and wool. Exact material compositions of non-textiles were unavailable. 79 non-textile swaths representing 13 common materials compose the non-textile data set. Some common non-textiles included wood, rocks, grass, plastic, and metal. Ten samples are collected from each swath measured, creating a total of 10 (79+80) = 1590 samples of spectra in the data set.

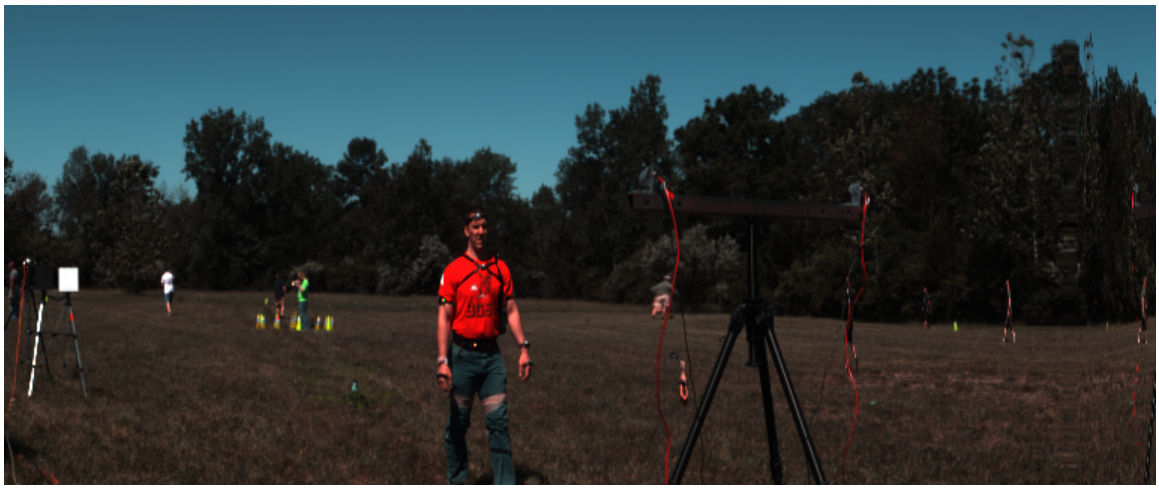


Figure 4.2. A color representation of the hyperspectral image used for detection in this thesis.

Remotely-Sensed Data Collection.

On 4 June 2013, a hyperspectral data collect with the AisaDUAL sensor was performed. Participants in the data collect were asked to walk in a predetermined pattern in an outdoor environment. At timed intervals, the participants were asked to stop and remain motionless so that the AisaDUAL sensors could pan across the scene to create a hyperspectral image. The hyperspectral image used in this thesis for classification is shown in Figure 4.2.

The image shows a woodland scene, with trees in the background and grass in the foreground. Eight dismounts are present in the scene, two of which are obscured by objects in the foreground. The remaining six dismounts are described as follows: a caucasian male with a red shirt is located in the foreground; a pair of dismounts surrounded by green traffic cones are located in the middleground; a dismount with a white shirt and blue shorts is in the background on the left; and two dismounts in the middleground/background are located to the right of the metal tripod in the foreground.

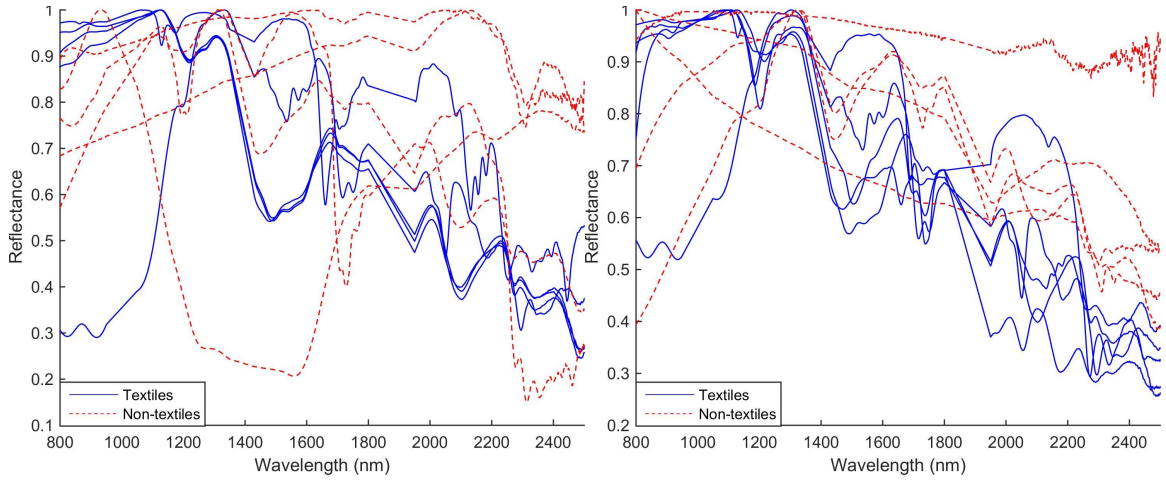
4.2 Data Pre-Processing

The data sets presented in Section 4.1 require pre-processing before they are used in the feature selection and classification processes. Interpolation, band elimination and normalization are used to standardize the data within the contact and remotely-sensed data sets.

Contact Data Pre-Processing.

The contact data is processed by eliminating portions of the reflectance curves corresponding to the visible wavebands 350nm-800nm and the atmospherically attenuated wavebands 1350nm-1430nm and 1800nm-1950nm. The wavebands in the range 950nm-1050nm are excluded due to scene overlap in the image cube in those wavebands (see Section 3.1), and the wavebands in the range 2455nm-2500nm are excluded to prevent feature selection of bands that lie outside the range of the AisaDUAL sensors.

After noise addition and normalization of the data set, the samples are split into a testing/training data set and a generalization data set. The generalization set contains approximately 38% of the original data set, and is comprised all 10 samples from 30 textile swaths and 30 non-textile swaths, for a total of 600 samples. The testing/-training set is used for feature selection and for optimization of the classifiers. The generalization set is withheld until after classifier optimization to determine generalization accuracy of the optimized classifiers. Samples of the normalized reflectance curves of the training/testing set and the generalization set are shown in Figure 4.3a and Figure 4.3b respectively.



(a) Training/Testing Set (b) Generalization Set

Figure 4.3. Samples of the training/testing and generalization data sets. Reflectance curves corresponding to textiles are shown in blue (solid) lines, and curves corresponding to non-textiles are shown in red (dashed) lines.

To calculate the noise standard deviation in a hyperspectral image as a function of wavelength, the known constant reflectance of a NIST-certified Spectralon white panel is measured with the AisaDUAL sensors. The variance between the pixels of the Spectralon panel collected had the same incident radiance conditions. The only differences between each pixel are attributed to atmospheric noise. Thus the noise variance in a given waveband is determined by calculating the variance in reflectance between the white Spectralon panel pixels in the waveband. The standard deviation of the noise as a function of wavelength is shown in Figure 4.4. Noise with the standard deviation shown in Figure 4.4 is added prior to normalization of the spectral samples.

Figure 4.4 indicates that the noise standard deviation varies greatly as a function of wavelength. Standard deviation initially decreases from 402nm (the smallest wavelength read by the sensor). However, the standard deviation begins an upward trend in the range 750nm through the maximum of 2455nm. The large spikes in the standard deviation shown in Figure 4.4 in the ranges 1350nm-1430nm and 1800nm-

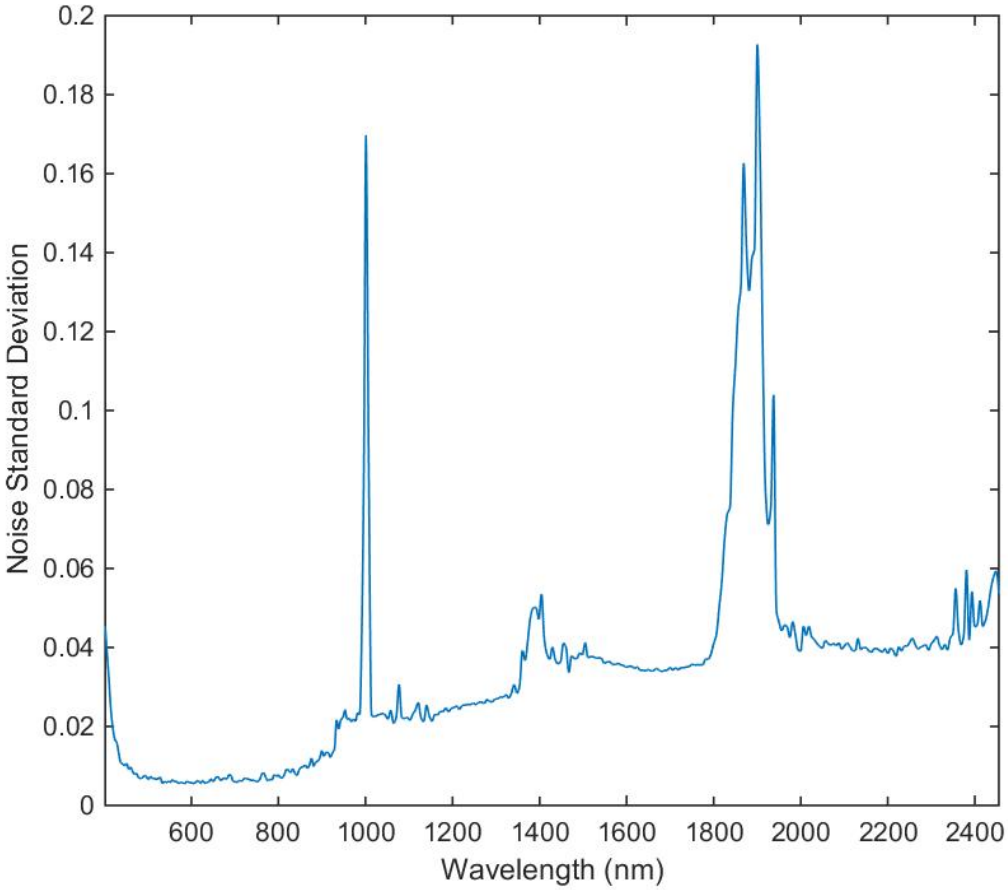


Figure 4.4. Noise standard deviation versus wavelength, calculated using Spectralon reflectance from hyperspectral image in Figure 3.4.

1950nm are due to the significant effects of atmospheric attenuation, while the spike at 950nm-1050nm is caused by the sensor's registration issues at that waveband.

To use the noise standard deviation curve, shown in Figure 4.4, with the contact data set, the curve is interpolated using a cubic spline method to the 1nm resolution of the FieldSpec® 3. Noise vectors are individually generated using Equation 3.3 and added to each contact data sample to produce a noisy contact data set.

Remotely-Sensed Data Pre-Processing.

The hyperspectral data cube represented in Figure 4.2 must be registered prior to detection. Registration is required because the horizontal offset between the image apertures (see Figure 3.3) causes parallax in the data cube.

Areas of interest in the scene are selected based on their material composition, distance from the sensor, and exposure to the sensor. The portions of the data cubes corresponding to areas of interest are independently registered for detection purposes, leaving the rest of the image unused for detection. The six dismounts described in Section 3.1 are chosen as areas of interest. It is desirable to have areas of interest without textiles present as part of remotely-sensed data set. Thus a patch of grass in the bottom right of the image, a portion of the metal tripod in the foreground, and the white reflectance panel on the left are also selected as areas of interest. Areas of interest are shown in Figure 4.5.

Because it is desired to analyze the accuracy of optimized classifiers on the hyperspectral image, the image data is ground-truthed by hand. As with the contact data set, pixels that are occupied by textiles are labeled with a "1," while pixels that are occupied by non-textiles are labeled with a "0." Pixels of a hyperspectral image can be occupied by more than one material in cases where a pixel's Field of View (FOV) is larger than the objects present in the pixel. Thus, unlike the samples of the con-

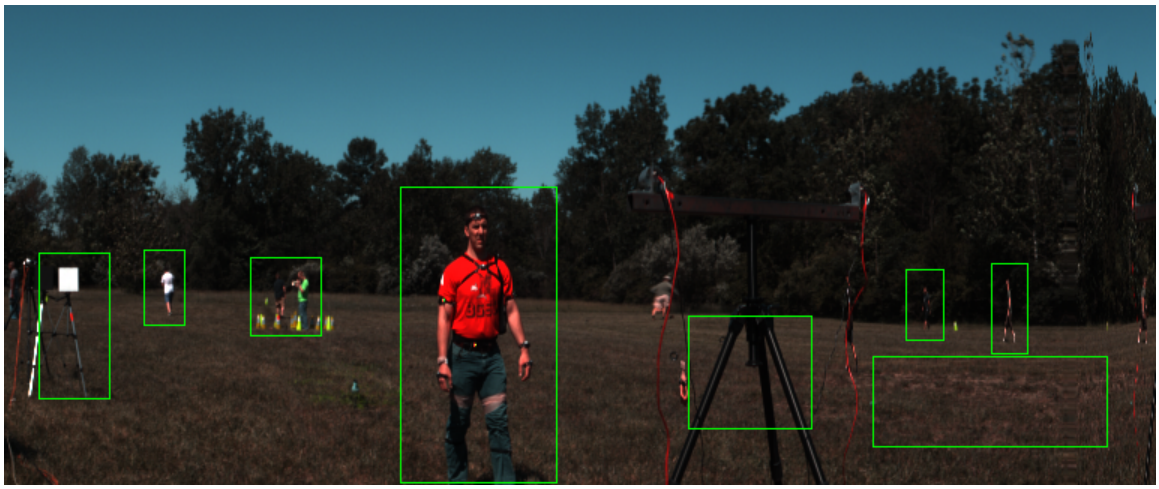


Figure 4.5. A color representation of the hyperspectral image used for detection in this thesis with areas of interest are outlined in green.

tact data set, the pixels of the hyperspectral image may contain spectra from both textile and non-textile materials. It is therefore necessary to determine whether these "mixed" pixels are considered textiles or non-textiles. For this research, a pixel is only given a textile label if close examination of the pixel indicates that it is mostly occupied by textile materials.

A mask that illustrates the labeling of pixels in the hyperspectral image is shown in Figure 4.6. Classifier performance is determined only using the areas of interest, so Figure 4.6 only shows white pixels for textile materials within those areas.

4.3 Feature Selection

The feature selection methods, FCBF and SFS, are performed using only the training/testing contact data set. Because the features in the ranges 350nm-800nm, 950-1050nm, 1350nm-1430nm, 1800nm-1950nm, and 2455nm-2500nm are excluded from the data set, they cannot be selected by FCBF or SFS.

For FCBF selection, four unique feature sets are produced: one corresponding to a noiseless data set normalized by the "max" normalization method; one corresponding

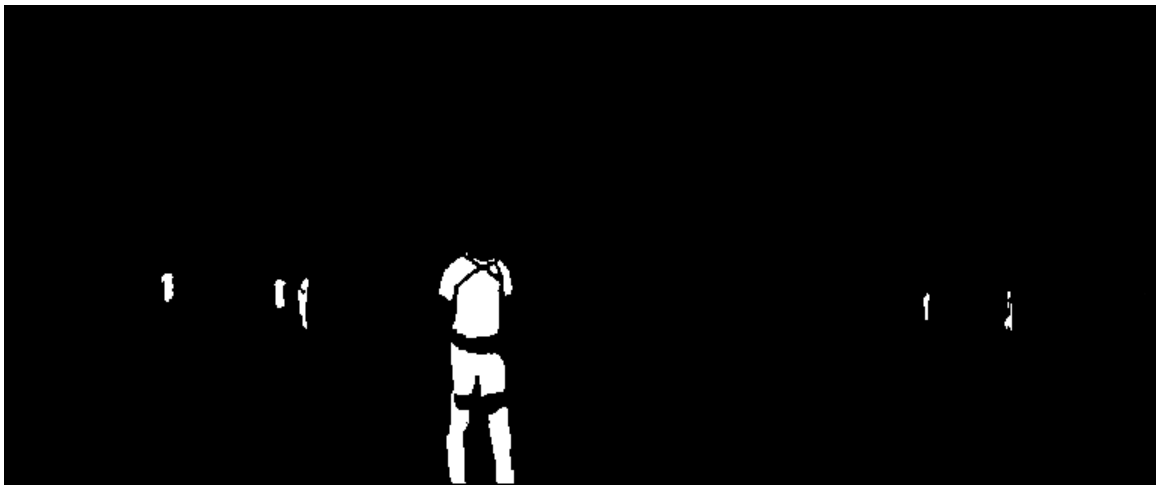


Figure 4.6. A truth mask of the pixels of the hyperspectral image. Black pixels indicate non-textiles, while white pixels indicate textiles.

Table 4.1. FCBF Feature Sets

Normalization Type	Data type	Feature Set wavelengths (nm)
Max	Noiseless	2425, 1060
	Noisy	1320, 2160, 815, 1965
L^2	Noiseless	1185
	Noisy	1195, 2000, 1790, 1650

to a noisy data set normalized by the "max" normalization method, one corresponding to a noiseless data set normalized by the "L^2" normalization method, and one corresponding to a noisy data set normalized by the "L^2" normalization method. The noise added to the training/testing set is generated based on the standard deviation curve in Figure 4.4 and the noise vector calculation in Equation 3.3.

A feature set is generated by peforming the feature selection method on the training/testing data. The entire training/testing set is input to Algorithm 1 in Section 2.2. The four FCBF feature sets determined are shown in Table 4.1.

Varying the noise of the data set and the normalization type generates a variety of SFS feature sets, similar to the process used to produce the FCBF feature sets in

Table 4.2. SFS Feature Sets

Classifier	Normalization Type	Data Type	Feature set wavelengths (nm)
MLP	Max	Noiseless	2425, 1645, 1350, 1215, 940
		Noisy	1325, 2230, 1965, 2075, 2000
	L^2	Noiseless	2220, 1605, 1705, 1785, 1470, 1660, 1550, 925, 2055, 2360
		Noisy	1205, 2000, 810, 2050, 830, 1440, 905
SVM	Max	Noiseless	1670, 1990, 1595, 1300, 1675, 1680, 1795, 1695, 1250
		Noisy	1305, 2015, 2125, 1965, 2425, 1650, 900, 1655, 1220
	L^2	Noiseless	2120, 2125, 2000, 2010
		Noisy	1310, 1270, 1340, 2190, 2365, 2165, 1135, 820, 855, 1100, 1650, 2240, 860, 1725

Table 4.1. However, because SFS features are in part determined by the classifier used, varying the classifier between SVM and MLP introduces another factor to be varied. Thus eight SFS feature sets are produced: four for MLPs and four for SVMs. The feature sets produced by peforming SFS with different classifiers, normalization types, and data noise settings are shown in Table 4.2. All feature sets in Table 4.2 are determined using Algorithm 2 in Section 3.5.

4.4 Classifier Optimization

Using the feature sets shown in Tables 4.1 and 4.2, MLP and SVM classifiers are used to classify data in the contact training/testing data set with noise added.

MLP performance depends on the number of hidden layers and the number of nodes in each hidden layer. Similarly, SVM performance depends on the kernel function used to transform the input data. Classifier parameters are optimized to produce a better Equal Weighted Accuracy (EWA) for the contact training/testing data set. Because it is desired to optimize the classifiers for a realistic scenario, all optimization is performed using the noisy contact training/testing data set.

The FCBF features in Table 4.1 are used with both the MLP and SVM classifiers. Because SFS relies on the classifier to produce a feature set, only the MLP features are utilized with the MLP classifier. Similarly, only the SVM features are utilized with the SVM classifier.

The effect of changing the hidden layer topology of the MLP classifier is explored by evaluating the performance of MLPs with different hidden layer topologies. Every hidden layer topology up to 3 hidden layers and up to 6 hidden nodes per layer is explored. Each of the 258 possible hidden layer topologies is trained and tested using 5-fold cross validation with the contact training/testing data set. The best testing set EWA from the 5 folds is recorded. The 5-fold cross validation process is repeated until 50 best testing set EWAs are recorded for each topology. Thus 258 structures * 50 repetitions * 5 folds = 64500 MLPs are created, but only 258 * 50 = 12900 of these produce the best testing EWA of their 5 fold grouping and are recorded with their winning EWA score. This methodology allows for determining the average performance of the best fold from a 5-fold cross validation procedure.

The topology that produces the highest mean EWA on the contact training/testing data set is determined. The set of topologies that produce a statistically identical result (according to a two-sided Wilcoxon Rank Sum Test (WRST) with a 95% confidence interval) with the highest mean topology is found. The topology from the statistically identical set with the smallest number of nodes is considered the best

Table 4.3. MLP Optimization Results

Normalization	Feature Set	Selected Topology	Average Training/Testing EWA
Max	FCBF (Noiseless)	[2 5 3 1]	85.2%
	FCBF (Noisy)	[4 6 1]	91.9%
	SFS (Noiseless)	[5 6 1 1]	90.8%
	SFS (Noisy)	[5 5 1]	91.7%
L^2	FCBF (Noiseless)	[1 2 1]	81.1%
	FCBF (Noisy)	[4 5 1]	89.4%
	SFS (Noiseless)	[10 6 1]	90.3%
	SFS (Noisy)	**[7 6 1]**	**93.5%**

network topology. Up to three runner-up topologies are also recorded, each one being the next-smallest in the group of statistically identical topologies. In instances where less than three runner-up topologies exist, as many that exist are recorded. The Receiver Operating Characteristic (ROC) curves of the runner-up topologies for the generalization data set and the image data set are provided in Appendix B.

A summary of the optimization results of the MLP classifier and corresponding EWA on the noisy training/testing contact data set is shown in Table 4.3. MLP topologies are denoted as vectors where the first element of the vector is the number of input nodes (features), the following elements are the numbers of hidden nodes in the consecutive hidden layers, and the last element is the number of output nodes. Thus a vector $[x, h_1, h_2, \cdots, h_N, o]$ represents a MLP with x inputs, h_1 nodes in the first hidden layer, h_2 nodes in the second hidden layer, h_N nodes in the N^{th} hidden layer, and o outputs. For all MLPs in this thesis, there is only one output, so only one output node is needed in the MLP.

SVM classifiers were optimized by finding the kernel function that produces the best results for the training/testing contact data set. The three kernel functions

Table 4.4. SVM Optimization Results

Normalization	Feature Set	Best Kernel Setting	Average Training/Testing EWA
Max	FCBF (Noiseless)	Gaussian	82.5%
	FCBF (Noisy)	Polynomial	90.3%
	SFS (Noiseless)	Gaussian	89.5%
	SFS (Noisy)	**Gaussian**	**93.1%**
L^2	FCBF (Noiseless)	Gaussian	79.8%
	FCBF (Noisy)	Gaussian	89.9%
	SFS (Noiseless)	Gaussian	89.9%
	SFS (Noisy)	**Gaussian**	**93.1%**

explored were Gaussian, polynomial, and linear. Each of the three kernel functions explored were trained and tested using 5-fold cross validation. In a fashion similar to the MLP optimization process, the best testing score from each set of 5 folds is recorded. Thus 3 kernel functions * 25 repetitions * 5 folds = 375 SVMs are produced, and only 3 * 25 = 75 SVMs are recorded with their winning EWA score. The kernel that produces the highest average EWA is considered the best. The ROC curves produced by the other kernels for the generalization data set and the image data set are provided in Appendix C.

Table 4.4 shows the optimization results for the SVM classifier, as well as the EWA of each classifier on the noisy training/testing data set.

The optimized classifiers from Table 4.3 and Table 4.4 are applied to the contact generalization set (with added noise), which was not used in the feature selection or optimization steps. The contact generalization set is composed entirely of samples from textile and non-textile swaths not represented in the contact training/testing set, and contains some textile material compositions not represented in the contact training/testing set. The EWAs of each of the MLP and SVM classifiers on the

Table 4.5. Optimized Classifier Performance (MLP)

Normalization	Classifier Description	Average Generalization EWA
Max	MLP FCBF (Noiseless)	80.0%
	MLP FCBF (Noisy)	**81.5%**
	MLP SFS (Noiseless)	76.7%
	MLP SFS (Noisy)	76.3%
L^2	MLP FCBF (Noiseless)	69.1%
	MLP FCBF (Noisy)	74.2%
	MLP SFS (Noiseless)	75.1%
	MLP SFS (Noisy)	79.8%

Table 4.6. Optimized Classifier Performance (SVM)

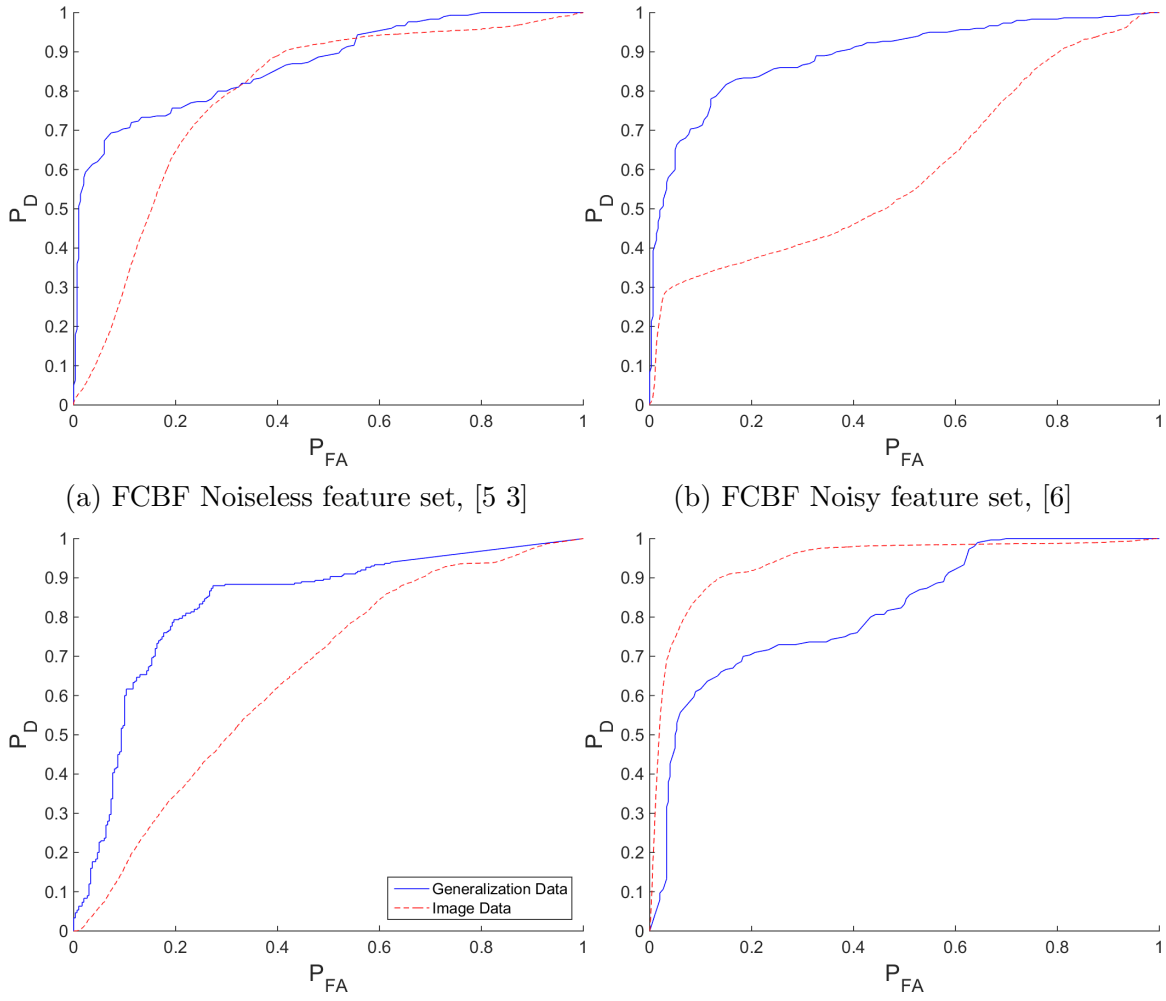
Normalization	Classifier Description	Average Generalization EWA
Max	SVM FCBF (Noiseless)	80.0%
	SVM FCBF (Noisy)	79.7%
	SVM SFS (Noiseless)	79.0%
	SVM SFS (Noisy)	77.1%
L^2	SVM FCBF (Noiseless)	68.4%
	SVM FCBF (Noisy)	74.7%
	SVM SFS (Noiseless)	76.0%
	SVM SFS (Noisy)	**80.5%**

training/testing set are presented in Table 4.5 and Table 4.6, respectively.

To show the performance of the optimized classifier settings for varying classification thresholds, a ROC curve must be produced for each of the 16 optimized classifiers. Each classifier optimization setting is associated with 50 (for MLPs) or 25 (for SVMs) different classifiers. Because a ROC curve is a function of a single threshold, one classifier must be chosen from each group to produce a ROC curve.

The classifier with the highest training/testing EWA out of its group is selected for ROC analysis. ROC curves are produced for both the generalization data set and the hyperspectral image data set.

In order to produce ROC curves, the "soft scores" of the classification results must be produced. In both the SVM and MLP classifiers, a classification decision is made depending on whether the soft score falls above or below a scalar threshold (see Section 3.4). A ROC curve can be produced by adjusting the threshold and recording the true positives and false positives at each threshold level using the soft score of each sample in the data set. The ROC curves for each of the 16 classifier groups are presented in Figures 4.7 - 4.10.



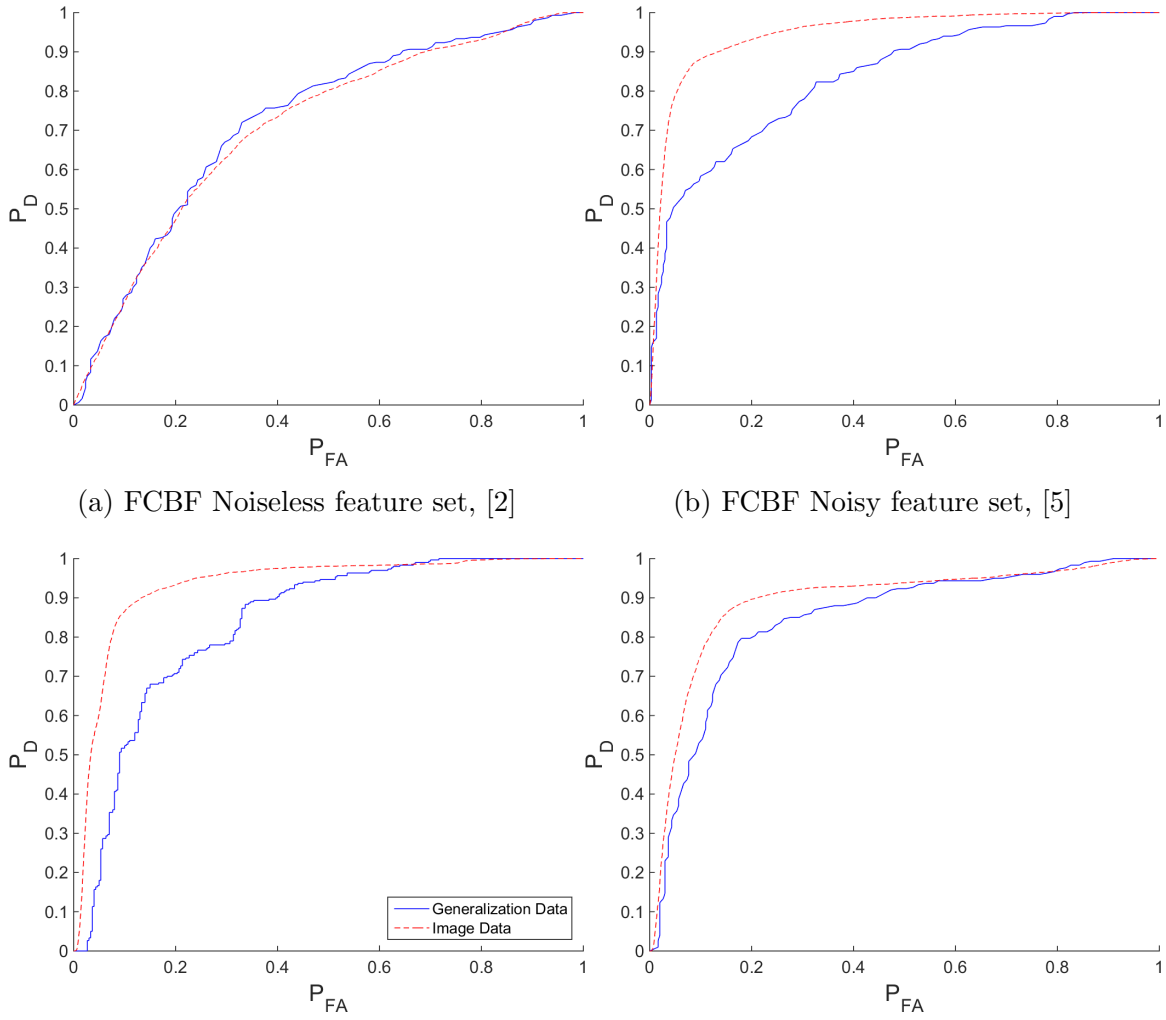
(a) FCBF Noiseless feature set, [5 3]

(b) FCBF Noisy feature set, [6]

(c) SFS Noiseless feature set, [6 1]

(d) SFS Noisy feature set, [5]

Figure 4.7. ROC curves of MLPs on contact generalization data and image data with Max Normalization. The ROC curves of the contact generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



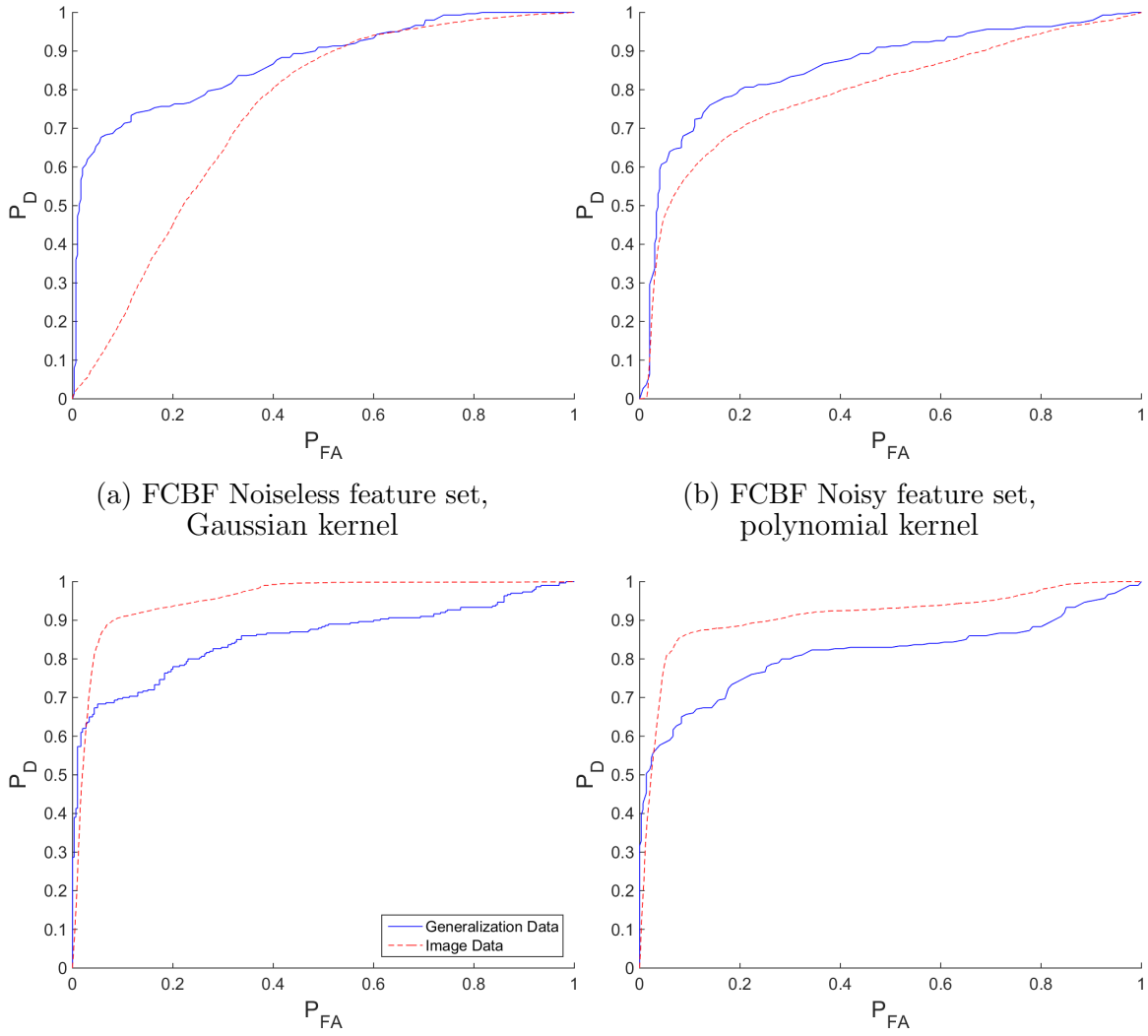
(a) FCBF Noiseless feature set, [2]

(b) FCBF Noisy feature set, [5]

(c) SFS Noiseless feature set, [6]

(d) SFS Noisy feature set, [6]

Figure 4.8. ROC curves of MLPs on contact generalization data and image data with L^2 Normalization. The ROC curves of the contact generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



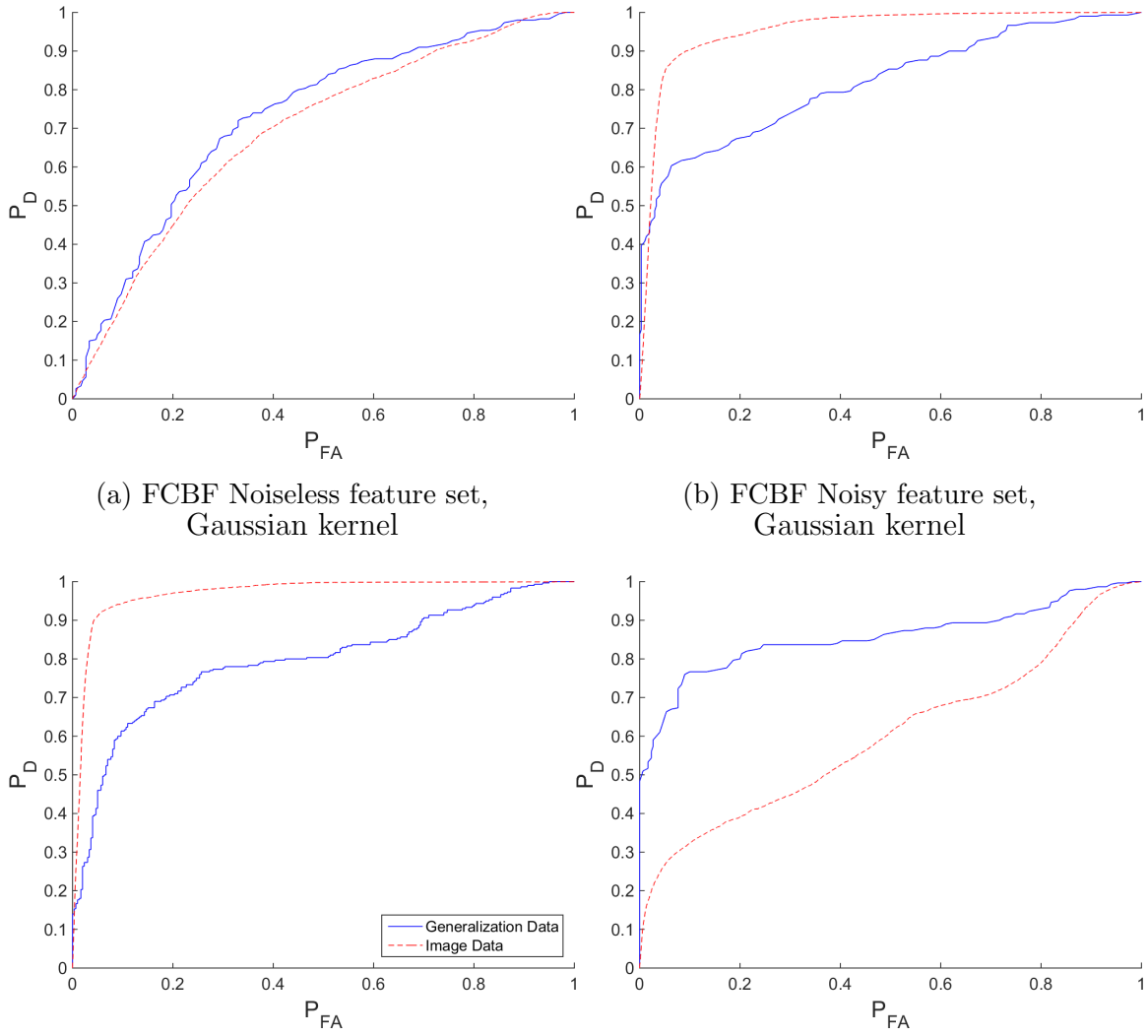
(a) FCBF Noiseless feature set, Gaussian kernel

(b) FCBF Noisy feature set, polynomial kernel

(c) SFS Noiseless feature set, Gaussian kernel

(d) SFS Noisy feature set, Gaussian kernel

Figure 4.9. ROC curves of SVMs on contact generalization data and image data with Max Normalization. The ROC curves of the contact generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) FCBF Noiseless feature set, Gaussian kernel

(b) FCBF Noisy feature set, Gaussian kernel

(c) SFS Noiseless feature set, Gaussian kernel

(d) SFS Noisy feature set, Gaussian kernel

Figure 4.10. ROC curves of SVMs on contact generalization data and image data with L^2 Normalization. The ROC curves of the contact generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).

The ROC curves in Figure 4.7 indicate that MLP classification performance on the max-normalized data varies significantly for both the generalization data set and the image data set. There is noticeably poor performance on the image data in Figure 4.7b, where P_D and P_{FA} are approximately equal for $P_D > 0.5$. The best image data performance among the max-normalized MLPs is the SFS noisy feature set in Figure 4.7d, where the P_D reaches 0.9 with $P_{FA} < 0.2$. Figure 4.8 presents ROC curves that are more consistent that those of Figure 4.7. The simple [1 2 1] MLP in Figure 4.7a has very similar ROC curves for both data sets. The image curves in Figure 4.8b-d are similar in shape, though high detection rates are achieved most quickly by the FCBF Noisy feature set of 1195nm, 2000nm, 1790nm, and 1650nm.

The SVM detectors depicted in Figures 4.9 and 4.10 exhibit different detection characteristics than the MLPs. SVMs tended to perform better than MLPs at very low levels of P_{FA}. Figures 4.9a,c,e and Figures 4.10b,d all show $P_D \geq 0.4$ for $P_{FA} < 0.03$ on the generalization data set. The most dramatic example of high performance at low P_{FA} on the generalization data set is Figure 4.10d, where $P_D = 0.5$ is achieved with a $P_{FA} = 0$. The performance of the SVM in Figure 4.10a is almost identical to that of Figure 4.9a, a consequence of them sharing the FCBF noiseless feature set, which contains only one feature (1185nm).

The overall performance of the classifiers is more easily compared with the Area Under the Curve (AUC) metric. AUC is a computation of the area under the ROC curve, and is inclusively bounded from 0 to 1 where a higher value indicates better classifier performance. The AUCs of each of the MLP and SVM classifiers are compared in Tables 4.7 and 4.8, respectively.

The topologies and operating parameters of the classifiers with the highest generalization data set and image data set AUCs (bolded with their winning AUC scores) are presented in Appendix D. To more intuitively illustrate the performance of the

Table 4.7. AUC of Optimized MLPs

Normalization	Classifier Description	Generalization AUC	Image AUC
Max	MLP FCBF (Noiseless)	0.869	0.789
	MLP FCBF (Noisy)	**0.892**	0.598
	MLP SFS (Noiseless)	0.832	0.652
	MLP SFS (Noisy)	0.817	0.940
L^2	MLP FCBF (Noiseless)	0.723	0.711
	MLP FCBF (Noisy)	0.835	**0.947**
	MLP SFS (Noiseless)	0.834	0.931
	MLP SFS (Noisy)	0.845	0.889

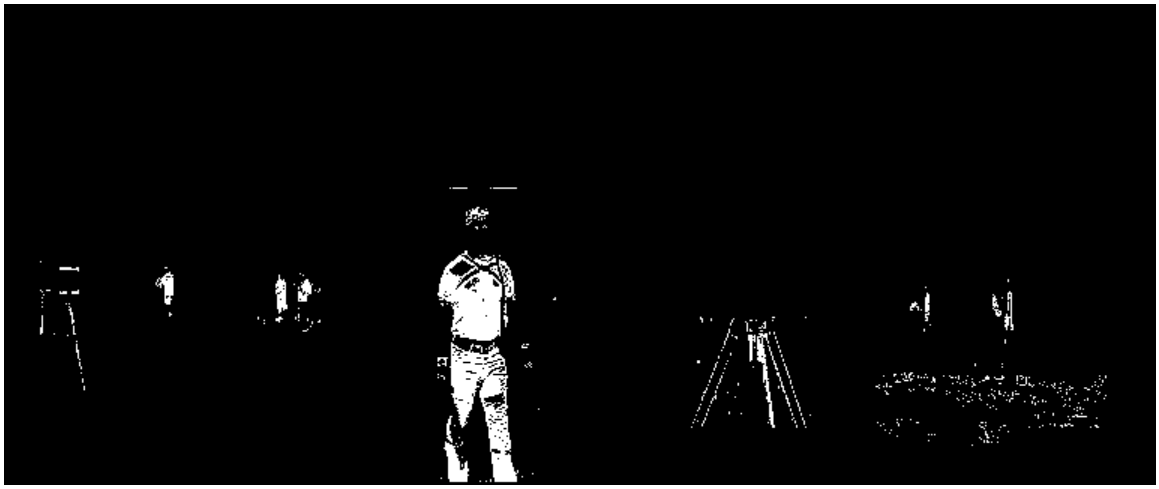
Table 4.8. AUC of Optimized SVMs

Normalization	Classifier Description	Generalization AUC	Image AUC
Max	**SVM FCBF (Noiseless)**	**0.872**	0.740
	SVM FCBF (Noisy)	0.858	0.796
	SVM SFS (Noiseless)	0.855	0.955
	SVM SFS (Noisy)	0.814	0.912
L^2	SVM FCBF (Noiseless)	0.728	0.692
	SVM FCBF (Noisy)	0.819	0.954
	SVM SFS (Noiseless)	0.797	**0.970**
	SVM SFS (Noisy)	0.856	0.602

best optimized image classifiers, the results of the MLP and SVM with the highest image AUCs are presented in detection masks. Results are thresholded to $P_D = 0.8$. Figure 4.11 presents the best MLP and SVM detection masks.

Figure 4.11 shows that the SVM outperformed the MLP due to its P_{FA} being approximately 55% that of the MLP. The MLP detector has many false alarms in the grass patch to the right of the metal tripod that the SVM does not. However, the images reveal that the detectors have common characteristics. Both pick up false alarms on the materials of the metal tripod in the foreground, as well as the smaller tripod holding the white Spectralon reflectance panel on the left of the image. In addition, both classifiers produce false alarms on the cones surrounding the pair of dismounts in the background. Also of note are the false alarms produced by the hair of the dismount in the foreground, including the eyebrows. This is attributable to the chemical similarity of human hair to the wool textiles in the training/testing set.

The MLP and SVM have misses in common as well. Both miss textiles where the textile surface is facing upward toward the sky, such as on the shoulders of the dismount in the foreground. Similarly, both have misses in areas of shadow. This indicates that both detectors are unable to identify textiles when they are exposed to more electromagnetic energy or less electromagnetic energy than normal.



(a) MLP detection mask, $P_{FA} = 0.0540$

(b) SVM detection mask, $P_{FA} = 0.0299$.

Figure 4.11. Detection masks of the hyperspectral image for MLP and SVM. Black pixels indicate non-textiles, while white pixels indicate textiles. Results are thresholded such that $P_D = 0.8$ for both images.

V. Conclusions and Future Work

Dismount detection has a wide variety of applications in security and search and rescue. Current dismount detection methods include the use of Synthetic Aperture Radar (SAR) [43] and spectral skin detection [62]. Spectral textile detection has advantages over these and other methods due to the abundance of textiles exposed on dismounts. However, there has been minimal investigation of the performance of spectral textile detection in a remote sensing environment.

To implement spectral textile detection, it is necessary to identify spectral features of textiles that allow textile materials to be uniquely identified among background spectra. Hyperspectral imagers collect electromagnetic radiation in hundreds of wavebands in the Visible/Near-Infrared (VNIR) and Short-Wave Infrared (SWIR) ranges. By applying feature selection methods to hyperspectral data, wavebands relevant to detecting textiles can be identified. Spectral detectors such as Spectral Matched Filter (SMF) cannot be used to detect textiles due to the variety of spectra textiles produce. More complex classifiers such as Support Vector Machines (SVMs) and Multi-Layer Perceptrons (MLPs), which are trained on labeled textile spectral data, can spectrally detect textiles.

5.1 Summary of Methodology and Results

This thesis presented a methodology of developing spectral textile detectors. A set of contact hyperspectral data containing textile and non-textile materials is collected. After dividing the data into a training data set and a generalization data set, feature selection is performed on the former to find features relevant to the textile detection problem. Different sets of features are created by varying noise settings, normalization types, and feature selection methods. MLP classifiers with differing topologies and

SVM classifiers with differing kernels were trained on the training data to determine the optimal MLP topology and SVM kernel. Classifiers with optimal settings were tested against the generalization set and a true remotely-sensed hyperspectral image.

The Area Under the Curve (AUC) metric is used to decide the best optimized classifiers. The best MLP and SVM results for the generalization set data were AUCs of 0.892 and 0.872, respectively. The best MLP and SVM results for the image data were AUCs of 0.947 and 0.970, respectively. The classifiers that produced these AUCs used only 2-4 features, and outperformed classifiers that made use of larger feature sets. This indicates that 2-4 features is sufficient to detect textiles in hyperspectral data.

The generally superior performance of the classifiers on the image data is best attributed to the smaller variety of textiles and nontextiles present in the image data set. The comparatively low performance in the simulated data set, which is composed of a larger variety of both textile and non-textile samples, shows that the generalization ability of the classifiers is not sufficient to identify textile compositions they have not been trained on. However, the higher AUCs on the image data and the detection masks in Figure 4.11 show that spectral textile detectors are reliable on more common textile materials. The SVM trained with the wavebands 2000nm, 2010nm, 2120nm, and 2125nm works best for the purpose of detecting dismounts in image data (AUC = 0.970).

5.2 Future Work

Spectral textile detection is not a well-explored method of dismount detection. This section introduces recommended avenues of further research.

The limited amount of remotely-sensed data available for this thesis does not provide sufficient information to determine the effectiveness of a spectral textile detector

on a wide variety of remote sensing scenarios. It is desirable to know a textile detector's performance in scenes where dismounts are partially obscured or in the shade. The problems associated with detecting textiles soiled with dirt, dust, and foliage can be examined. The work of Chan [11] analyzes the effectiveness of skin detection algorithms in aquatic conditions. Similar research into wet or submerged textile detection is necessary if a textile detector is to be used in an aquatic environment.

This thesis uses processed reflectance data that must be calculated from radiance measurements by placing an object of known reflectance in the scene. In a realistic remote-sensing scenario, it is infeasible to have objects of known reflectance. Moreover, the processing time associated with calculating reflectance in a scene significantly slows the detection process. Beisley [5] produced a reliable way of using raw radiance data, rather than reflectance data, in spectral skin detectors. Beisley's method can be implemented for use with textile detection.

Yeom [85] proposes a method of using spectral features in the VNIR and SWIR domains to detect certain FOI associated with a known dismount of interest (DOI) among other textile samples. The textile detectors used in this research can aid in the detection of fabrics of interest (FOI). By combining a universal textile detector from this thesis with an FOI detector, it is possible to detect an FOI (and thus a DOI) in a remote sensing scenario.

The shortcomings of the detectors in this thesis reveal ways to make spectral textile detection more reliable. Many false alarms are produced by foliage in a scene. Further research is needed to determine a method of mitigating false alarms due to trees, bushes, and grass in a scene.

It may also be possible to design a textile detector for only one textile material e.g. polyester with better performance characteristics than the detectors in this thesis. A polyester detector would lack the generalization of a textile detector, but could

provide better detection ability in situations where all dismounts are wearing textiles containing polyester.

At the time of writing, no method of integrating a variety of spectral detectors for dismount detection has been produced or explored. Skin detection efforts have largely dominated spectral dismount detection work. A combination of skin, hair, and textile spectral detectors would produce a more robust dismount detector than a detector that only searches for one type of human signature. A multi-signature dismount detector can search a scene for the presence of multiple human spectral signatures, using aggregate knowledge from many detectors to recognize the presence of a dismount.

List of Acronyms

Acronym	Definition
ASD	Analytical Spectral Devices
AUC	Area Under the Curve
EWA	Equal Weighted Accuracy
FCBF	Fast Correlation-Based Filter
FOV	Field of View
HSI	Hyperspectral Imager
IG	Information Gain
ILF	Induced Local Field
LM	Levenberg-Marquardt
MLP	Multi-Layer Perceptron
MRMR	Minimal-Redundancy-Maximal-Relevance
ROC	Receiver Operating Characteristic
SAR	Synthetic Aperture Radar
SBS	Sequential Backward Selection
SFBS	Sequential Floating Backward Selection
SFFS	Sequential Floating Forward Selection
SFS	Sequential Forward Selection
SID	Spectral Information Divergence
SMF	Spectral Matched Filter

SU	Symmetrical Uncertainty
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine - Recursive Feature Elimination
SWIR	Short-Wave Infrared
VNIR	Visible/Near-Infrared
WRST	Wilcoxon Rank Sum Test

Appendix A. List of Materials in Training/Testing and Generalization Sets

Table A.1. Training/Testing Set

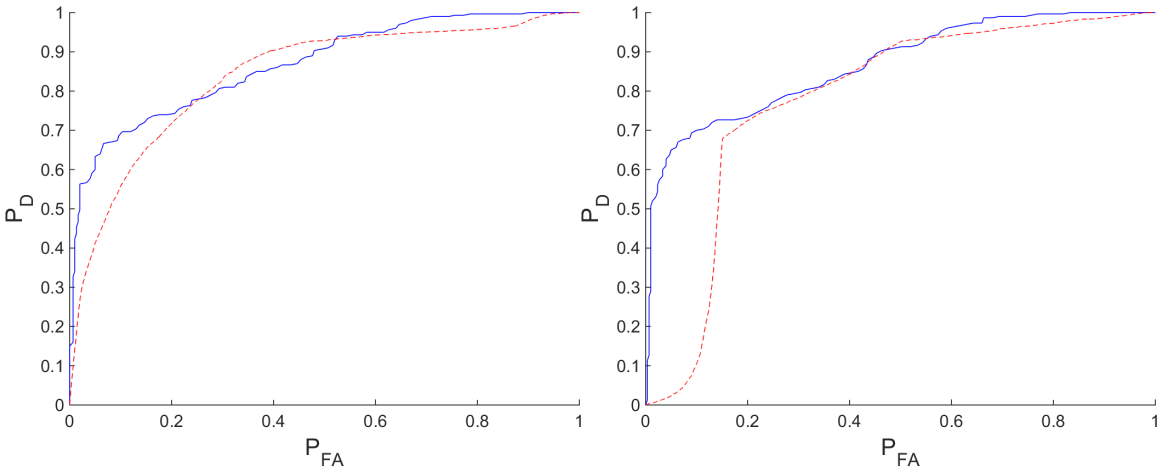
Textiles	Non-textiles
50% acrylic 40% polyester 10% rayon	asphalt (x5)
50% cotton 50% linen	brick (x2)
50% cotton 50% polyester (x2)	car surface (x2)
53% linen 47% rayon	metal (x9)
55% linen 45% rayon	grass (x8)
55% polyester 45% rayon	wood (x6)
55% polyester 45% wool (x2)	plastic (x8)
60% cotton 37% polyester 3% spandex	concrete (x2)
60% cotton 40% polyester	paper towel
60% wool 32% polyester 8% rayon	rock (x2)
65% polyester 33% rayon 2% spandex	tree bark (x4)
65% polyester 35% cotton	
67% polyester 30% rayon 3% spandex	
70% cotton 28% polyester 2% spandex (x2)	
70% rayon 28% polyester 2% spandex	
80% polyester 20% cotton	
85% polyester 15% cotton	
90% cotton 10% polyester	
95% polyester 5% rayon	
95% polyester 5% spandex	
96% polyester 4% spandex (x2)	
97% cotton 3% spandex (x2)	
98% cotton 2% spandex	
98% polyester 2% spandex	
99% viscose 1% other	
100% cotton (x9)	
100% polyester (x9)	
100% nylon (x2)	

Table A.2. Generalization Set

Textiles	Non-textiles
54% linen 46% rayon	asphalt (x2)
58% cotton 39% polyester 3% spandex	grass (x6)
58% linen 42% cotton	wood (x2)
58% polyester 42% rayon	metal (x8)
60% cotton 40% polyester	tree bark
65% polyester 35% rayon	plastic (x4)
70% cotton 28% polyester 2% spandex	concrete (x2)
70% polyester 20% acrylic 5% wood 5%misc	leaf
76% rayon 21% polyester 3% spandex	brick(x2)
80% polyester 20% wool	tire (x2)
84% polyester 14% rayon 2% spandex	
91% rayon 9% spandex	
95% acrylic 5% spandex	
95% rayon 5% spandex	
96% rayon 4% spandex	
97% cotton 3% spandex	
100% acrylic	
100% wool (x3)	
95% acrylic 5% spandex	
100% cotton (x4)	
100% nylon (x2)	
100% polyester (x3)	

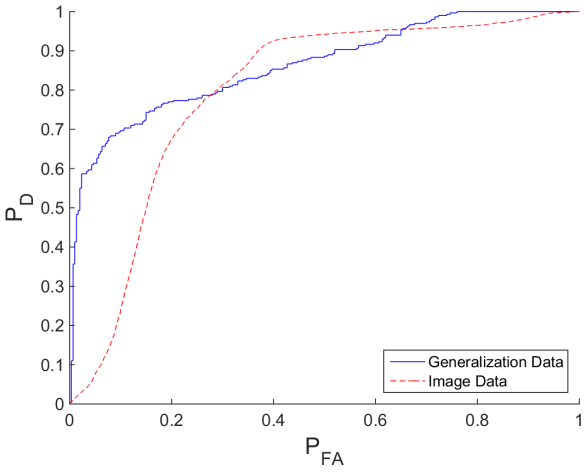
Appendix B. Additional Multi-Layer Perceptron (MLP) Receiver Operating Characteristic (ROC) curves

Figures B.1 through B.7 present Receiver Operating Characteristic (ROC) curves for the "runner-up" Multi-Layer Perceptron (MLP) topologies described in Section 4.4. The Area Under the Curves (AUCs) are included as AUC_{GEN} for the generalization data set and AUC_{IM} for the image data set. Figure B.7 shows results for only one runner-up MLP because only one runner-up topology existed for the L^2-normalized noisy feature set. Because there were no runner-up topologies for the L^2-normalized noiseless feature set, no plots for that feature set are presented.



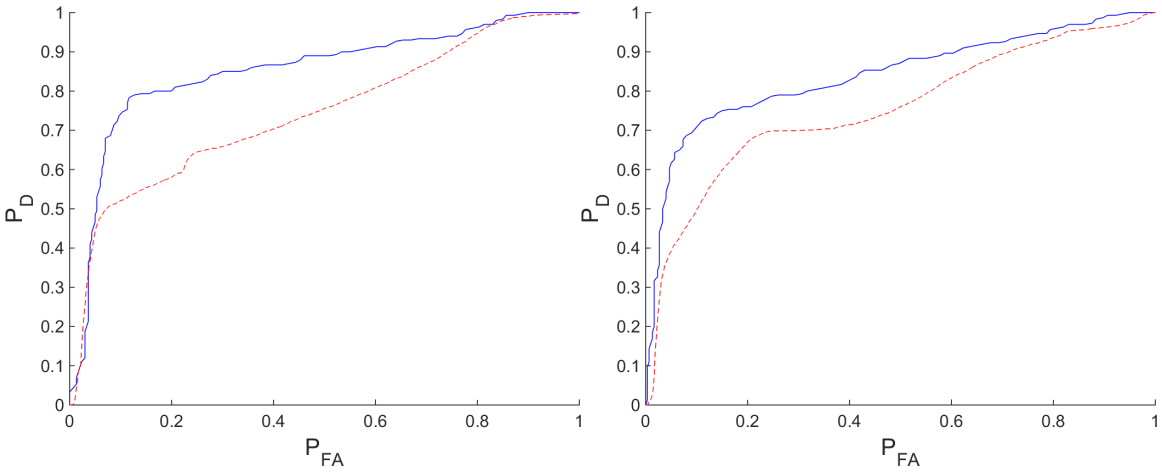
(a) [2 4 4 1], $\text{AUC}_{GEN} = 0.868$, $\text{AUC}_{IM} = 0.838$

(b) [2 4 5 1], $\text{AUC}_{GEN} = 0.868$, $\text{AUC}_{IM} = 0.781$



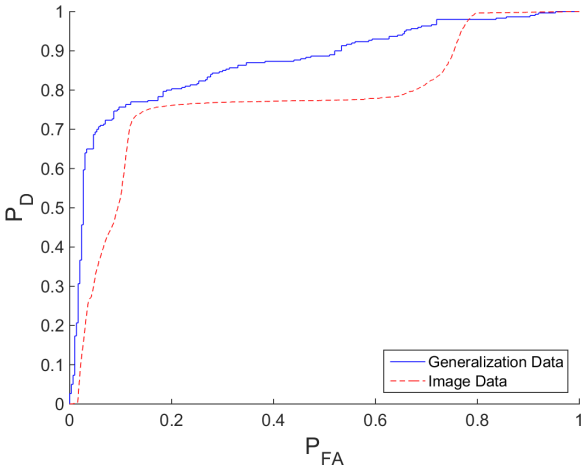
(b) [2 5 4 1], $\text{AUC}_{GEN} = 0.863$, $\text{AUC}_{IM} = 0.796$

Figure B.1. ROC curves of selected hidden layer networks for the noiseless max-normalized FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



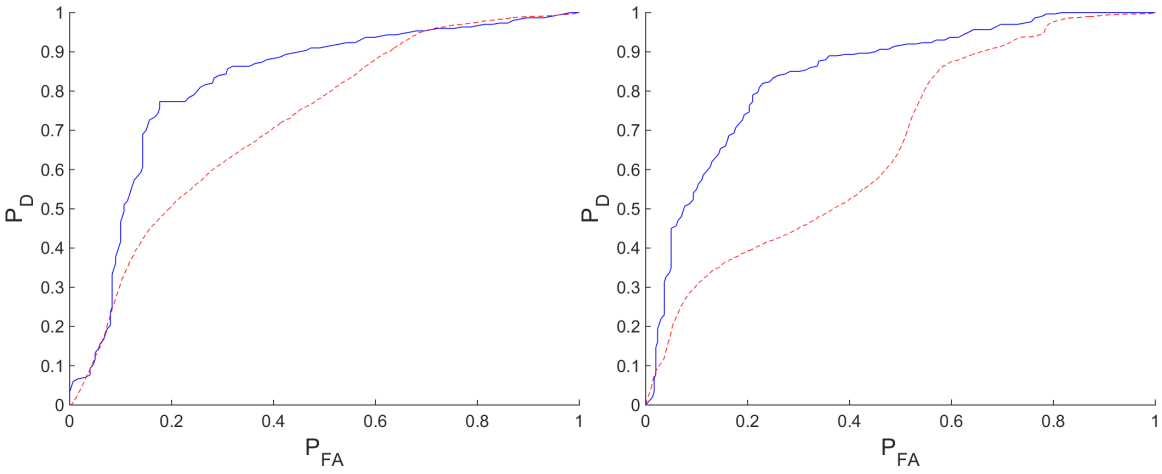
(a) [4 6 4 1], $\text{AUC}_{GEN} = 0.852$, $\text{AUC}_{IM} = 0.743$

(b) [4 6 3 1 1], $\text{AUC}_{GEN} = 0.841$, $\text{AUC}_{IM} = 0.756$



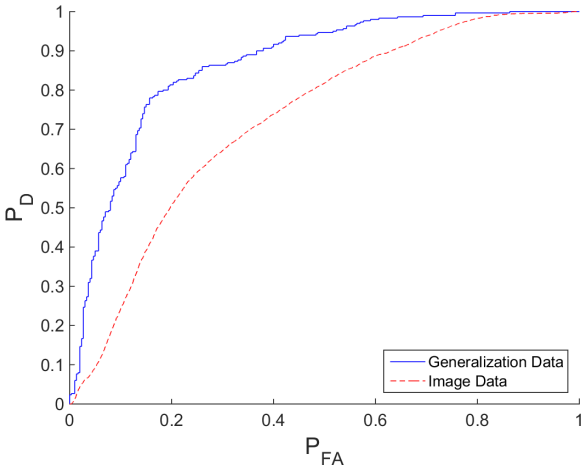
(c) [4 5 6 1], $\text{AUC}_{GEN} = 0.872$, $\text{AUC}_{IM} = 0.778$

Figure B.2. ROC curves of selected hidden layer networks for the noisy max-normalized FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



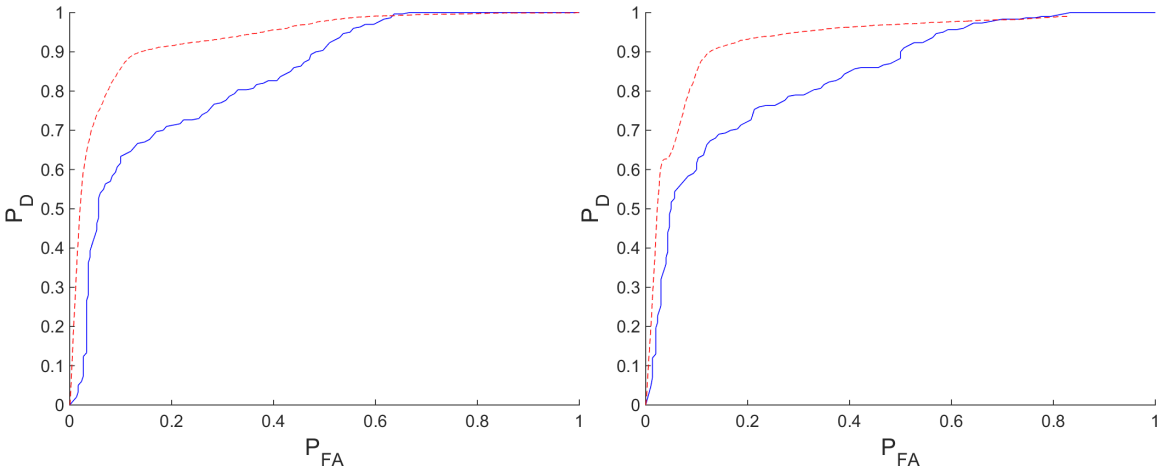
(a) [5 5 2 1], $\text{AUC}_{GEN} = 0.819$, $\text{AUC}_{IM} = 0.723$

(b) [5 5 4 1], $\text{AUC}_{GEN} = 0.844$, $\text{AUC}_{IM} = 0.663$



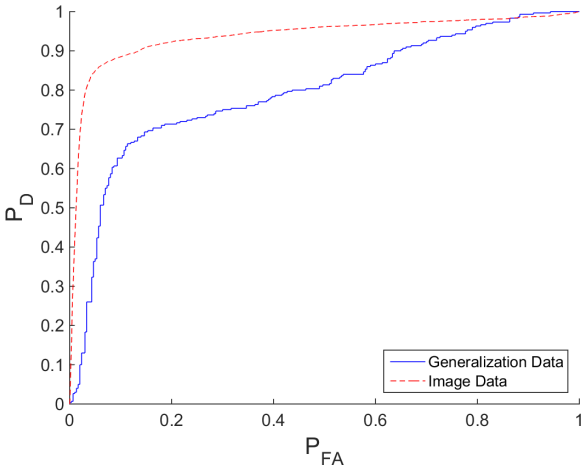
(c) [5 6 2 1 1], $\text{AUC}_{GEN} = 0.866$, $\text{AUC}_{IM} = 0.726$

Figure B.3. ROC curves of selected hidden layer networks for the noiseless max-normalized SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



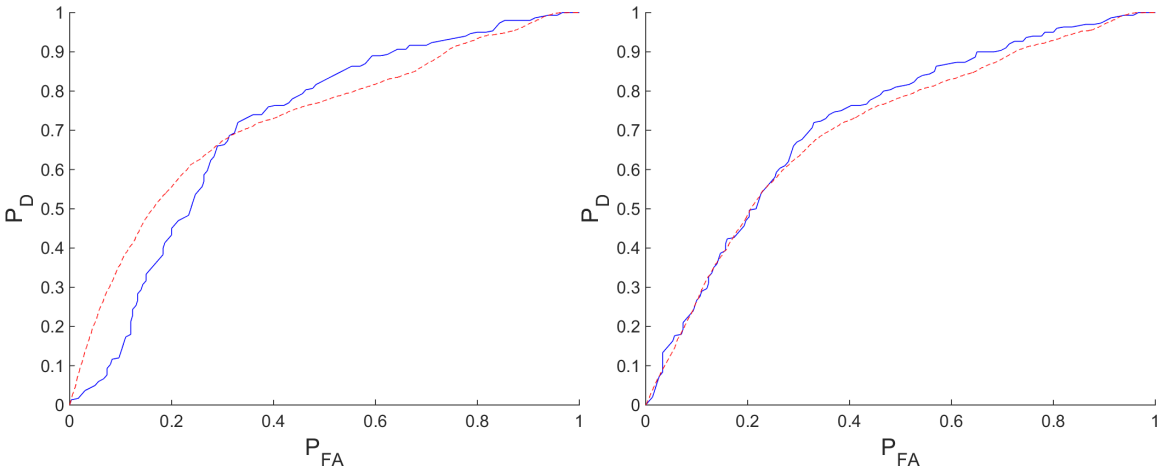
(a) [5 4 1 1], $\text{AUC}_{GEN} = 0.838$, $\text{AUC}_{IM} = 0.935$

(b) [5 5 1 1], $\text{AUC}_{GEN} = 0.841$, $\text{AUC}_{IM} = 0.929$



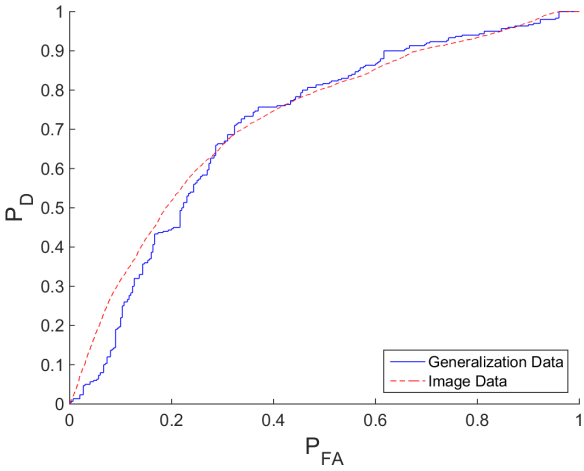
(c) [5 4 2 1], $\text{AUC}_{GEN} = 0.797$, $\text{AUC}_{IM} = 0.938$

Figure B.4. ROC curves of selected hidden layer networks for the noisy max-normalized SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



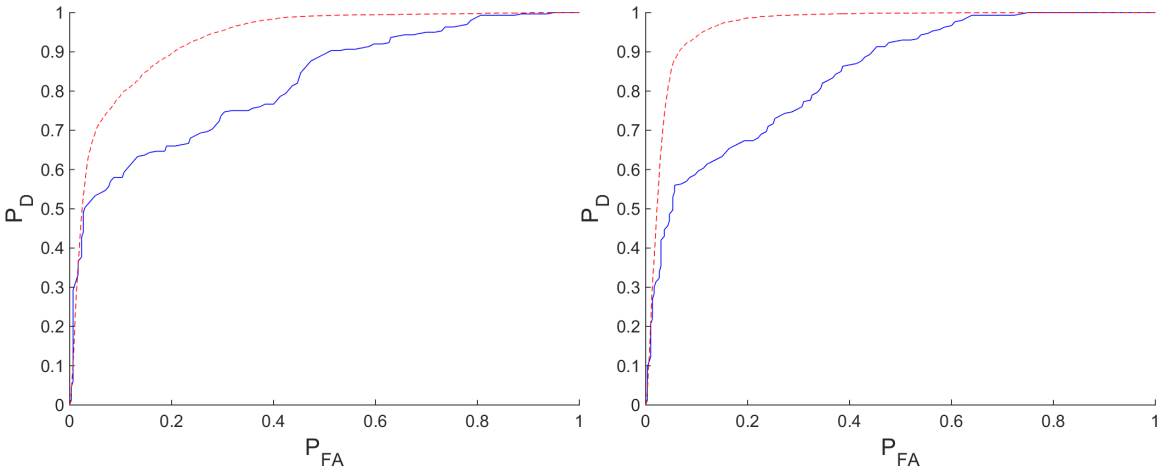
(a) [1 3 3 1], $\text{AUC}_{GEN} = 0.709$, $\text{AUC}_{IM} = 0.723$

(b) [1 3 4 1], $\text{AUC}_{GEN} = 0.723$, $\text{AUC}_{IM} = 0.704$



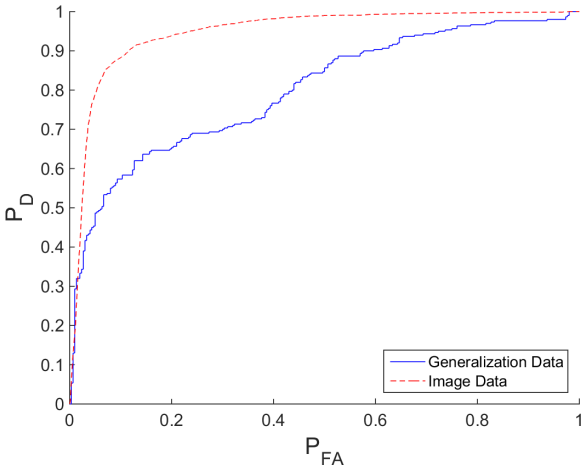
(c) [1 1 3 3 1], $\text{AUC}_{GEN} = 0.712$, $\text{AUC}_{IM} = 0.724$

Figure B.5. ROC curves of selected hidden layer networks for the noiseless L^2-normalized FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) [4 5 5 1], $\text{AUC}_{GEN} = 0.818$, $\text{AUC}_{IM} = 0.931$

(b) [4 4 61 1], $\text{AUC}_{GEN} = 0.844$, $\text{AUC}_{IM} = 0.966$



(c) [4 6 5 1], $\text{AUC}_{GEN} = 0.799$, $\text{AUC}_{IM} = 0.947$

Figure B.6. ROC curves of selected hidden layer networks for the noisy L^2-normalized FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).

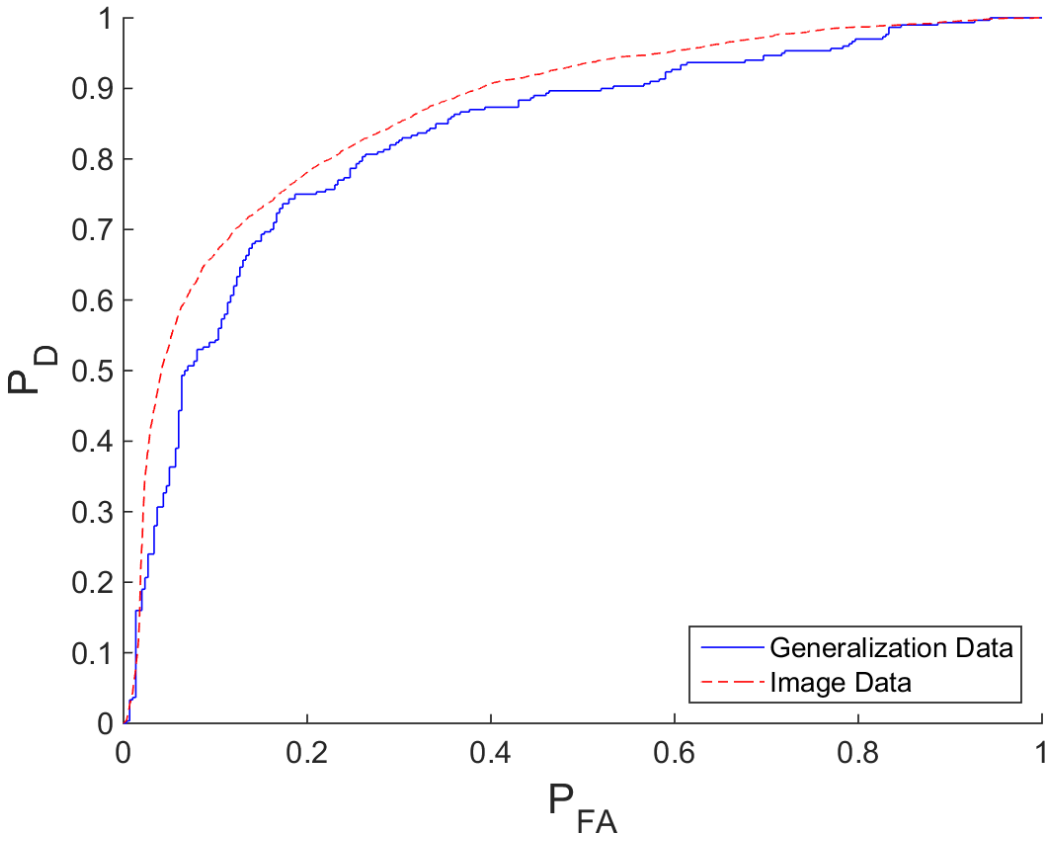
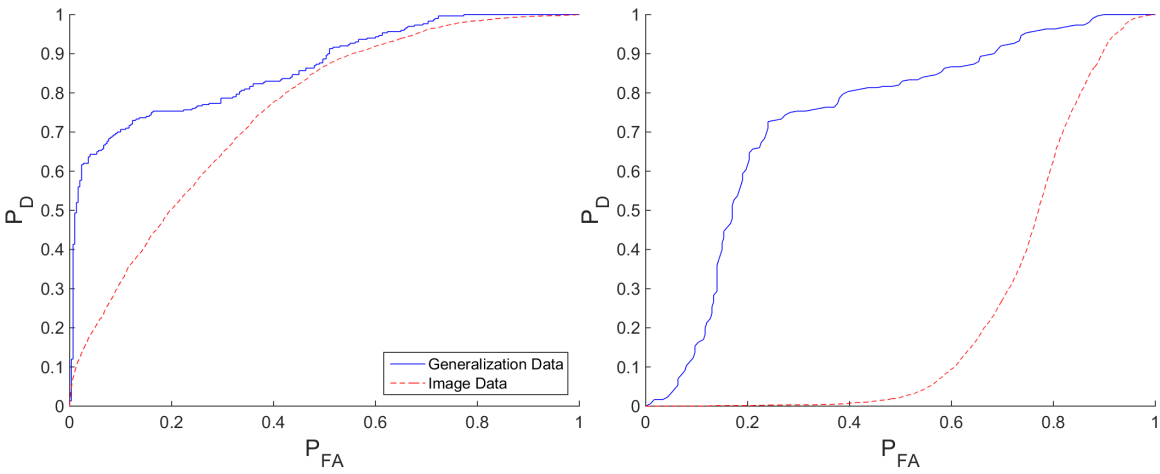


Figure B.7. ROC curves of [7 6 5 1] network for the noisy L^2-normalized SFS feature set. $AUC_{GEN} = 0.832$, $AUC_{IM} = 0.868$. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).

Appendix C. Additional Support Vector Machine (SVM) ROC curves

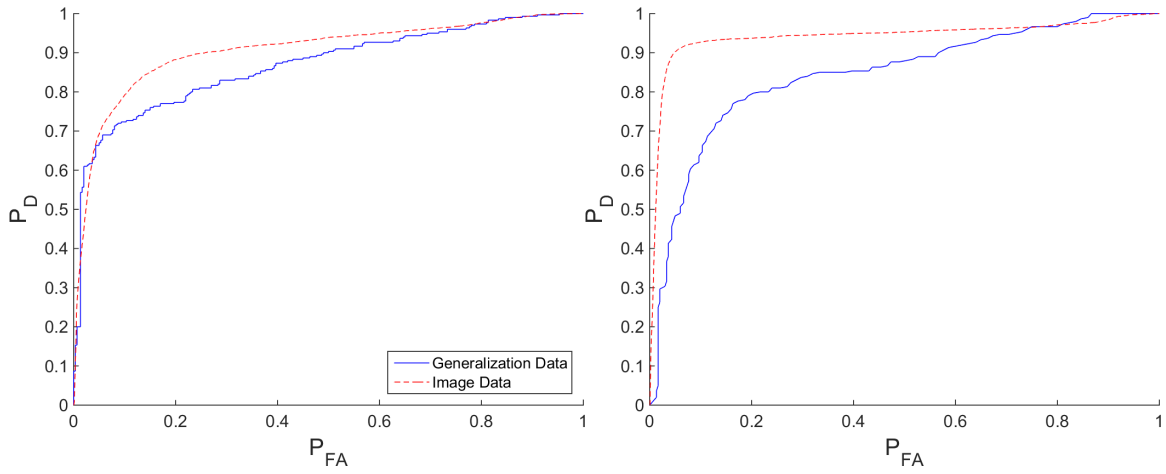
Figures C.1 through C.8 present Receiver Operating Characteristic (ROC) curves for the kernels not selected for optimization with each feature set. The Area Under the Curves (AUCs) are included as AUC_{GEN} for the generalization data set and AUC_{IM} for the image data set.



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.864$, $\text{AUC}_{IM} = 0.752$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.738$, $\text{AUC}_{IM} = 0.243$

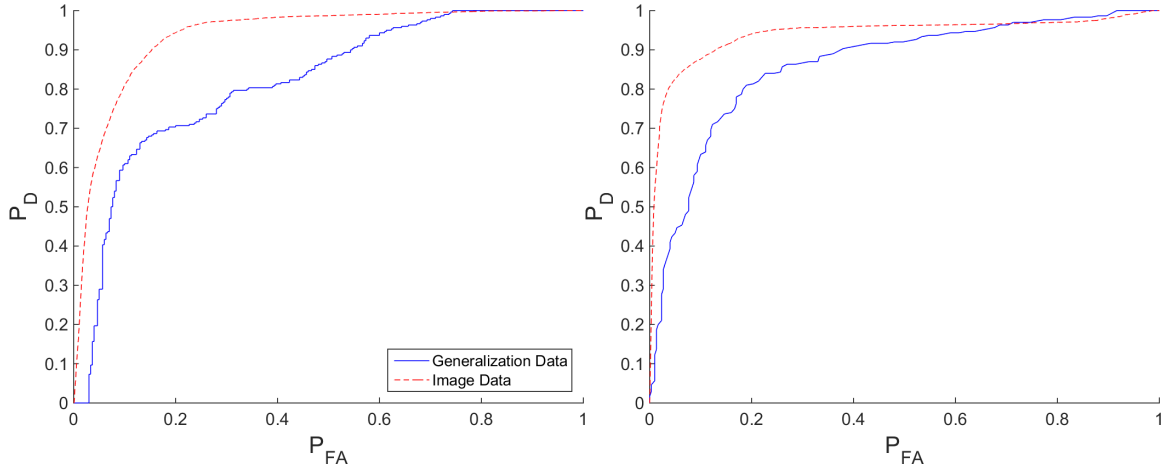
Figure C.1. ROC curves of SVM kernels not selected by optimization for the max-normalized noiseless SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Gaussian kernel
$\text{AUC}_{GEN} = 0.868$, $\text{AUC}_{IM} = 0.905$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.844$, $\text{AUC}_{IM} = 0.942$

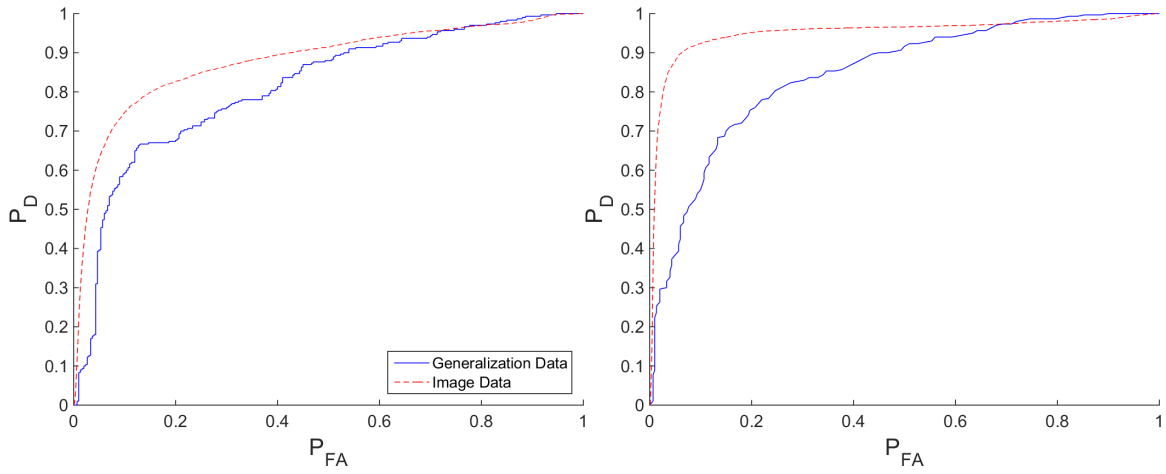
Figure C.2. ROC curves of SVM kernels not selected by optimization for the max-normalized noisy FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.817$, $\text{AUC}_{IM} = 0.936$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.861$, $\text{AUC}_{IM} = 0.940$

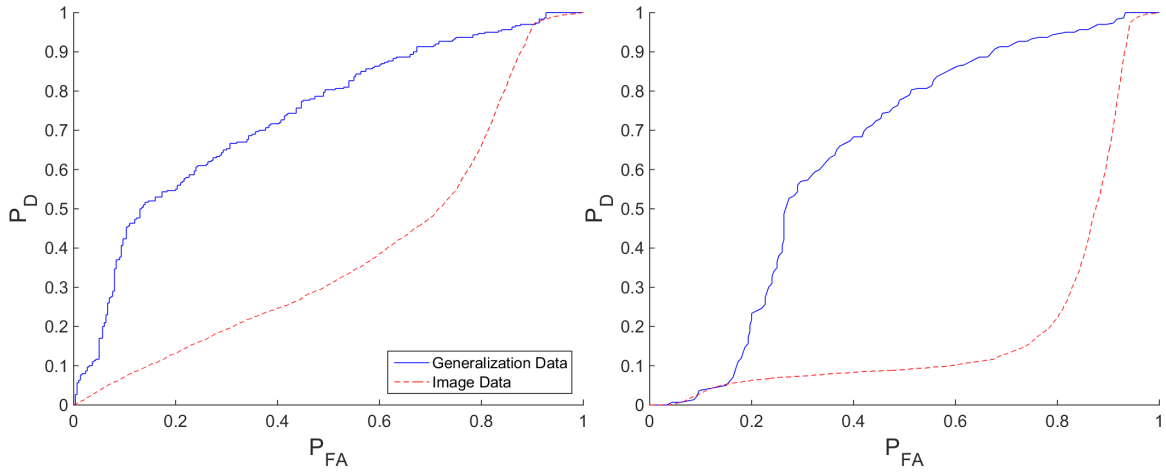
Figure C.3. ROC curves of SVM kernels not selected by optimization for the max-normalized noiseless SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.812$, $\text{AUC}_{IM} = 0.879$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.847$, $\text{AUC}_{IM} = 0.951$

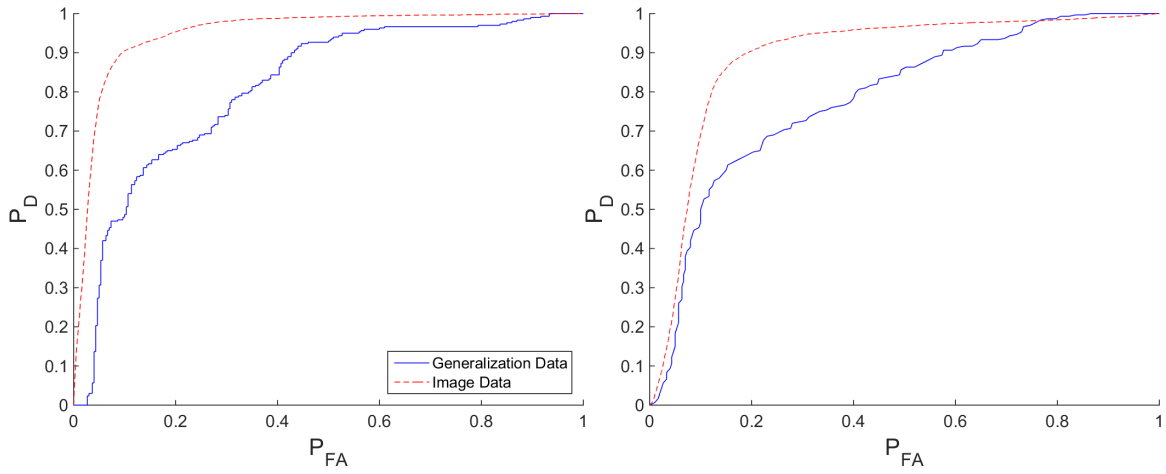
Figure C.4. ROC curves of SVM kernels not selected by optimization for the max-normalized noisy SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.736$, $\text{AUC}_{IM} = 0.393$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.642$, $\text{AUC}_{IM} = 0.198$

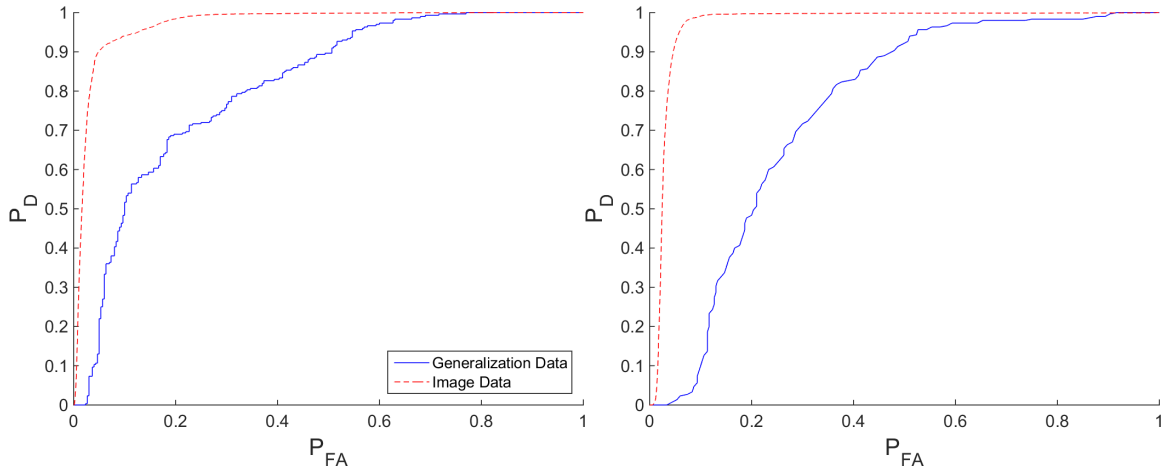
Figure C.5. ROC curves of SVM kernels not selected by optimization for the L^2-normalized noiseless FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.812$, $\text{AUC}_{IM} = 0.951$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.785$, $\text{AUC}_{IM} = 0.890$

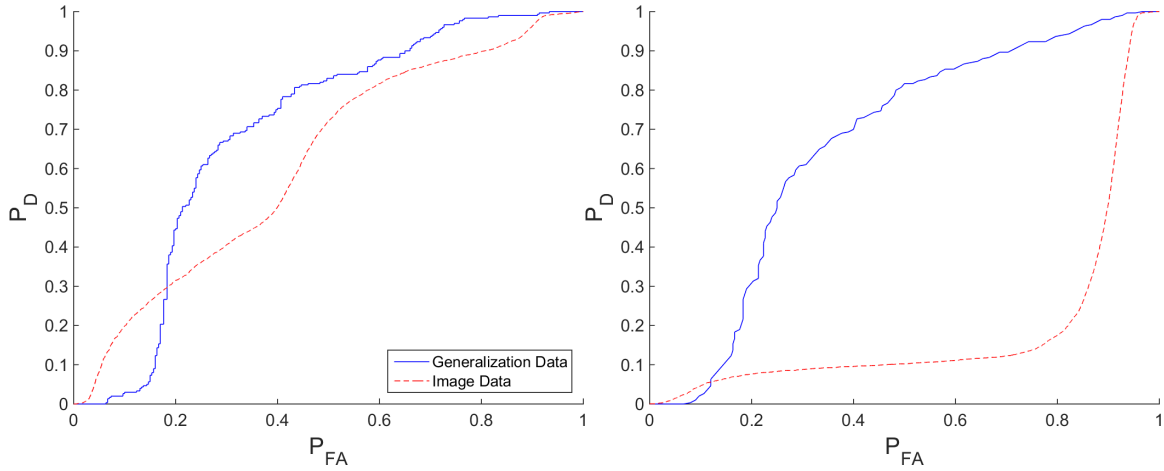
Figure C.6. ROC curves of SVM kernels not selected by optimization for the L^2-normalized noisy FCBF feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.813$, $\text{AUC}_{IM} = 0.971$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.753$, $\text{AUC}_{IM} = 0.970$

Figure C.7. ROC curves of SVM kernels not selected by optimization for the L^2-normalized noiseless SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).



(a) Polynomial kernel
$\text{AUC}_{GEN} = 0.694$, $\text{AUC}_{IM} = 0.618$

(b) Linear kernel
$\text{AUC}_{GEN} = 0.663$, $\text{AUC}_{IM} = 0.188$

Figure C.8. ROC curves of SVM kernels not selected by optimization for the L^2-normalized noisy SFS feature set. The ROC curves of the generalization data set are shown in blue (solid line), while the ROC curves of the image data set are shown in red (dashed line).

Appendix D. Structures, Weights, and Biases of Selected Classifiers

The structures, weights, and biases of the Multi-Layer Perceptrons (MLPs) and the training parameters of the Support Vector Machines (SVMs) with the highest Area Under the Curve (AUC) for the image data set and the generalization data set are enumerated in this appendix.

4.1 Highest-rated Image MLP

Inputs i_1, i_2, i_3, and i_4 are the L^2-normalized measurements at the wavebands 1195nm, 1650nm, 1790nm, and 2000nm, respectively. The structure of the MLP is shown in Figure D.1. Weights and biases are given in Tables D.1-D.3. The hyperbolic tangent function (Equation 3.6) was used as the activation function.

Table D.1. First hidden layer weights

		From Node			
		i_1	i_2	i_3	i_4
To Node	a_1	69.6585	-30.8848	11.2531	-42.1305
	a_2	-50.6595	43.4163	-51.7249	19.3712
	a_3	-36.1604	-7.3086	69.8189	-7.7391
	a_4	25.7578	25.0540	14.9634	-24.1422
	a_5	3.9878	-26.0244	-44.8837	-0.2193

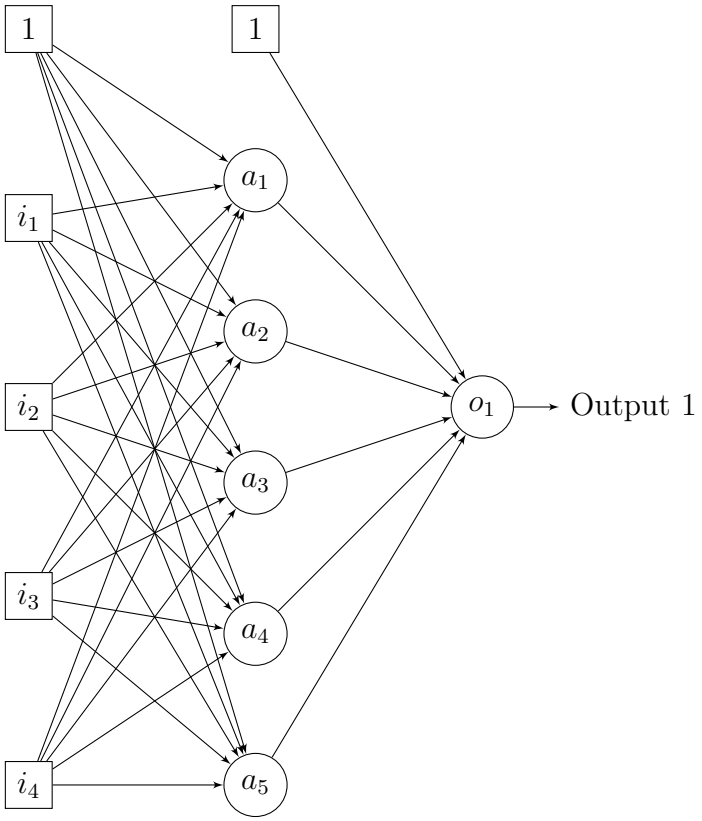


Figure D.1. Topology of the MLP with the highest AUC on the image data set.

Table D.2. Output layer weights

	From Node				
	a_1	a_2	a_3	a_4	a_5
To node o_1	-42.9296	-17.8731	-30.8848	-9.4568	-23.4303

Table D.3. Node Biases

Node	Bias Value
a_1	-2.0554
a_2	1.5152
a_3	-0.8788
a_4	5.9057
a_5	3.4241
o_1	-11.3893

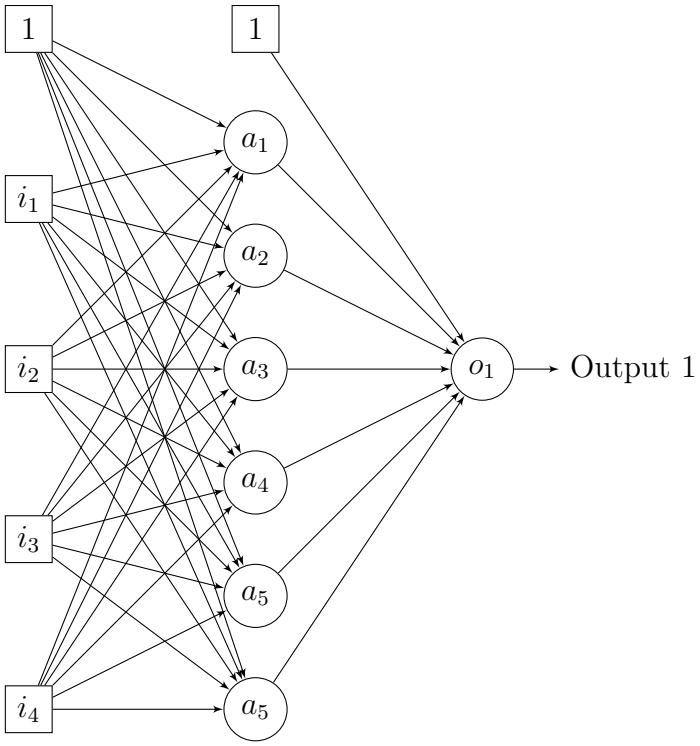


Figure D.2. Topology of the MLP with the highest AUC on the generalization data set.

4.2 Highest-rated Generalization MLP

Inputs i_1, i_2, i_3, and i_4 are the max-normalized measurements at the wavebands 815nm 1320nm 1965nm and 2160nm, respectively. Weights and biases are given in Tables D.4-D.6. The structure of the MLP is shown in Figure D.2. The hyperbolic tangent function (Equation 3.6) was used as the activation function.

Table D.4. First hidden layer weights

To Node \ From Node	i_1	i_2	i_3	i_4
a_1	3.1601	7.1008	-12.2838	0.9043
a_2	8.4708	-2.3693	0.8890	-0.7705
a_3	-2.0970	2.8557	-1.4650	2.4206
a_4	1.7818	1.1368	1.0767	7.2745
a_5	-13.3151	1.0922	8.5924	3.7210
a_6	3.7567	1.8495	-4.0021	3.4102

Table D.5. Output layer weights

	From Node a_1	a_2	a_3	a_4	a_5	a_6
To node o_1	-4.6552	16.2230	15.9406	-11.2606	7.3836	-8.6741

Table D.6. Node Biases

Node	Bias Value
a_1	-4.6129
a_2	-4.2407
a_3	-0.8277
a_4	-7.2986
a_5	3.2779
a_6	-1.4365
o_1	-9.6123

Table D.7. Settings for SVM with highest AUC on Image Data Set

Parameter	Value
autoscale	true
boxconstraint	1
kernelcachelimit	5000
kktviolationlevel	0
method	SMO
maxiter	400000
tolkkt	1e-3

4.3 Highest-rated Image SVM

The features of the SVM are 2000nm, 2010nm, 2120nm, and 2125nm. Normalization was performed using the L^2 method. The SVM used the Gaussian kernel. The other settings entered into the MATLAB® svmtrain function are enumerated in Table D.7.

Table D.8. Settings for SVM with highest AUC on Generalization Data Set

Parameter	Value
autoscale	true
boxconstraint	1
kernelcachelimit	5000
kktviolationlevel	0
method	SMO
maxiter	400000
tolkkt	1e-3

4.4 Highest-rated Generalization SVM

The features of the SVM are 1060nm and 2425nm. Normalization was performed using the max method. The SVM used the Gaussian kernel. The other settings entered into the MATLAB® svmtrain function are enumerated in Table D.8.

Bibliography

1. "Cellulose as a nanostructured polymer: A short review". *BioResources*, 3(4):1403 – 1418.

2. "Unit - Chemistry of Garments: Animal Fibres". `http://wwwchem.uwimona.edu.jm/courses/CHEM2402/Textiles/Animal_Fibres.html`. Accessed: 2014-03-12.

3. version 8.4.0 (R2014b), MATLAB. *svmtrain*. The MathWorks Inc., Natick, Massachusetts, 2014.

4. Alam, S., O. Odejide, O. Olabiyi, and A. Annamalai. "Further results on area under the ROC curve of energy detectors over generalized fading channels". *Sarnoff Symposium, 2011 34th IEEE*, 1–6. May 2011.

5. Beisley, Andrew P. *Spectral Detection of Human Skin in VIS-SWIR Hyperspectral Imagery Without Radiometric Calibration*. Master's thesis, 2012.

6. Bernard, I. "Multilayer perceptron and uppercase handwritten characters recognition". *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*, 935–938. Oct 1993.

7. Beveridge, J.R. and E.M. Riseman. "How easy is matching 2D line models using local search?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6):564–579, Jun 1997. ISSN 0162-8828.

8. Blum, Avrim L. and Pat Langley. "Selection of relevant features and examples in machine learning". *Artificial Intelligence*, 97:245–271, 1997.

9. Breve, F.A, M.P. Ponti-Junior, and N. Mascarenhas. "Multilayer Perceptron Classifier Combination for Identification of Materials on Noisy Soil Science Multispectral Images". *Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on*, 239–244. Oct 2007. ISSN 1530-1834.

10. Camps-Valls, G. and L. Bruzzone. "Kernel-based methods for hyperspectral image classification". *Geoscience and Remote Sensing, IEEE Transactions on*, 43(6):1351–1362, June 2005. ISSN 0196-2892.

11. Chan, Alice W. *An Assesment of Normalized Difference Skin Index Robustness in Aquatic Environments*. Master's thesis, 2014.

12. Chang, Chein-I. "Spectral information divergence for hyperspectral image analysis". *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, volume 1, 509–511 vol.1. 1999.

13. Chang, Chein-I and D.C. Heinz. "Constrained subpixel target detection for remotely sensed imagery". *Geoscience and Remote Sensing, IEEE Transactions on*, 38(3):1144–1159, May 2000. ISSN 0196-2892.

14. Chen, Jin, Xiuping Jia, Wei Yang, and Bunkei Matsushita. "Generalization of Subpixel Analysis for Hyperspectral Data With Flexibility in Spectral Similarity Measures". *Geoscience and Remote Sensing, IEEE Transactions on*, 47(7):2165–2171, July 2009. ISSN 0196-2892.

15. Civco, Daniel L. "Artificial neural networks for land-cover classification and mapping". *International journal of geographical information systems*, 7(2):173–186, 1993. URL `http://dx.doi.org/10.1080/02693799308901949`.

16. Clark, Jeffrey D. "Artificial Intelligence 823 lecture notes", 2013.

17. Cook, J. Gordon. *Handbook of Textile Fibers*. Merrow Publishing Co. Ltd, 2 edition, 1960.

18. Cooksey, Catherine and David Allen. "Investigation of the potential use of hyperspectral imaging for stand-off detection of person-borne IEDs", 2011. URL `http://dx.doi.org/10.1117/12.883502`.

19. Cooksey, Catherine C., Jorge E. Neira, and David W. Allen. "The evaluation of hyperspectral imaging for the detection of person-borne threat objects over the 400nm to 1700nm spectral region", 2012. URL `http://dx.doi.org/10.1117/12.919432`.

20. Dalton, P M, J H Nobbs, and R W Rennell. "The influence of moisture content on the colour appearance of dyed textile materials. Part 1-dyeing methods and reflectance measurements on wool". *Journal of the Society of Dyers and Colourists*, 111(9):285–287, 1995. ISSN 1478-4408. URL `http://dx.doi.org/10.1111/j.1478-4408.1995.tb01743.x`.

21. Dash, M. and H. Liu. "Feature Selection for Classification". *Intelligent Data Analysis*, 1:131–156, 1997.

22. Ding, C. and H. Peng. "Minimum redundancy feature selection from microarray gene expression data". *Journal of Bioinformatics and Computational Biology*, 3(2):185–205, June 2004. ISSN 0162-8828.

23. Dreyer, P. (Jutland Telephone. "Classification of land cover using optimized neural nets on SPOT data". May 1993.

24. Duan, K. and J.C. Rajapakse. "A variant of SVM-RFE for gene selection in cancer classification with expression data". *Computational Intelligence in Bioinformatics*

and Computational Biology, 2004. CIBCB '04. Proceedings of the 2004 IEEE Symposium on, 49–55. Oct 2004.

25. Dubas, Stephan T., Panittamat Kumlangdudsana, and Pranut Potiyaraj. "Layer-by-layer deposition of antimicrobial silver nanoparticles on textile fibers". *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 289(13):105 – 109, 2006. ISSN 0927-7757. URL `http://www.sciencedirect.com/science/article/pii/S0927775706002858`.

26. Duch, Wodzisaw, Jacek Biesiada, Tomasz Winiarski, Karol Grudziski, and Krzysztof Grbczewski. "Feature Selection Based on Information Theory Filters". Leszek Rutkowski and Janusz Kacprzyk (editors), *Neural Networks and Soft Computing*, volume 19 of *Advances in Soft Computing*, 173–178. Physica-Verlag HD, 2003. ISBN 978-3-7908-0005-0. URL `http://dx.doi.org/10.1007/978-3-7908-1902-1_23`.

27. Friedman, Nir, Dan Geiger, Moises Goldszmidt, G. Provan, P. Langley, and P. Smyth. "Bayesian Network Classifiers". *Machine Learning*, 131–163. 1997.

28. Frohlich, H., O. Chapelle, and Bernhard Scholkopf. "Feature selection for support vector machines by means of genetic algorithm". *Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on*, 142–148. Nov 2003. ISSN 1082-3409.

29. Ghosh, Sbuhas and James Rodgers. "NIR Analysis of Textiles". Donald A. Burns and Emil W. Ciurczak (editors), *Handbook of Near-Infrared Analysis, 3rd Edition*. CRC Press, Boca Raton, 2008.

30. Gibbons, J.D. and S. Chakraborti. *Nonparametric Statistical Inference, Fourth*

Edition: Revised and Expanded. Taylor & Francis, 2014. ISBN 9780203911563. URL http://books.google.com/books?id=kJbVO2G6VicC.

31. Goldstein, J. Scott, M.L. Picciolo, M. Rangaswamy, and J.D. Griesbach. "Detection of dismounts using synthetic aperture radar". *Radar Conference, 2010 IEEE*, 209–214. May 2010. ISSN 1097-5659.

32. Gonzalez, J.A., Michael J. Mendenhall, and E. Merenyi. "Minimum Surface Bhattacharyya feature selection". *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop on*, 1–4. Aug 2009.

33. Grant, L.L. and G.K. Venayagamoorthy. "Cellular Multilayer Perceptron for Prediction of Voltages in a Power System". *Intelligent System Applications to Power Systems, 2009. ISAP '09. 15th International Conference on*, 1–6. Nov 2009.

34. Gualtieri, J Anthony and Robert F Cromp. "Support vector machines for hyperspectral remote sensing classification". *The 27th AIPR Workshop: Advances in Computer-Assisted Recognition*, 221–232. International Society for Optics and Photonics, 1999.

35. Guha, D.R. and S.K. Patra. "Cochannel Interference Minimization Using Wilcoxon Multilayer Perceptron Neural Network". *Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on*, 145–149. March 2010.

36. Guorong, Xuan, Chai Peiqi, and Wu Minhui. "Bhattacharyya distance feature selection". *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, 195–199 vol.2. Aug 1996. ISSN 1051-4651.

37. Guyon, Isabelle, Jason Weston, Stephen Barnhill, Vladimir Vapnik, and Nello Cristianini. "Gene selection for cancer classification using support vector machines". *Machine Learning*, 2002.

38. Hagan, M.T. and M.B. Menhaj. "Training feedforward networks with the Marquardt algorithm". *Neural Networks, IEEE Transactions on*, 5(6):989–993, Nov 1994. ISSN 1045-9227.

39. Hameed, M.A, M. A Malik, S.F. Sayeedunnisa, and H. Imroze. "An Effective Hybrid Algorithm in Recommender Systems Based on Fast Genetic k-means and Information Gain". *Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on*, 860–865. Nov 2012.

40. Haran, T. *Short-Wave Infrared Diffuse Reflectance of Textile Materials*. Master's thesis, 2008.

41. Haykin, Simon. *Neural Networks and Learning Machines (3rd Edition)*. Prentice Hall, 3 edition, November 2008. ISBN 0131471392.

42. Haykin, S.S. *Neural Networks and Learning Machines*. Number v. 10 in Neural networks and learning machines. Prentice Hall, 2009. ISBN 9780131471399. URL `http://books.google.com/books?id=K7P36lKzI\_QC`.

43. Hersey, R.K., W.L. Melvin, and E. Culpepper. "Dismount modeling and detection from small aperture moving radar platforms". *Radar Conference, 2008. RADAR '08. IEEE*, 1–6. May 2008. ISSN 1097-5659.

44. Herweg, J., J. Kerekes, and M. Eismann. "Hyperspectral imaging phenomenology for the detection and tracking of pedestrians". *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 5482–5485. July 2012. ISSN 2153-6996.

45. Herweg, J., J. Kerekes, E. Ientilucci, and M. Eismann. "Spectral variations in HSI signatures of thin fabrics for detectin and tracking of pedestrians". *SPIE Defense, Security, and Sensing*, volume 80400G-80400G. 2011.

46. Hodson, John. "The Analysis Of Textile Materials And Textile Additives By Fourier Transform Infrared (FTIR) Photoacoustic Spectroscopy (PAS)", 1985. URL `http://dx.doi.org/10.1117/12.970740`.

47. Huang, C, LS Davis, and JRG Townshend. "An assessment of support vector machines for land cover classification". *International Journal of Remote Sensing*, 23(4):725–749, 2002.

48. Hughes, M., C.A.S. Hill, and J.R.B. Hague. "The fracture toughness of bast fibre reinforced polyester composites Part 1 Evaluation and analysis". *Journal of Materials Science*, 37(21):4669–4676, 2002. ISSN 0022-2461.

49. Hyuk-Gyu, Cho, Heum Park, and Hyuk-Chul Kwon. "Similarity Measurement among Sectors Using Extended Relief-F Algorithm for Disk Recovery". *Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on*, volume 2, 790–795. Nov 2008.

50. Inc, ASD. *RS3 User Manual*. 1 edition, 2007.

51. Inc, ASD. *FieldSpec 3 User Manual*. 1 edition, 2008.

52. Inc, ASD. *ViewSpec Pro User Manuall*. 1 edition, 2008. URL `http://dx.doi.org/10.1117/12.970740`.

53. Jiao, Hongzan, Yanfei Zhong, and Liangpei Zhang. "Artificial DNA Computing-Based Spectral Encoding and Matching Algorithm for Hyperspectral Remote Sensing Data". *Geoscience and Remote Sensing, IEEE Transactions on*, 50(10):4085–4104, Oct 2012. ISSN 0196-2892.

54. Khokher, Muhammad Rizwan, Abdesselam Bouzerdoum, and Son Lam Phung. "Crowd Behavior Recognition Using Dense Trajectories". *Digital lmage Computing: Techniques and Applications (DlCTA), 2014 International Conference on*, 1–7. Nov 2014.

55. Kovac, J., P. Peer, and F. Solina. "Human skin color clustering for face detection". *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, volume 2, 144–148 vol.2. Sept 2003.

56. Kwon, Heesung and N.M. Nasrabadi. "Hyperspectral Target Detection Using Kernel Spectral Matched Filter". *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, 127–127. June 2004.

57. Lee, Joo-Young and Jeong-Wha Choi. "Estimation of Regional Body Surface Area Covered by Clothing". *Journal of the Human-Environment System*, 12(1):35–45, 2009.

58. Liu, Li, Li Yan, Yaocheng Xie, Songzhang Li, Ge Xia, and Libin Zhou. "Content measurement of textile mixture by Fourier transform near infrared spectroscopy", 2009. URL `http://dx.doi.org/10.1117/12.838029`.

59. Morton, W.E. and J.W.S. Hearle. *Physical Propterties of Textile Fibres*. Halsted Press, 2 edition, 1975.

60. Muhammed, H.H., P. Ammenberg, and E. Bengtsson. "Using feature-vector based analysis, based on principal component analysis and independent component analysis, for analysing hyperspectral images". *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, 309–315. Sep 2001.

61. Nasrabadi, N.M. "Regularized Spectral Matched Filter for Target Detection in

Hyperspectral Imagery". *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 4, IV – 105–IV – 108. Sept 2007. ISSN 1522-4880.

62. Nunez, Abel S. *A Physical Model of Human Skin and its Application for Search and Rescue*. Ph.D. thesis, Air Force Institute of Technology, 2009.

63. Online., Merriam-Webster. *parallax*. Merriam-Webster. Accessed: 2015-01-24.

64. Parvinzadeh, M. and H. Najafi. "Textile Softeners on Cotton Dyed with Direct Dyes: Reflectance and Fastness Assessments". *Tenside Surfactants Detergents*, 45(1):13 – 16, 2008.

65. Peng, H., Fulmi Long, and C. Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, Aug 2005. ISSN 0162-8828.

66. Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 3 edition, 2007. ISBN 0521880688, 9780521880688.

67. Pudil, P., F.J. Ferri, J. Novovicova, and J. Kittler. "Floating search methods for feature selection with nonmonotonic criterion functions". *Pattern Recognition, 1994. Vol. 2 - Conference B: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International. Conference on*, volume 2, 279–283 vol.2. Oct 1994.

68. Rasekh, E., I. Rasekh, and M. Eshghi. "PWL approximation of hyperbolic tangent and the first derivative for VLSI implementation". *Electrical and Computer Engineering (CCECE), 2010 23rd Canadian Conference on*, 1–4. May 2010. ISSN 0840-7789.

69. Reyes-Aldasoro, C.C. and A. Bhalerao. "The Bhattacharyya space for feature selection and its application to texture segmentation". *Pattern Recognition*, 39(5):812 – 826, 2006. ISSN 0031-3203. URL `http://www.sciencedirect.com/science/article/pii/S0031320305004590`.

70. Robila, S.A. and A. Gershman. "Spectral matching accuracy in processing hyperspectral data". *Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on*, volume 1, 163–166 Vol. 1. July 2005.

71. Schaepman-Strub, G, T Painter, S Huber, S Dangel, ME Schaepman, J Martonchik, and F Berendse. "About the importance of the definition of reflectance quantities-results of case studies".

72. Schaepman-Strub, G, ME Schaepman, TH Painter, S Dangel, and JV Martonchik. "Reflectance quantities in optical remote sensingDefinitions and case studies". *Remote sensing of environment*, 103(1):27–42, 2006.

73. Sebe, N., I. Cohen, T.S. Huang, and T. Gevers. "Skin detection: a Bayesian network approach". *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, 903–906 Vol.2. Aug 2004. ISSN 1051-4651.

74. Serpico, S.B. and L. Bruzzone. "A new search algorithm for feature selection in hyperspectral remote sensing images". *Geoscience and Remote Sensing, IEEE Transactions on*, 39(7):1360–1367, Jul 2001. ISSN 0196-2892.

75. Song, Fengxi, Zhongwei Guo, and Dayong Mei. "Feature Selection Using Principal Component Analysis". *System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2010 International Conference on*, volume 1, 27–30. Nov 2010.

76. Thai, B. and G. Healey. "Invariant subpixel material detection in hyperspectral imagery". *Geoscience and Remote Sensing, IEEE Transactions on*, 40(3):599–608, Mar 2002. ISSN 0196-2892.

77. Uto, Kuniaki, Yukio Kosugi, Toru Murase, and Sigenori Takagishi. "Hyperspectral band selection for human detection". *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, 501–504. June 2012. ISSN 1551-2282.

78. Utschick, W., P. Nachbar, C. Knobloch, A. Schuler, and J.A. Nossek. "The evaluation of feature extraction criteria applied to neural network classifiers". *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, 315–318 vol.1. Aug 1995.

79. Wang, Hongxia, Kejian Yang, Feng Gao, and Jun Li. "Normalization Methods of SIFT Vector for Object Recognition". *Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2011 Tenth International Symposium on*, 175–178. Oct 2011.

80. Waxman, A, D. Fay, P. Ilardi, P. Arambel, Xinzhuo Shen, J. Krant, T. Moore, B. Gorin, S. Tilden, B. Baron, J. Lobowiecki, R. Jelavic, C. Verbiar, J. Antoniades, M. Baumback, D. Hass, J. Edwards, S. Henderson, and D. Chester. "Fused EO/IR detection & tracking of surface targets: Flight demonstrations". *Information Fusion, 2009. FUSION '09. 12th International Conference on*, 1439–1443. July 2009.

81. Wood, Frank. "Principal component analysis". 2009.

82. Wu, Zhongli, Bin Zhang, Yongli Zhu, Wenqing Zhao, and Yamin Zhou. "Transformer Fault Portfolio Diagnosis Based on the Combination of the Multiple

Bayesian Classifier and SVM". *Electronic Computer Technology, 2009 International Conference on*, 379–382. Feb 2009.

83. Yan, Li and Li Liu. "Quantitative prediction of cotton and wool mixture materials by BP neural network and NIR spectrometry", 2010. URL `http://dx.doi.org/10.1117/12.869394`.

84. Yao, Haibo and Lei Tian. "A genetic-algorithm-based selective principal component analysis (GA-SPCA) method for high-dimensional data feature extraction". *Geoscience and Remote Sensing, IEEE Transactions on*, 41(6):1469–1478, June 2003. ISSN 0196-2892.

85. Yeom, Jennifer S. *Textile Fingerprinting for Dismount Analysis in the Visible, Near, and Shortwave Infrared Domain*. Master's thesis, 2014.

86. Yu, Lei and Huan Liu. "Feature selection for high-dimensional data: A fast correlation-based filter solution". 856–863. 2003.

87. Yuan, Yuan, Guokang Zhu, and Qi Wang. "Hyperspectral Band Selection by Multitask Sparsity Pursuit". *Geoscience and Remote Sensing, IEEE Transactions on*, 53(2):631–644, Feb 2015. ISSN 0196-2892.

88. Zhang, Lefei, Liangpei Zhang, Dacheng Tao, Xin Huang, and Bo Du. "Hyperspectral Remote Sensing Image Subpixel Target Detection Based on Supervised Metric Learning". *Geoscience and Remote Sensing, IEEE Transactions on*, 52(8):4955–4965, Aug 2014. ISSN 0196-2892.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2015	Master's Thesis	Sept 2013 — Mar 2015

4. TITLE AND SUBTITLE

Spectral Textile Detection in the VNIR/SWIR Band

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Arneal, James A., Second Lieutenant, USAF

5d. PROJECT NUMBER

15G258

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-15-M-049

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
711th Human Performance Wing
Lead Scientist, Dr. Darrell F. Lochtefeld
2800 Q Street, Bldg 824
Wright Patterson AFB, OH 45433
(937) 255-2570 darrell.lochtefeld@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RHXBA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A.
Approved for Public Release; distribution unlimited.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Dismount detection, the detection of persons on the ground and outside of a vehicle, has applications in search and rescue, security, and surveillance. Spatial dismount detection methods lose effectiveness at long ranges, and spectral dismount detection currently relies on detecting skin pixels. In scenarios where skin is not exposed, spectral textile detection is a more effective means of detecting dismounts. This thesis demonstrates the effectiveness of spectral textile detectors on both real and simulated hyperspectral remotely sensed data. Feature selection methods determine sets of wavebands relevant to spectral textile detection. Classifiers are trained on hyperspectral contact data with the selected wavebands, and classifier parameters are optimized to improve performance on a training set. Classifiers with optimized parameters are used to classify contact data with artificially added noise and remotely-sensed hyperspectral data.
The performance of optimized classifiers on hyperspectral data is measured with Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. The best performance on the contact data is 0.892 and 0.872 for Multilayer Perceptrons (MLPs) and Support Vector Machines (SVMs), respectively. The best performance on the real remotely-sensed data is AUC = 0.947 and AUC = 0.970 for MLPs and SVMs, respectively. The difference in classifier performance between the contact and remotely-sensed data is due to the greater variety of textiles represented in the contact data. Spectral textile detection is more reliable in scenarios with a small variety of textiles.

15. SUBJECT TERMS

textile detection, dismount detection, hyperspectral imaging, feature selection, classification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt Col Jeffrey D. Clark (ENG)
U	U	U	UU	127	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4614 jeffrey.clark@afit.edu