3-21-2013

# Spectral Detection of Acute Mental Stress with VIS-SWIR Hyperspectral Imagery

Elizabeth K. Norvell


# SPECTRAL DETECTION OF ACUTE MENTAL STRESS WITH VIS-SWIR HYPERSPECTRAL IMAGERY

THESIS

Elizabeth K. Norvell, Second Lieutenant, USAF

AFIT-ENG-13-M-38

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

***AIR FORCE INSTITUTE OF TECHNOLOGY***

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

AFIT-ENG-13-M-38

# SPECTRAL DETECTION OF ACUTE MENTAL STRESS WITH VIS-SWIR HYPERSPECTRAL IMAGERY

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Insitute 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

Elizabeth K. Norvell, B.S.E.E.

Second Lieutenant, USAF

March 2013

AFIT-ENG-13-M-38

# SPECTRAL DETECTION OF ACUTE MENTAL STRESS WITH VIS-SWIR HYPERSPECTRAL IMAGERY

Elizabeth K. Norvell, B.S.E.E.
Second Lieutenant, USAF

Approved:

| | |
|---|---|
| /signed/ | 4 Mar 2013 |
| Lt Col Jeffrey D. Clark, PhD (Chairman) | Date |
| /signed/ | 4 Mar 2013 |
| Richard K. Martin, PhD (Member) | Date |
| /signed/ | 4 Mar 2013 |
| Michael E. Miller, PhD (Member) | Date |

AFIT-ENG-13-M-38


## Abstract

The ability to identify a stressed person is becoming an important aspect across different work environments. Especially in higher-stress career fields, such as first responders and air traffic controllers, mental stress can inhibit a person's ability to accomplish their job. A person's efficiency and psychological state in the work environment can be impeded due to poor mental health. Stress can result in harmful effects on the body, both physically and mentally, including depression, lack of sleep, and fatigue, which can lead to reduced work productivity.

Research is being conducted to detect stress in workload-intensive environments. This thesis implements an imaging approach that utilizes hyperspectral data across the visible through shortwave infrared electromagnetic spectrum. The data is applied to the feature selection algorithms ReliefF, Support Vector Machine Attribute Evaluator (SVM AE), and Non-Correlated Aided Simulated Annealing Feature Selection-Integrated Distribution Function (NASAFS-IDF) to obtain features that discriminate between the classes, "stress" and "non-stress." This data is classified using naive Bayes, Support Vector Machine (SVM), and decision tree methodologies.

The feature set and classifier that produce the highest classification results are calculated using percent accuracy and area under the curve (AUC). The reported results are divided into contact and non-contact (NC) validation sets. The contact validation returned a high accuracy of 96.30% and high AUC of 0.979. Validation on NC models returned a high accuracy of 99.64% and high AUC of 0.998.

**Acknowledgments**

I would like to first thank my advisor, Lt Col Jeffrey Clark, for his motivation, support, and dedication to seeing my successful completion of AFIT. I also appreciate the contributions and inputs of Dr. Michael Miller during the experimental and data analysis phases of the thesis. Lastly, I would not have made it through this program without the help and encouragement of my friends during coursework and struggles.

Elizabeth K. Norvell

## Table of Contents

Page

## List of Figures

## List of Tables

Table Page

## List of Abbreviations

| Abbreviation | | Page |
|---|---|---|
| HOL | Hemoglobin Oxygen Level | 23 |
| Hb | Hemoglobin | 23 |
| $(2D)^2PCA$ | 2-Directional, 2-Dimensional Principal Component Analysis | 23 |
| $HbO_2$ | Hemoglobin Oxygen Saturation | 26 |
| EO | Electro-Optic | 26 |
| $SO_2$ | Oxygen Saturation | 27 |
| WEKA | Waikato Environment for Knowledge Analysis | 28 |
| ROC | Receiver Operating Characteristic | 39 |
| TP | True Positive | 39 |
| FN | False Negative | 39 |
| FP | False Positive | 39 |
| TN | True Negative | 39 |
| AUC | Area Under the Curve | 40 |
| FOV | Field-of-View | 45 |
| WR | White Reference | 51 |
| VNIR | Visible-to-Near-Infrared | 51 |
| A/D | Analog-to-Digital | 51 |
| DC | Dark Current | 52 |
| CSV | Comma-Separated Value | 53 |

# SPECTRAL DETECTION OF ACUTE MENTAL STRESS WITH VIS-SWIR HYPERSPECTRAL IMAGERY

## I. Introduction

Recognizing stress is an important aspect of monitoring a subject's productivity and psychological status in the work environment. Levels of stress vary depending on the individual and the tasks encountered. Those in physically and emotionally demanding career fields, such as emergency personnel or air traffic controllers, often experience a higher workload and elevated stress levels compared to others in lower-stress environments. While certain aspects of stress can be positive, such as increased physical strength and alertness, negative results of stress, such as depression and reduced work efficiency are also possible [6]. This is due to the fact that a high level of stress causes physiological changes, releasing chemicals in the body that then affect cognitive processes and internal functions [6].

Because the negative aspects of stress outweigh the positives and inhibit work productivity, research is being conducted to examine the potential of detecting stress in workload-intensive environments [3, 4]. One way to detect stress is by imaging individuals in stressful situations with different types of cameras that can capture heat dispersal, reflected energy, or radiance. A thermal imager detects stress due to an increased temperature [4]. To capture reflected energy and radiance involves the use of hyperspectral imagers that capitalize on the change in reflectance signature when stressed.

Previous researchers have discovered the possible implementation of hyperspectral imaging (HSI) in stress detection [3, 4]. Yuen et al. [3] began exploration of HSI and its potential in the area of stress detection. The research accomplished in this thesis continues

HSI as a stress identification method and includes the use of a non-contact (NC) probe t o provide a non-invasive form of stress detection.

## 1.1 Problem Statement

Hyperspectral cameras collect radiance from a scene across the visible through the shortwave infrared (SWIR) spectrum [1]. The radiance is converted to reflectance, providing a spectral response of wavelengths versus reflectance. Radiance is a measure of the quantity of electromagnetic radiation that is emitted from the imaged surface [75]. Reflectance is the ratio of the radiation reflected from a surface to the total amount of radiation on the surface [75]. Every object has its own unique spectral response and is evident from the reflectance it produces. For example, the reflectance signature of skin differs from that of wood or clothing; across the spectrum, each object has peaks and valleys at different wavelengths.

The reflectance of skin has the same general shape independent of skin tones [7]. Nunez and Mendenhall [7] created an algorithm that successfully detects human skin among a cluttered background using HSI. Examining the unique reflectance of skin could prove viable in determining a method to perform stress detection.

Current research discusses the result of an increase in blood volume in skin at the onset of stress [3]. This property leads to a change in the reflectance of the skin, though only initial experiments have been conducted on this discovery [3, 4]. There are several steps required to confirm the hypothesis that HSI can be applied to stress detection. These include 1) Collecting hyperspectral data on a variety of skin tones under various stress levels; 2) Evaluating the features to determine discriminating wavelengths; and 3) Applying mathematical algorithms to separate stress from non-stress based on the selected features. Feature selection is important due to the high dimensional data produced by HSI. Hyperspectral data covers a large spectrum; the data collected in this work spans from 350-2500nm, with a sampling interval of 1nm, which equals 2,150 wavelengths.

Feature selection identifies features with the greatest discrimination between classes. Class distinction is determined by the amount of separation between the two classes, stress and non-stress. This thesis will test different feature selection and classification algorithms as they apply to hyperspectral data for the purpose of mental stress detection in a person.

## 1.2 Justification

HSI is non-invasive and collects details that are not distinguishable to the naked eye or to other types of technology, such as thermal and photoplethysmographic (PPG) imaging. Common methods to detect stress include an oximeter [9], which measures blood oxygen levels, a polygraph [8], which measures blood pressure, pulse, respiration, and skin conductivity, and a blood pressure monitor, which can detect potential stress as it relates to a change in blood pulse [10]. Thermal and PPG imaging have been implemented for stress identification [2, 4]; however, HSI is able to characterize features of the human skin with the ability to see below the epidermis [11]. HSI is a different type of imager than thermal or PPG imagers; it is focused across the visible to SWIR electromagnetic spectrum. Thermal imaging utilizes thermal heat produced and PPG imaging focuses on blood flow throughout the body [2]. PPG technology produces pixilated images showing blood pulse throughout the body, while HSI shows the reflectance of the skin.

Applying a stress detection method to the workforce and real-world scenarios necessitates a non-invasive technique. In the application of emergency personnel, pilots, air traffic controllers, or deep-sea divers, these individuals cannot be connected to a standstill device that impedes their agility. HSI can collect and process information using non-contact devices, to aid in stress detection.

## 1.3 Assumptions

Due to the complexity of detecting stress, several assumptions are made. These assumptions include an increase in heart rate (HR), blood volume, and blood oxygen levels

at the onset of stress. On average, a non-stressed subject has a temperate heart rate of 60 beats per minute (bpm). This provides a baseline reference for the HR measurements. The baseline is a reference of what the reflectance signature should look like when the subject is not stressed and can be used to compare the reflectance signature of a subject under stress.

Hyperspectral data from a spectroradiometer can be collected using two basic differing fore optics: a contact probe and a NC probe. The contact probe has a built in illumination source that spans the wavelength range of 350-2500nm. The NC fore optic can successfully take images in the sunlight without any extra lighting, but requires artificial light (spanning the range of 350-2500nm) if used indoors.

## 1.4 Approach

A subject's biological response to stress results in increased blood flow and blood oxygenation [4]. Therefore, the radiance from human skin under stress should hold a unique spectral distribution as blood volume and oxygen levels change with stress.

This study will be performed indoors using artificial light, which is designed to represent sunlight. Due to imperfections in the light source, when using the NC fore optic, power is attenuated in the lower (350nm) and higher (2500nm) spectral range. The signal-to-noise ratio (SNR) in these regions is not low enough to affect the results and will be considered negligible.

The contact sensor is placed on the skin in the area of the carotid artery. This site is chosen due to its ease of access for current testing and future implementation. The carotid is one of the largest arteries in the cardiovascular system and holds a strong pulse close to the surface of the skin, which increases imaging accuracy. The NC optic is positioned to collect the reflectance of the skin in the area of the carotid artery also. The spectral responses from these collections are applied to feature selection and classification algorithms in order to detect stress.

## 1.5 Equipment

Collecting hyperspectral data of the human stress response necessitates imaging of human skin. To accomplish these collections, an Analytic Spectral Devices (ASD) FieldSpec3®Pro spectroradiometer is used [12, 75]. The FieldSpec3 measures wavelength, absolute reflectance, radiance, and irradiance.

The spectrum output of wavelengths and reflectance values from the spectroradiometer is recorded and processed with $RS^3$ 5.7 software [12]. This software enables the user to optimize the FieldSpec3 instrument and collect various data types. This software converts the wavelength and reflectance data to a format compatible with the computer program MATLAB® [12]. MATLAB® is used to preprocess the data.

To bring about an accepted level of mental stress, subjects will interact with the Air Force Multi-Attribute Test Battery (AF_MATB) [5]. The AF_MATB provides a method to manipulate a subject's task load and impose different levels (high, medium, low) of mental stress, though only the "high" level will be applied in this thesis [5]. The original MATB software has become a mainstay for psychological research regarding cognitive workload and the version used in this research has updated software to be compatible with modern operating systems [5]. Subjects will use a standard laptop keyboard in addition to a USB joystick to perform the tasks.

## 1.6 Results

This thesis presents classification results for stress detection using the feature selection algorithms ReliefF [18–20], Support Vector Machine Attribute Evaluator (SVM AE) [21, 22, 34, 45], and Non-Correlated Aided Simulated Annealing Feature Selection - Integrated Distribution Function (NASAFS-IDF) [35, 77] and classification algorithms naive Bayes [23–25], Support Vector Machine (SVM) [26, 28, 29, 45], and decision tree [30–33]. Each algorithm is evaluated on datasets comprising subject's normalized reflectance and variance. Data is collected, trained, tested, and validated for each case: a

contact probe and a NC fore optic. The top performing feature selection and classification algorithm pairs are determined by average percent accuracy and average area under the curve (AUC), calculated from a receiver operating characteristic (ROC) curve. For validation using contact data with models trained/tested on contact data, NASAFS-IDF and SVM AE feature sets with a decision tree and naive Bayes classifier were found to have the highest accuracy and AUC. For validation using NC data with models trained/tested on NC data, SVM AE and ReliefF feature sets with a SVM and decision tree classifier return the highest results.

# II. Background

CERTAIN concepts of stress and stress detection are introduced in order to understand the work accomplished in this thesis. Stress detection has been researched for several years, with a variety of different attempts at achieving a successful model. There are some successes at detecting stress using various imaging techniques which are used as a starting point for the theories presented in this thesis and discussed in the following sections [3, 4, 36, 37]. Section 2.1 provides a brief introduction to the physiological processes that produce stress detection characteristics. Different methods that have been used for stress and human skin detection are reviewed in Section 2.2. The correlation between stress attributes and detection methods is presented in Section 2.3. Sections 2.4 and 2.5 detail the feature selection and classification algorithms implemented in this thesis and Section 2.6 addresses the classification training and output. Section 2.7 addresses the feature generation method applied to the hyperspectral data, which produces stress detection attributes.

## 2.1 Biological Effects of Stress

Stress is a rapid transformation of bodily chemicals that results from a perceived threat or possible danger and that causes physical and physiological changes to the human body [4]. Stress manifests in two forms: physical and emotional. Physical and emotional stressors are caused by different effects and produce different physiological conditions. Physical stress occurs when the body is directly affected by a physical outside source [3]. Common physical stressors include exercise, external strain, and environmental conditions, e.g. heat, cold, or noise. Emotional stress is produced when the brain is overwhelmed with psychological processes, as opposed to physical effectors [3]. An emotional stressor affects the cognitive or emotional systems. A stressor involving the cognitive system manifests

in the form of anxiety, for example, the reaction to a quiz or speech, whereas a stressor involving the emotional system is produced by a reaction to a physiological event, for example, an argument producing anger or intimidation [3]. In both of these forms of stress, the body reacts with a surge of adrenaline to the bloodstream, which is produced by the hypothalamus, the pituitary gland, and adrenal gland secretions [3]. This leads to the fight-or-flight response, which induces biological changes; for example:

- accelerated heartbeat;
- elevated blood pressure;
- increase in red and white blood cells released by the spleen to deliver more oxygen to the body;
- redirected blood to augment the brain, muscles, and heart;
- nutrient dispersal for increased muscle capability;
- blood vessel constriction in many parts of the body, such as the skin, stomach, and intestines; and
- increased sweat [4].

The responses listed above represent the general reaction of a person when stress is experienced. The degree of change for each physiological response varies from person to person. However, previous clinical research [51] discovered that the combination of the first five items results in an elevated blood volume during stress with an approximate 100% increase.

Heart rate (HR) and heart rate variability (HRV) are also used as indicators of stress [71]. Much research has been accomplished that indicates as mental workload increases, HR increases [66, 71] and HRV decreases [66, 67, 71]. There is not a medically

proven value or range of beats per minute (bpm) for which stress is determined to have occurred, therefore the recorded HR will be compared to a baseline HR as a baseline indication of stress. HRV (averaged over a five minute period) is a better indication of stress because it is relatively constant at a resting state, especially compared to HR, which is continuously vacillating [66]. HRV is the variations of instantaneous HR and respiratory rate (RR) intervals [68]. RR fluctuation is one of the most commonly investigated components of HRV because it primarily reflects respiration-driven vagal modulation of sinus arrhythmia [69]. Respiratory sinus arrhythmia (RSA) is an innately occurring variation in HR during the breathing cycle [70]. It is represented as an increase in HR upon inhalation and decrease in HR upon exhalation [70]. This thesis utilizes a 3-lead electrocardiogram (ECG) to continuously record HR and HRV throughout the experiment. The ECG data is used to validate a subject's state of stress.

## 2.2 Detection Methods

Detection of stress is accomplished by using various sensors to examine the physiological changes listed in Section 2.1. Physiological changes occur simultaneously when a body is stressed; this causes a rise in body temperature, rapid blood pulsations, and a 100-200% increase in blood oxygenation [4]. Common methods used to detect these three changes include a blood pressure monitor [10, 65], a thermometer [4], and an oximeter [3, 9]. A blood pressure monitor detects the blood pulsation, which spikes at the onset of stress [4]. A thermometer measures body temperature, providing a baseline and the recognition of a change in thermal heat [4]. An oximeter measures blood oxygen levels, which when elevated, indicate stress [3]. These methods identify stress, but require direct contact, which is physically intrusive. Advanced imaging systems that can detect these changes through non-contact (NC) methods include thermal infrared sensors [4, 13–15], photoplethysmographic (PPG) imaging [2, 16, 17], and hyperspectral

imaging (HSI) [1, 3, 7, 11, 38–40, 42–44]. These three approaches differ in their techniques of detection, however, they all produce attributes that can be used for stress detection.

### *2.2.1 Thermal Imaging.*

Thermal imaging [4, 13–15] is a technique that uses special cameras that are sensitive to very small spectral changes. This process records radiation levels across the wavelengths in the infrared spectrum. Radiation levels increase as temperature elevates, which allows this method to be effective at stress detection. At the onset of stress, temperature throughout the body tends to increase, which reflects a change in radiation; this change can be captured by thermal imaging [4]. Dr. Ioannis Pavlidis, of the HoneyWell Corp Laboratory, discovered that stress caused from a sudden excitement or a startle resulted in raised blood volume levels in the facial region [4]. Using thermal imaging, an increase in blood volume to the surface of the skin is detected as "hot" pixels in the image. Physical and emotional stressors cause different physiological changes in the body as shown in Fig. 2.1 [4]. This experiment demonstrates that there is a higher temperature increase in the forehead region due to emotional stress than due to physical stress.

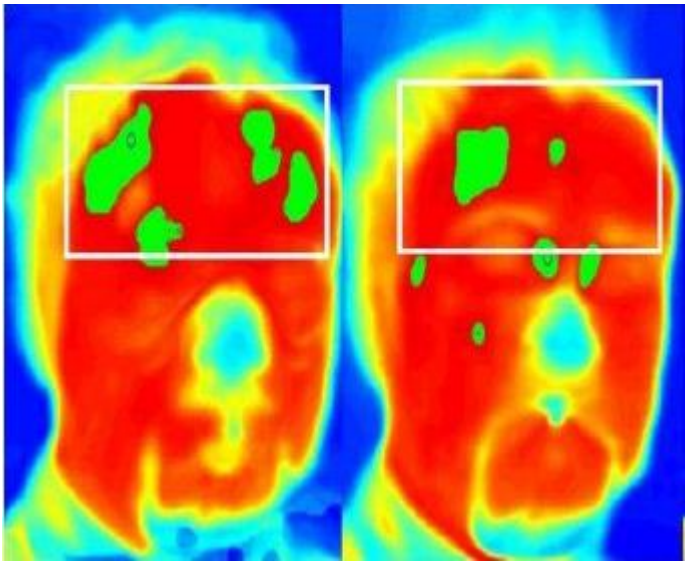


**Figure 2.1: The dual image shows the result of an experiment conducted by Yuen, et al., using thermal imaging to show the outcome of different types of stressors. The left frame is a thermal image captured after the subject experiences emotional stress. A high proportion of "hot" pixels (green) are detected in the forehead region. In the right frame, a physical stressor is administered and a thermal image is taken. The resulting thermal detection shows some "hot" pixels in the forehead, but not as many as that of emotional stress. There are other "hot" pixels located in other regions of the face when stress occurs [4].**

### *2.2.2 Photoplethysmographic Imaging.*

Photoplethysmographic imaging [2, 16, 17], with high spatial resolution, is used to remotely record a blood pulse throughout the body. This method uses a light source for illumination and a photodetector to record small changes in the reflected energy due to the changing light intensity [2]. A relationship exists between the intensity modulations of reflected light from the skin due to a person's heartbeat. Kamshilin et al., from the University of Finland, developed a method to show blood pulsations, represented as light-intensity modulations, by manipulating image data in mathematical software, such as MATLAB® [2]. After recording several frames of data over a period of time, as in Fig. 2.2, the authors create a reference function, $R_C(t)$, as in Eq. (2.3). This reference function is of a specific region-of-interest (ROI) and results from averaging pixels for each reference frame. Using this information, the fast Fourier transform, in Eq. (2.1), is applied to obtain a cardiac pulsation and breathing reference function,

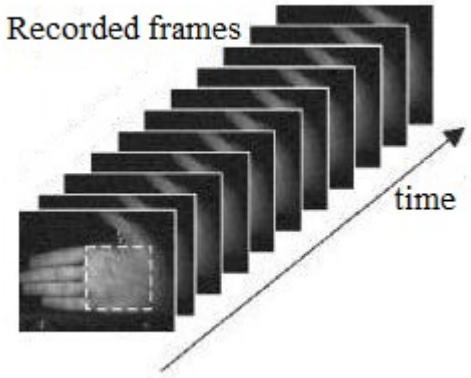



**Figure 2.2: This image, taken from Kamshilin's et al. work on PPG imaging, provides an illustration of the recorded frames in a specific ROI are for a given timeframe. The pixels in each recorded frame ROI averaged to result in one pixel for that image, creating a vector of mean valued pixels. Using Fourier analysis, the cardiac and respiration cycles are detected from this information [2].**

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi k \frac{n}{N}} \quad k = 0, ..., N-1 \tag{2.1}$$

where $x_n$ is a continuous-time signal, $k$ is the frequency increment, $n$ is the incrementing sample number, and $N$ is the total number of samples in the transform.

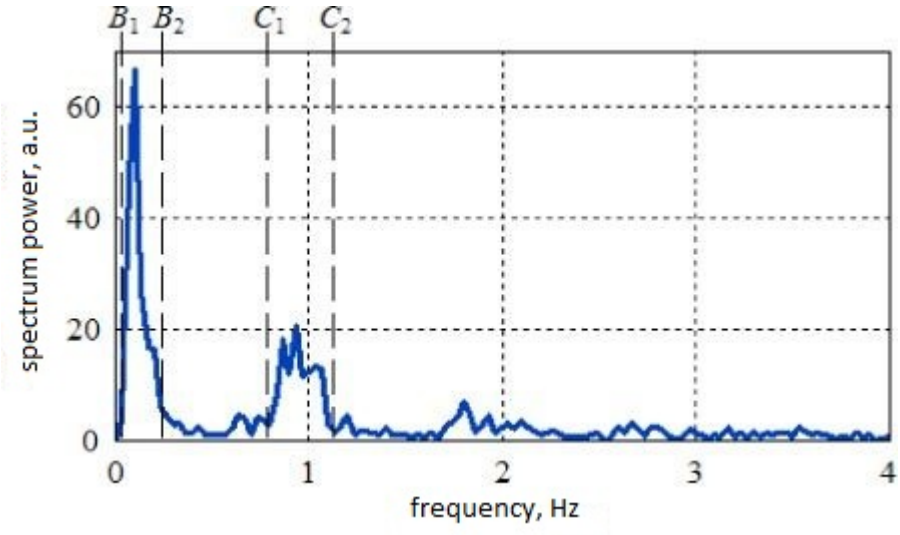



**Figure 2.3: A continuation of Kamshilin's work, this is the power spectrum resulting from the fast Fourier transform applied to the type of data shown in Fig 2.2. Two prominent spikes are indicated in the graph, the first one at approximately 0.13Hz, which corresponds to the respiration rate, and the second at approximately 1.0Hz, which corresponds to the heartbeat rate. The researchers select and use these two narrow bands, $B_1$-$B_2$ and $C_1$-$C_2$, as reference functions for the breathing cycle, and the cardiac cycle, respectively [2].**

The specific frequency bands associated with heartbeat and breathing are selected and all other bands are truncated. The power spectrum showing the frequency bands can be seen in Fig. 2.3 with the breathing bands noted by $B_1$ and $B_2$ and heartbeat by $C_1$ and $C_2$. The reference function is reconstructed with the inverse Fourier transform,

$$f(t) = \int_{-\infty}^{\infty} f(\xi) e^{j2\pi\xi t} d\xi, \tag{2.2}$$

where $f(\xi)$ is the continuous-time function represented in the Fourier domain, $t$ is a time, and $\xi$ is the frequency sample, which is then representative of heart pulsations [2]. A normalized reference function with $N$ samples and a frequency component, $f$, multiplied by a time variable, $t$, is

$$R_C(t) = \frac{2}{N} \exp(j2\pi f t). \tag{2.3}$$

Next, $R_C(t)$ is multiplied by the set of image frames of the ROI and corresponding pixel value at time $t$ creating a correlation matrix, $S_C(x, y)$. Figure 2.4 gives a visualization of the

mathematical process conducted. The correlation matrix is equal to

$$S_C(x, y) = \sum_t I(x, y, t) R_C(t) \tag{2.4}$$

where $R_C(t)$ is the reference function and $I(x, y, t)$ is the pixel coordinate $(x, y)$ at time $t$, which is equal to

$$I(x, y, t) = A(x, y)\cos[i2\pi ft + \psi] + B(x, y). \tag{2.5}$$

In Eq. (2.5), $A(x, y)$ is the amplitude of the pixel values at frequency $f$ and time $t$, $\psi$ is the phase of the pixel oscillations, and $B(x, y)$ is the mean pixel value. The correlation matrix synchronously corresponds to the time variation of the pixel values with the heartbeat. This matrix represents the PPG image since the modulated amplitude of the reflected light is represented by each pixel value in the matrix. The authors determined that the blood pulsations do not always occur with the same phase. Therefore, they implemented a new series of frames, resulting in a new matrix of values that determine the phase shift [2],

$$H_c(x, y, t) = Re[S_C(x, y)]\cos[\phi(t)] + Im[S_C(x, y)]\sin[\phi(t)] \tag{2.6}$$

where $\phi(t) = 2\pi f_C t$ and $f_C$ is the mean rate of heartbeats.

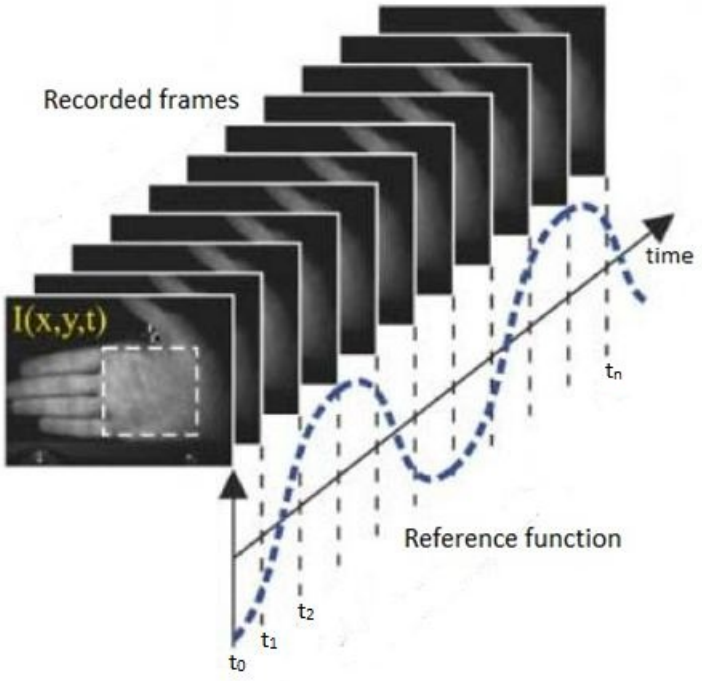



**Figure 2.4: The images from Kamshilin et al. [2] of the palm with outlined ROI represent the recorded frames that are multiplied by the reference function to create a correlation matrix, $R_C(t)$. $R_C(t)$ is multiplied by the individual frame at each time increment. According to Eq. 2.4, the image frame is modeled as the function $I(x, y, t)$, which contains the pixel value coordinates $(x, y)$ at time $t$ [2].**

In Cui et al. [36], PPG technology is used to examine the reflectance of blood and tissue in human skin. The results of the mathematical equation for modulation of light,

$$\frac{\delta R}{R} = \frac{-1}{\sqrt{w(w+2k)}}\left[\frac{w}{k}(k_t - k_b) + (w_{ba} - w_t)\right]\delta V_b, \tag{2.7}$$

show that blood-oxygen levels are less correlated with peak modulation of light in the red and infrared range than modulation in the 630-940nm range [36]. In Eq. (2.7), $(k, w)$ are the scattering and absorption coefficients of human tissue respectively, $(k_b, w_{ba})$ are the scattering and absorption coefficients of blood in the skin tissue respectively, $(k_t, w_t)$ are the scattering and absorption coefficients of bloodless tissue respectively, and $\delta V_b$ is a dimensionless quantity pertaining to the fractional volume of blood per volume of tissue [36].

Experimental data, collected using a contact probe, showed that non-pulsating blood resulted in a decrease in reflectance for wavelengths 450-600nm [36]. The authors also discovered that skin pigmentation does not affect the shape of the modulation spectrum, since coloring only occurs in the epidermis, where no blood exists [36]. In general, Cui et al. show that longer wavelengths result in deeper light penetration and that electromagnetic energy ranging from 510-590nm provides the maximum pulsation modulation based on reflectance measurements [36].

In a similar study performed using PPG imaging, specific cardiac and respiratory bands across the spectrum were identified in the visible and near-infrared (NIR) range. Corral et al. examined bands containing the greatest reflectance power relating to the heart and breathing rates by imaging the forehead region [37]. The authors observed the power spectrum output after filtering and extracting the desired features, which was obtained from narrow frequency bands instead of single frequencies [37]. Figure 2.5 is a block diagram detailing this process.

The recorded raw data was filtered to improve the signal-to-noise ratio (SNR) for each parameter, HR and RR. This is performed using a 6th-order high-pass filter with $f_c = 0.416$Hz to obtain the HR parameter and a 6th-order band-pass filter with $f_l = 0.133$Hz and $f_h = 0.5$Hz for the RR parameter [37].

To extract the HR and RR parameters, the authors obtained the power spectrum of each parameter from 380-980nm [37]. Peak power occurred at approximately 590nm for the HR and 710nm for the RR [37]. These numbers were verified from the average rates recorded by the oximeter. Because the HR and RR were not constant, small bands surrounding the mean frequency add to the frequency range. The cardiac representation adds ±3 bpm and breathing adds ±1.5 repetitions per minute (rpm) [37]. After the bands were determined, peak power is identified. The peak power is removed and the next highest value in the power spectrum represents the peak noise signal.

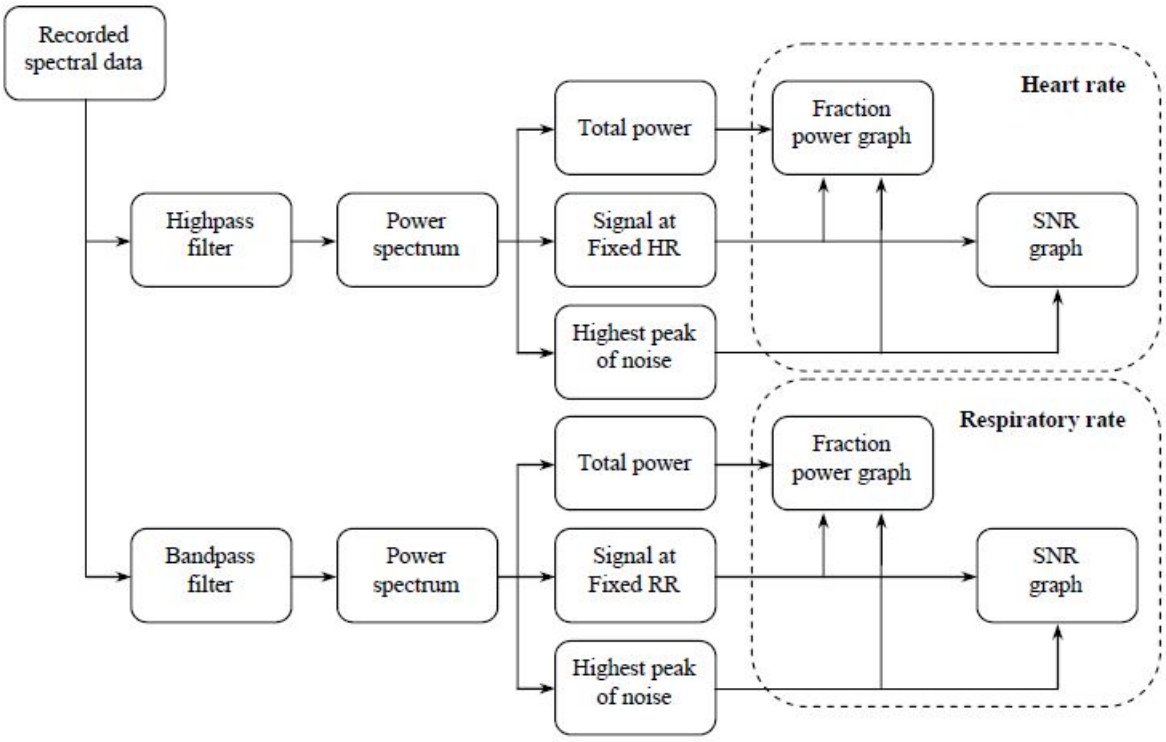
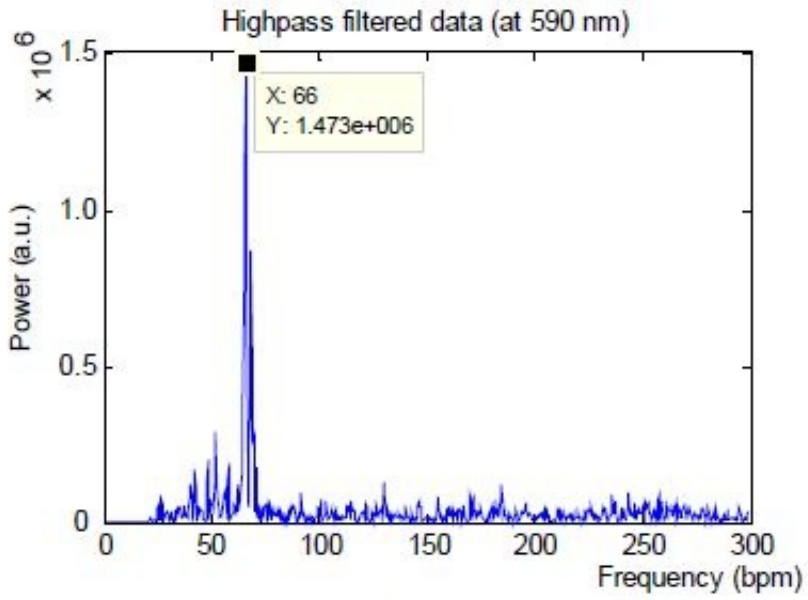
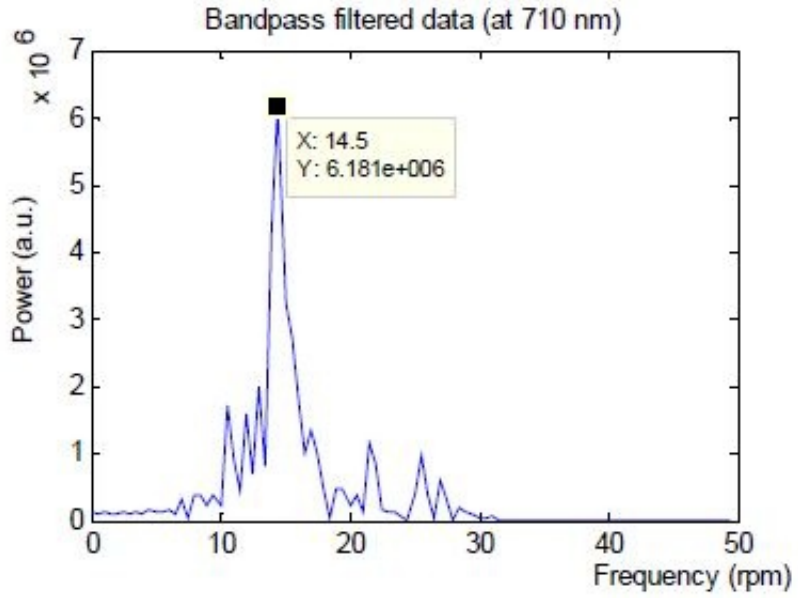



**Figure 2.5: This block diagram, from Corral's et al. [37] work on the PPG process, represents the task flow required to identify maximum power signals from the HR and RR. The filtered data was the recorded raw input. Both the HR and RR step through similar processes, but slightly different values were used for identification. The HR bands are located around 590nm and the peak respiratory bands around 710nm [37].**

Figure 2.6 shows the results of the power spectrum plotted versus wavelength [37]. With reference values identified, Corral et al. examined the visible to NIR wavelengths to determine the maximum power for each parameter [37]. Signal power was computed across the bands 66 ± 3 bpm for cardiac and 14.5 ± 1.5 rpm for respiratory, then this value is subtracted from the filtered data to extract the appropriate signal [37]. They calculate the noise power using the maximum peak power values of the remaining bands and determine the SNR for both the HR and the RR parameters [37]. Figure 2.7(b) and 2.8(b) highlight wavelength bands that have the highest SNR [37]. The selected bands were 480-610nm and 800-925nm for HR detection and 450-490nm and 680-900nm for RR detection [37]. These bands were identified from the requirement that the SNR be at least 50% greater than the maximum noise power, providing an SNR of 1.5 [37].

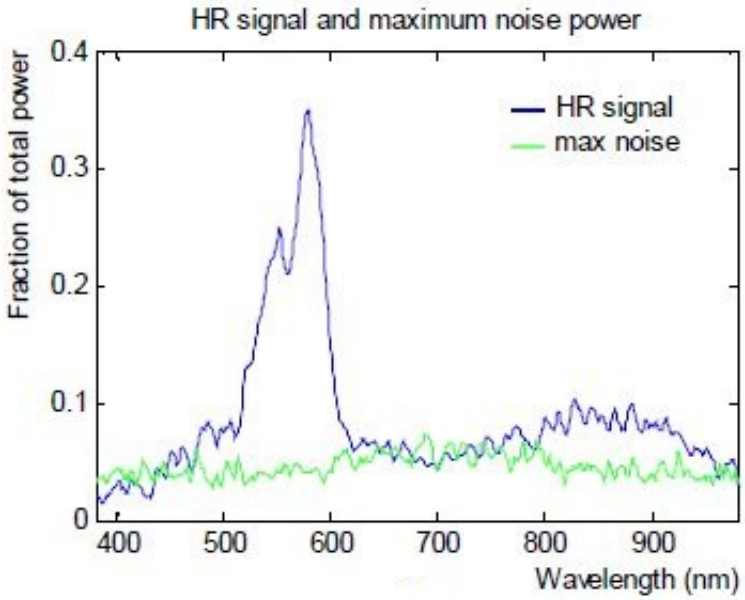



(a)

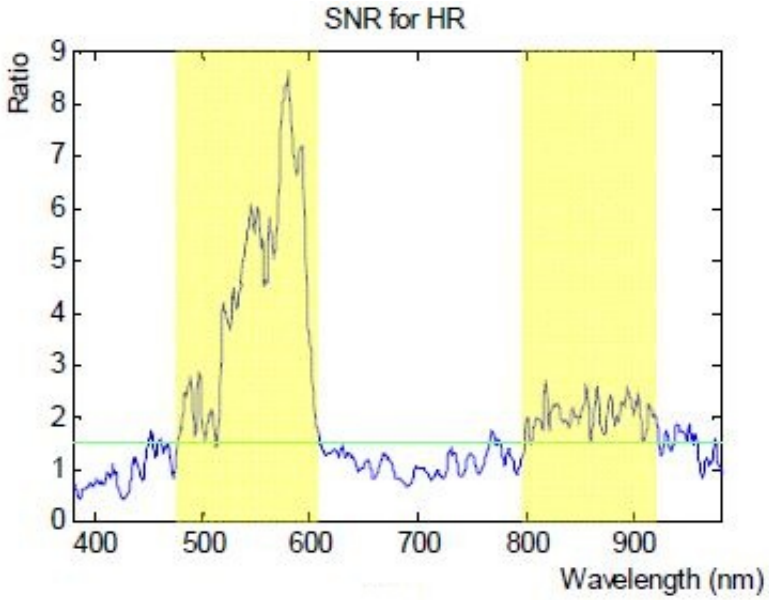



(b)

**Figure 2.6: Taken from Corral et al. [37], this shows the power spectrum plotted with respect to the frequency. (a) shows the HR with the highest spectral power for filtered data at 590nm at 66 bpm. The RR in (b) has a peak at 710nm at 14.5 rpm.**

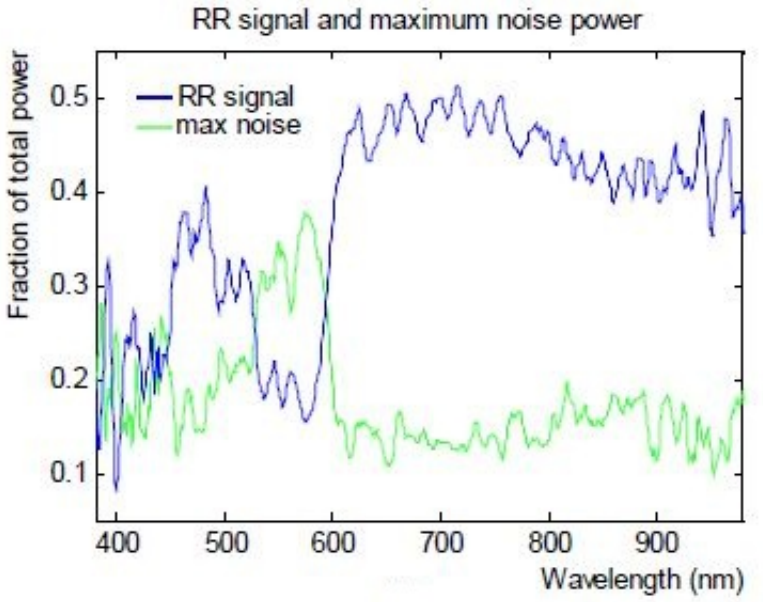



(a)

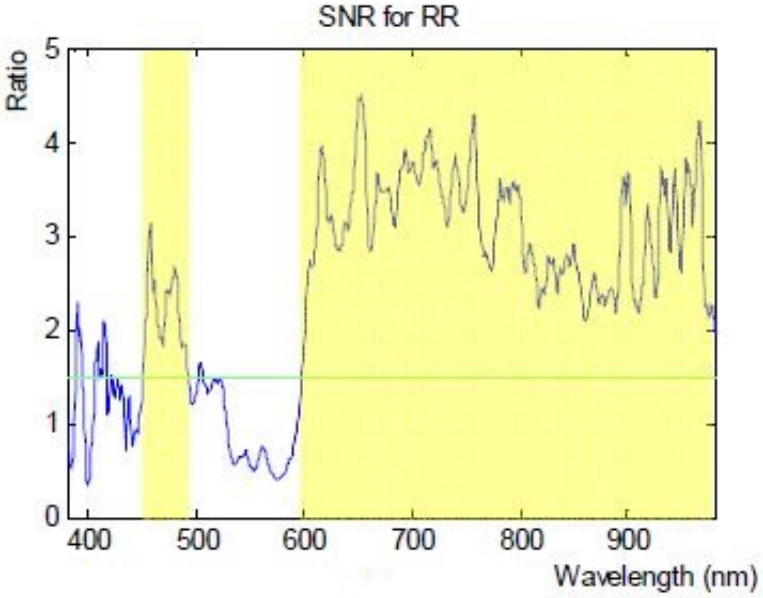



(b)

**Figure 2.7: The two graphs from Corral et al. [37] show the final process for optimal wavelength determination. The goal is to find the band of wavelengths that have the highest SNR. (a) is the HR signal and maximum noise power. (b) shows the bands that have a SNR of at least 1.5, which means there is 50% greater signal power than noise power. There are two bands that meet this criteria: 480-610nm and 800-925nm [37].**




(a)




(b)

**Figure 2.8: Taken from Corral et al. [37], the two graphs show the final process to determine optimal wavelengths applied to the RR signal. This process is similar to finding the optimal wavelength for HR in that the goal is to find the highest SNR. (a) shows the RR signal and maximum noise power. (b) shows the bands that hold a SNR of at least 1.5, resulting in a signal power that is 50% greater than noise power; these bands are 450-490nm and 600-980nm [37].**

### *2.2.3 Hyperspectral Imaging.*

Hyperspectral imaging [1, 3, 7, 11, 38–40, 42–44] is a form of data collection that uses hyperspectral cameras to collect radiance from a scene. Hyperspectral data contains high spectral resolution information across the spectrum, from the visible to the shortwave infrared [1]. This type of data displays unique characteristics of materials often missed by multispectral images.

The reflectance of an object can be determined from its radiance [1]. Absolute reflectance in a scene is characterized by Spectralon calibration panels. These panels allow for proper normalization of the reflectance data. Huynh et al. discuss the images as a pixel-based classification task, where each pixel has a different spectral signature [1]. These spectral signatures are representative of the various materials in an image, each with their own distinguishing characteristics across the wavelengths collected. Figure 2.9 shows an example of a spectral response from an HSI collection on human skin, where four subjects were imaged, each producing slightly different spectral responses depending on the amount of melanin in their skin [38].

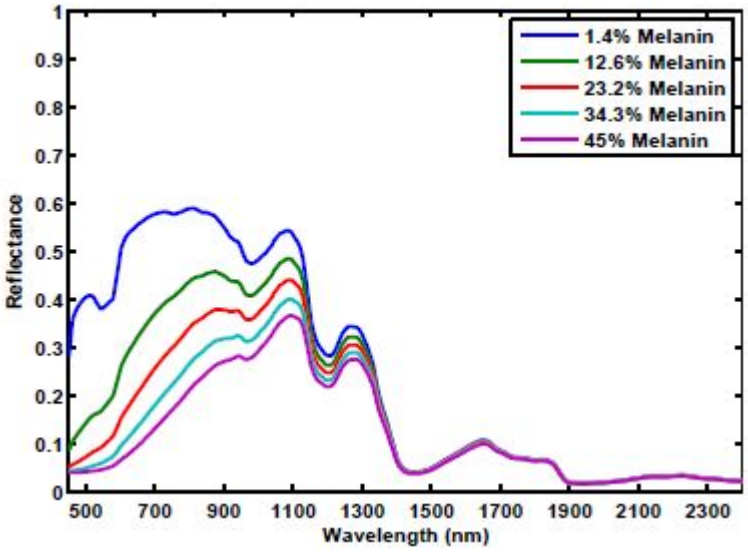



**Figure 2.9: This image, taken from Beisley's thesis on dismount detection shows the reflectance response of four different dismounts with various skin pigments. This graph shows that as melanin increases, the reflectance decreases [38].**

#### *2.2.3.1 Hyperspectral Imaging Applications.*

Hyperspectral imaging is implemented in a variety of research areas; this includes various skin-related imaging applications, such as skin detection in a cluttered scene [42], differentiating between ethnicities by examining varying properties of skin reflectance [1], detecting various physical properties located in the skin and organ tissues [39], and non-skin-related applications, such as imaging for food quality and safety [52–55]. In the medical field, HSI is used to classify and detect blood vessels during surgery [39, 56, 57]. Current research shows that HSI has potential to assist physicians and surgeons by providing previously unavailable information about their patients [42, 59–61]. The hyperspectral information collected on a patient allows researchers to examine correlations to medical problems, possibly aiding future diagnoses [39, 58, 62]. For example, a group at the National Institutes of Health, Laboratory of Chemical Physics, examined brain and breast tissue using NIR imaging and identified distinguishing features of organs and bodily tissues [42].

Currently, there is a focus on conducting research of skin detection and recognition using HSI. By employing HSI techniques, many discriminating features can be examined, providing an improved detection method [3, 11]. Pan et al. [11] are finding that with HSI, deeper skin layers can be imaged, producing results that are more distinguishing than those from surface level collection. NIR wavelengths are able to penetrate the skin deeper than the visible wavelengths [11]. The penetration depth is determined by the thickness of the skin tissue at which light intensity is reduced to 37% of that at the surface [11]. By imaging the subsurface skin features, distinguishing skin types becomes less complicated because such features cannot easily be altered [11].

Researchers at Purdue University, in conjunction with Polytechnic Universities in China, are using feature band selection to determine spectral response wavelengths that contain the most information regarding distinguishable characteristics of skin [40]. Some

of the features that they discovered include blood flow, hemoglobin oxygen level (HOL), water concentration, melanin concentration, aging, perspiration, and cosmetic makeup [40]. From these common skin characteristics, the most contributing types of light-absorbing chemical compounds of skin tone are carotene, melanin, and hemoglobin [40]. Of these, carotene plays a relatively insignificant role when compared to hemoglobin (Hb) [40]. Melanin is mostly a product of environmental factors, such as sunlight exposure, and therefore, it is not a major contributor to a common skin signature [40]. Figure 2.10 shows the absorption characteristics of hemoglobin and melanin across wavelengths 400-700nm [40]. Di et al. note two peak hemoglobin absorption bands, around 540nm and 580nm [40]. They did not account for the peak around 420nm due to a low SNR ratio of their system. The researchers implemented three types of 2-directional, 2-dimensional principal component analysis ($(2D)^2PCA$) feature selection methods that were successfully used for facial recognition in images [41]. These methods validated that the selected absorption bands are in fact the most significant. $(2D)^2PCA$ confirmed that these wavelengths result in a higher degree of facial recognition than using a single band or the entire band [40].

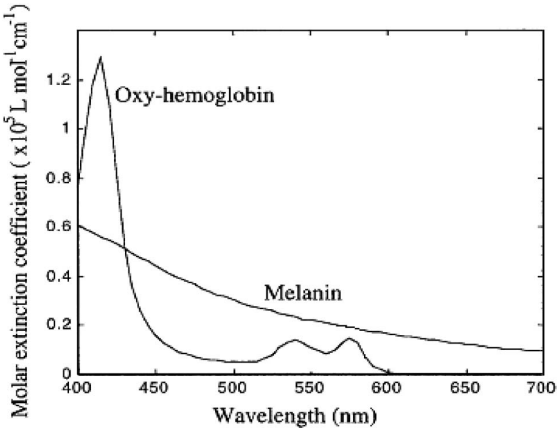



**Figure 2.10: This graph, taken from Di's et al. [40] work on hyperspectral facial recognition, shows the absorption characteristics of hemoglobin and melanin in *in vivo* human skin. Notice two small peaks on the oxy-hemoglobin line at the 540nm and 580nm range. These correspond to hemoglobin absorption bands. The peak at 420nm is not considered because of a low SNR at this band. The melanin curve shows that at lower wavelengths, the skin absorbs more light, resulting in a higher melanin reading [40].**

## 2.3 Hyperspectral Imaging for Stress Detection

Hyperspectral imaging produces data that identifies characteristics that are not visible to the naked eye. Medicinal fields are using this type of data for classification and identification of diseases, as well as for differentiation of physiological conditions that are not easily distinguished by the human eye [43, 58, 60]. HSI has been proven to be a viable collection method for accurate target detection [1, 3, 11, 39, 43]. HSI's application to security, surveillance, and target acquisition by Yuen et al. achieved 100% success of target detection in a field of vegetation, but only 60% success in a desert environment [43]. A relatively new use of HSI involves the classification and detection of human stress.

Researchers are currently looking at the changes of blood oxygenation in the facial region at the onset of stress [3, 4]. Yuen et al. [3] conducted an experiment to identify the affects of stress on blood pressure, coronary venous flow, oxygen extraction, and oxygen consumption by controlled adrenaline injections in a dog. These results are displayed in Fig. 2.11 [3]. The adrenaline injection (2 $\mu$g/kg·min) is shown with the arrowed point in the graph. Observations included a dramatic increase in blood pressure (top), an approximate increase of 90% of oxygen in the blood (second), and a drop in the oxygen extraction ratio (third) because the oxygenation consumption (fourth) remained relatively unchanged [3]. This work indicates stress can be successfully diagnosed based on the blood oxygenation levels [3].

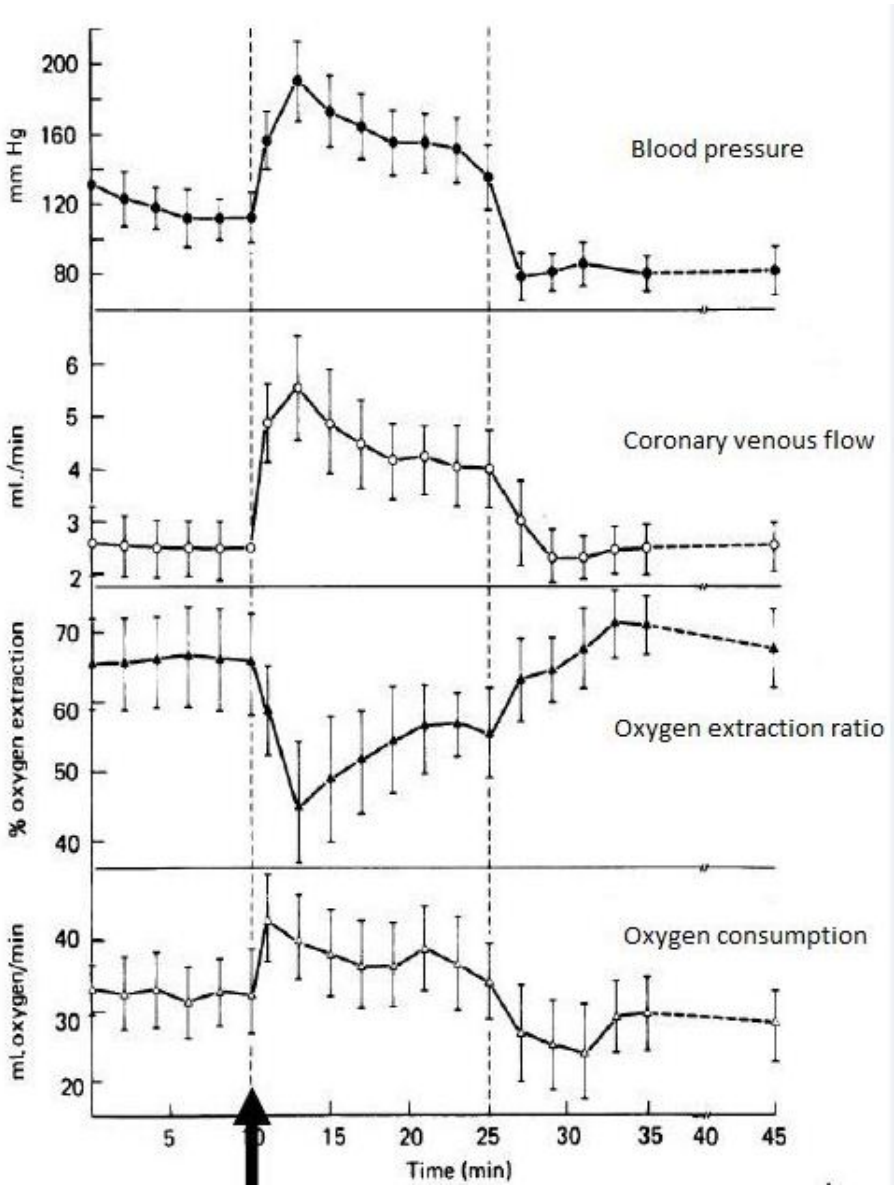



**Figure 2.11: Yuen et al. conducted an experiment to observe the affects of stress on blood pressure (top), coronary venous flow (second), oxygenation extraction (third), and oxygen consumption (fourth). The experiment involved injecting controlled amounts of adrenaline (2 $\mu$g/kg·min at the arrow) into a dog while making observations. The researchers noticed an increase in blood pressure, a 90% increase of oxygen in the blood, and a decrease in the oxygen extraction ratio, which is attributed to the fact that the oxygenation consumption of tissues remained relatively constant [3]. Overall, there is an increase in blood oxygenation of approximately 100-200% [3]. These observations support the theory of stress detection using HOL [3].**

Emotional and physical stress results in a surge of adrenaline into the bloodstream, aids increased activity of the brain, muscles, and heart [3]. Along with this physiological change, there is an elevated HOL (approximately twice the usual amount), which is the

ratio of hemoglobin oxygen saturation ($HbO_2$) to the total concentration of Hb and $HbO_2$ in the blood [3]. $HbO_2$ is formed when oxygen-depleted Hb binds to oxygen; each of these molecules has differing optical properties that result in distinct electro-optic (EO) characteristics [3]. Hb is naturally a purple-blue color and when it binds to oxygen, creating $HbO_2$, it becomes bright red [3]. The chemical difference between the two molecules is exhibited by their molar extinction coefficients, as displayed in Fig. 2.12 [3]. The two molecules have peak absorption regions in the 410nm range and 550nm range [3].

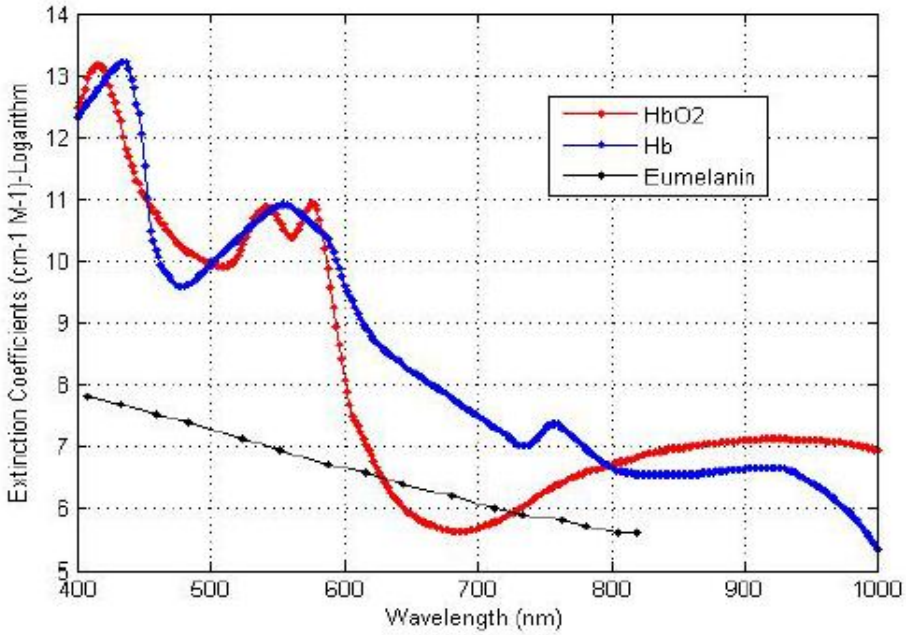



**Figure 2.12: These are the molar extinction coefficients (proportional to absorptivity) of $HbO_2$, Hb, and melanin. This chart shows that $HbO_2$ absorbs electromagnetic waves at wavelengths around 410, 545, and 578nm, and Hb around wavelengths 415 and 555nm [3]. It shows that melanin's absorption varies linearly with wavelength [3].**

Due to its optical properties, blood oxygenation, measured by HOL, provides researchers with a possible method to detect stress [3]. HOL can be determined from pixel reflectance values in a hyperspectral image [3]. In [3], researchers developed two algorithms for stress detection using HOL and the Beer-Lambert formulation. The Beer-

Lambert formula is,

$$\mathbf{A} = \sum \alpha_{\mathbf{i}} \mathbf{C}_{\mathbf{i}}, \tag{2.8}$$

where **A** is the attenuation of probing light, $\alpha$ represents the wavelength dependent absorption coefficients, and **C** represents the molecular concentrations of Hb and $HbO_2$ for each sample $i$ [3]. The researchers were able to identify both physical and emotional stress using this technique, though they noted a need for further enhancement of the model due to subjective baseline information [3]. The baseline recordings showed variations in the $HbO_2$, Hb, and oxygen saturation ($SO_2$) concentrations between and among individual subjects [3]. It was also noted that each collection depends on a subject's personal health, mood, and activity level at the time of the recording [3]. Figure 2.13 shows two images representing the change in $HbO_2$ in the facial region as a subject undergoes emotional stress [3].

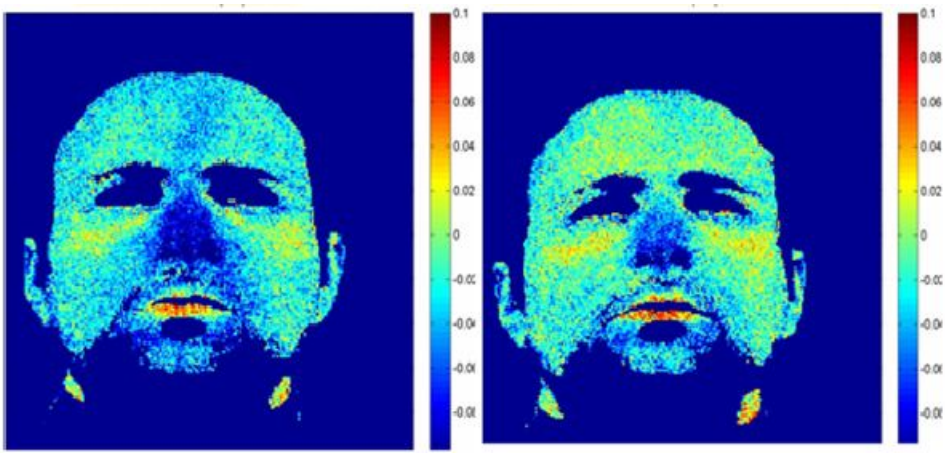



**Figure 2.13: This displays the change in $HbO_2$ located in the facial region as a subject undergoes emotional stress: (left) baseline, (right) emotional stress [3]. The subject was imaged at rest (left), then imaged after making a speech in order to bring about emotional stress. It is observed that there is an increase in "hot" pixels in the regions of the forehead, cheek, and lip, indicated by the yellow to red coloration. "Hot" pixels represent an increase in skin temperature.**

Additionally, research involving blood volume levels has been proven successful using remote HSI [3, 4]. Yuen et al. [4] found that the facial region shows a distinctive increase

in HOL that can be detected through HSI. One study that examined $HbO_2$ in relation to sickle cell disease patients found that there is a linear relationship between $HbO_2$ in skin tissue and oxygen saturation in venous blood [44]. They examined $HbO_2$ in the small blood vessels that are responsible for distribution of blood within tissues. The authors concluded that this can correlate $HbO_2$ with oxygen saturation of venous blood in underlying skin tissue. They also discovered that the normal percentage of skin $HbO_2$ is about $77.5 \pm 0.2\%$ for African-Americans and similarly, $78.2 \pm 0.2\%$ for Caucasians [44].

Overall, there has only been initial research accomplished on stress detection using HSI. Of the research accomplished, there has not been a grave attempt at feature reduction, via feature selection or classification using common machine learning techniques. This thesis applies three feature selection algorithms, three classification algorithms, and a feature generation technique to hyperspectral data. These are discussed in the following sections.

## 2.4 Feature Selection Algorithms

Feature selection algorithms create a subset of a particular dataset that is comprised of attributes that best discriminate between classes [76]. This process can reduce the cost of classification by using fewer features and can lead to superior classification accuracy by discarding irrelevant features [76]. Three feature selection algorithms are implemented in this thesis: ReliefF [18–20], Support Vector Machine Attribute Evaluator (SVM AE) [21, 22, 34, 45], and Non-Correlated Aided Simulated Annealing Feature Selection–Integrated Distribution Function (NASAFS-IDF) [35, 77]. The first two are processed in Waikato Environment for Knowledge Analysis (WEKA) and NASAFS-IDF is processed in MATLAB®. WEKA is a machine learning program that runs on Java [34]. WEKA has numerous data mining tools for analysis and modeling. This thesis uses WEKA to discover discriminating features, and to build, train, and test classification models.

### *2.4.1 ReliefF.*

ReliefF [18–20] uses supervised learning to determine feature ranking based on an assigned feature weight value. The algorithm calculates a weight for a particular feature based on the distance between the nearest within-class sample and nearest out-of-class sample [20]. Relief is a two-class methodology, where ReliefF is extended to multiple classes; for example, if there are $C$ classes, Relief distinguishes Class A from all other classes, opposed to ReliefF, which distinguishes Class A from Class B from Class C, etc. [20]. The weighted output falls into the range of [-1,1], with 1 as the most favorable rank [20].

ReliefF randomly chooses one sample from the dataset and calculates the Euclidean distance between the chosen sample and the remaining samples [20]. This distance measurement determines which samples are labeled as a "hit" or "miss" [49]. A "hit" is considered a sample that is in the same class as the selected sample and also has a minimum Euclidean distance among samples in the same class [20]. A "miss" is a sample from a different class that has a minimum Euclidean distance among the samples of that class [20]. The weight vector for a specific randomly selected sample, $R$, is calculated as

$$W[A] = W[A] - \sum_{j=1}^{k} \frac{diff(A;R;H_j)}{mk} + \tag{2.9}$$

$$+ \left[ \sum_{C \neq class(R)} \frac{P(C)}{1 - P(class(R))} \right] * \sum_{j=1}^{k} \frac{diff(A;R;M_j)}{mk}, \tag{2.10}$$

where $W[A]$ is the current weight, $A$ is a feature vector, $m$ is the number of randomly selected samples, which is one in Eq. (2.10), $k$ is the user-defined number of nearest hits ($H$) or misses ($M$), $R$ is the selected sample, $P(C)$ is the probability of each class, $P(class(R))$ is the probability of the class of the sample selected, and the $diff(\cdot)$ function calculates differences between features [49].

### *2.4.2 Support Vector Machine Attribute Evaluator.*

The SVM algorithm [21, 22, 26, 28, 29, 34] is designed to locate the maximum separation between the two classes. This is accomplished by selecting samples from each class that become a support vector defining a maximum margin between the classes [21]. Fig. 2.14 shows a two-class problem of circles and squares where the shaded squares and shaded circle are support vectors for their class [21].

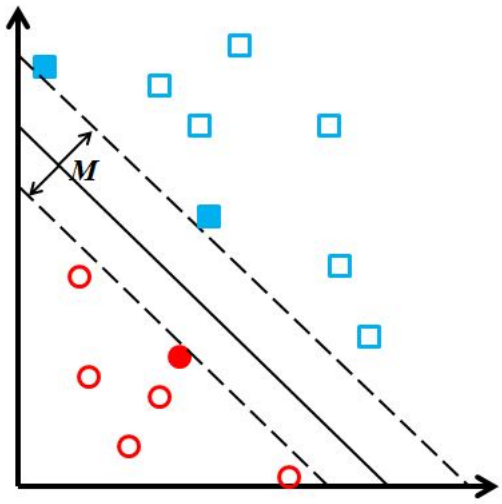



**Figure 2.14: The three data points that are shaded in are the support vectors for this data set. A support vector is a data point that exists on the very edge of the decision boundary margin, thus defining the width of the margin [21].**

Equation (2.11) represents the constraint of a support vector, where $y_i$ is the class label of the data point, either +1 or -1, $x_i$ the selected data point, $w$ is the normal vector to the hyperplane, indicated by $M$ in Fig. 2.14, and $b$ is a constant [21],

$$y_i(w \cdot x_i + b) = 1. \tag{2.11}$$

A point is a support vector if Eq. (2.11) is satisfied, whereas points lying within the margin width will have a value between 0 and 1, 0 being on the decision line and 1 being at the very edge of the margin [26]. A point at the very edge of the margin is a support vector.

The SVM AE feature selection algorithm is an extension of the SVM classifier. The SVM AE selects features that correspond to samples chosen as support vectors. Therefore, the features that correspond to the two shaded data points in Fig. 2.14 would be selected with the SVM AE as relevant features.

### *2.4.3 Non-Correlated Aided Simulated Annealing Feature Selection - Integrated Distribution Function.*

Non-Correlated Aided Simulated Annealing Feature Selection-Integrated Distribution Function (NASAFS-IDF) [35, 77] is a stochastic feature selection algorithm that implements simulated annealing to optimize a heuristic. This algorithm has been applied to hyperspectral data to select discriminating features for textiles [35]. The output of NASAFS-IDF is a feature set for each class containing a user-defined number of features [35]. The algorithm produces a feature set for each class because it uses a one-versus-all methodology [35]. Therefore, a class-specific feature set best distinguishes that class from the others. Because the datasets in this thesis only have two classes, either feature set produced by NASAFS-IDF should be able to be used to discriminate between the two classes.

There are three stages to NASAFS-IDF: selection, evaluation, and competition [35]. In the selection stage, the algorithm chooses a feature set at random from the available attributes of a sample. The heuristic calculation is accomplished in the evaluation stage and optimized by the simulated annealing method in the competition stage. The heuristic is calculated using a distance measure between classes and the covariance value of the selected feature sets [35]. This calculation determines if a given feature set is a good discriminator between classes [35]. A new feature set is determined by a random pick and replaces a feature of the previous set. This new feature set is sent to the heuristic to repeat the process as outlined above. The selection, evaluation, and competition process is repeated until convergence occurs, which is defined as meeting a minimum error

requirement [35]. Due to the stochastic nature of this process, a Monte Carlo method is employed to repeat the entire process [35]. The output is a histogram of features as selected by the prior process due to the Monte Carlo algorithm. The features among all sets are plotted as a histogram in order to evaluate the feature ranking [35]. Features that were chosen more often during the selection, evaluation, and competition process have a higher magnitude on the histogram [35]. These features are evaluated as superior discriminators and result in the final feature set. Another aspect of NASAFS-IDF is that it is programmed to choose highly discriminating features from across the dataset, rather than choosing features in close proximity to one another, as ReliefF and other common feature selection methods often do [35].

## 2.5 Classification Algorithms

For this thesis, three classification algorithms are implemented in WEKA: naive Bayes [23–25], SVM [21, 22, 26, 28, 29, 34, 45], and a decision tree [30–33]. The classifiers are applied to datasets of only the selected features from ReliefF, SVM AE, and NASAFS-IDF.

### *2.5.1 Naive Bayes Classifier.*

Naive Bayes classifier [23–25] is based on Bayesian theory that utilizes prior probabilities to distinguish between classes. This method bases its classification on the assumption of independence between features. Naive Bayes looks at each individual feature's contribution to the classification independent of the other features [23].

Naive Bayes creates a model based on the probability of data being in a particular class and the likelihood of future data being in that class. The general formula is

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}},$$

and

$$p(X|Y_1, \cdots, Y_n) = \frac{p(X)p(Y_1, ..., Y_n|X)}{p(Y_1, ..., Y_n)}. \tag{2.12}$$

The mathematical version of naive Bayes in Eq. (2.12) relies on a dependent class variable, $X$, which is conditional on a feature set, $Y_1, ..., Y_n$, of size $n$, and where $p(\cdot)$ represents the probability. The "naive" assumption of the algorithm means each feature ($Y_i$) is conditionally independent of every other feature ($Y_j$) for $i \neq j$, in class $X$ [24]. This assumption results in the following distribution [24],

$$p(X|Y_1, ..., Y_n) = \frac{1}{Z}p(X)\prod_{i=1}^{n} p(Y_i|X) \tag{2.13}$$

where $Z$ is a constant [49].

### *2.5.2 Support Vector Machine Classifier.*

The SVM classifier [21, 22, 26, 28, 29, 34, 45] implements supervised learning to distinguish patterns. SVM builds a model to classify samples into one of two classes [45]. The goal of SVM in a two-class problem is to achieve a maximized margin between classes [45].

With linearly separated data, there are several ways to classify samples. In Fig. 2.15, there are many different classifiers that would successfully differentiate between circles and squares, but only one of these is optimal: the one that maximizes the margin between the two classes [26]. Therefore, the SVM algorithm finds this optimal decision boundary using support vectors, as in Fig. 2.14 [26].

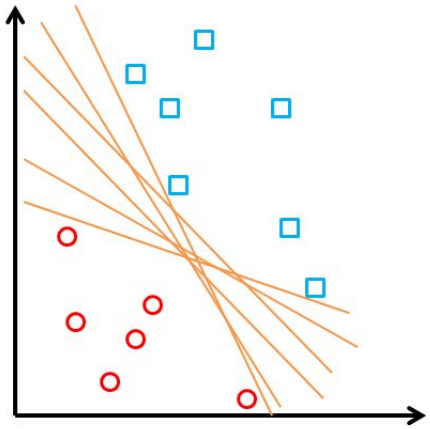


**Figure 2.15: For most linearly separable classes, there will be a number of different options for a decision boundary. The solid lines (red) show the possibilities of classifiers, but only one maximizes the distance between the two classes (circles, squares), which creates an optimal decision boundary. The SVM classifier maximizes the margin between the two classes [26].**

A linearly separable two-class case is the simplest example that shows how the SVM operates. For this case, $x \in R^n$, $y \in \pm 1$, where the classification equations are represented as follows:

$$w^T x_i + b > +1 \text{ for } d_i = +1,$$
$$w^T x_i + b \leq -1 \text{ for } d_i = -1,$$

therefore,

$$d_i(w^T x_i + b) \geq +1 \; \forall \; i, \tag{2.14}$$

where $x_i$ is an input sample, $w$ is a weight, which is normal to the decision line, $b$ is a bias, and $d_i$ is the desired output [45]. The decision boundary is any hyperplane that satisfies the constraint [45]

$$\mathbf{w^T x} - b = 0. \tag{2.15}$$

The line formed from Eq. (2.15) represents the center line in Fig. 2.14.

To achieve the widest margin between classes (the distance defined by $M$ in Fig. 2.14), the width from the hyperplane to each support vector is maximized [45]. This margin width, $M$, of the boundary is the maximum distance between the hyperplanes created by support

vectors for Class 1, $x^+$ and Class 2, $x^-$ [45]. The margin width is equal to

$$M = 2\sqrt{\mathbf{w}^T\mathbf{w}} \tag{2.16}$$

where **w** is a weight vector that is normal to the separating hyperplane [45].

Non-linearly separable patterns, as in Fig 2.16, require a kernel trick to achieve classification [27]. The kernel trick involves preprocessing the data by mapping the input data points, $(x_1, x_2)$,

$$K(x_1, x_2) \rightarrow (\Phi(x_1) \cdot \Phi(x_2)), \tag{2.17}$$

where $\Phi$ represents the kernel function [27].

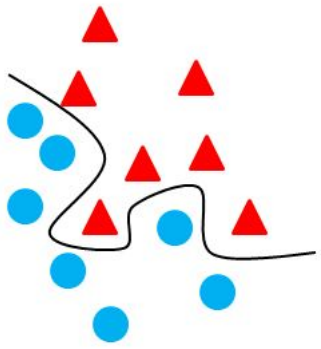


**Figure 2.16: The data collection is melded together such that the two classes, red triangles and blue circles, are not linearly separable. Therefore, different methods, such as the kernel trick and Lagrange multipliers are used to preprocess the data, allowing the SVM algorithm to accomplish separation [27].**

There are several different types of kernel functions. The default kernel in WEKA, which is implemented in this thesis, is the polynomial kernel. The PolyKernel is equal to

$$K(x_1, x_2) =< x, y >^p \tag{2.18}$$

where $< \cdot >$ represents the inner product and $p$ is an exponential, that has a default of 1.0. By choosing an optimal kernel function, the feature space becomes linearly separable, as in Fig. 2.17. From this point, the SVM algorithm can proceed.

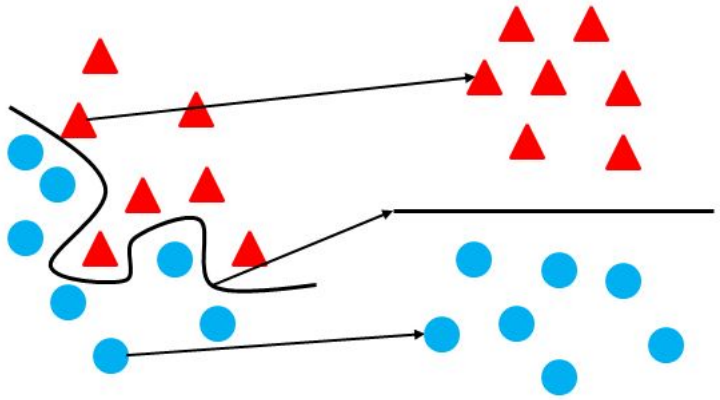


**Figure 2.17: Shows that non-linear preprocessing data can help transform the input space to a new feature space that has linearly separable data points.**

### *2.5.3 Decision Tree Classifier.*

Decision tree representation [30–33] uses a supervised learning method and has the ability to produce a binary output in terms of classification or regression. Decision trees are commonly employed due to their rules based methodology [30]. The algorithm is used to discern the class of a sample by stepping through decisions based on threshold values that are set based on the separation of the dataset. Figure 2.18 shows an example of a basic decision tree [30]. In the tree structure, the decision node ("root") represents a specific attribute, the branches represent the value of that attribute, and the leaves at the end of the tree assign the classification value [30]. The goal of the tree is to step through each attribute, moving down the branches to come closer to the leaf node that finalizes classification [30].

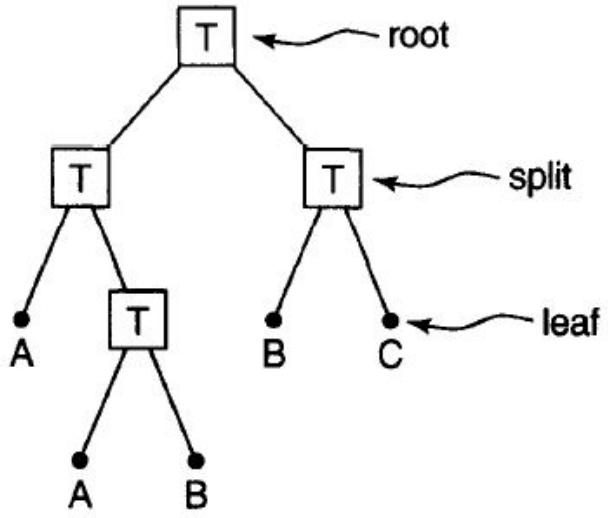



**Figure 2.18: This is an example of a basic decision tree. There are four decision nodes and five leaf nodes. The decision nodes pose the question to the attribute, thus calling for a decision to be made that leads to another decision node or a leaf node. The leaf nodes result in the classification of the sample [30].**

Similar to other classifiers, the decision tree algorithm accepts an input vector, which contains numerous different features [32]. The decision tree algorithm is most often developed as a top-down (greedy) search [33]. It starts at the first decision node and proceeds down through the different features until it reaches a leaf node [33].

Entropy is used to select the feature with the most information in regards to producing efficient results for decision trees [33]. Entropy is defined as a measure of the unpredictability of a variable and is graphed in Fig. 2.19 for a two-class problem [63]. Entropy calculates the amount of uncertainty in a set of outcomes from a random drawing [63]. If all samples belong to the same class, the entropy is 0 because there is no uncertainty of the outcome. Given a binary output $y \in \pm 1$, entropy falls between $[0, 1]$ and is equal to

$$H_n(p_1, p_2, ..., p_n) = -\sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \tag{2.19}$$

where $p_i$, $(i = 1, 2, ..., n)$ represents the probabilities and

$$\sum_{i=1}^{n} p_i = 1, \quad 0 < p_i < 1, \tag{2.20}$$

and $n$ is the number of samples [63].

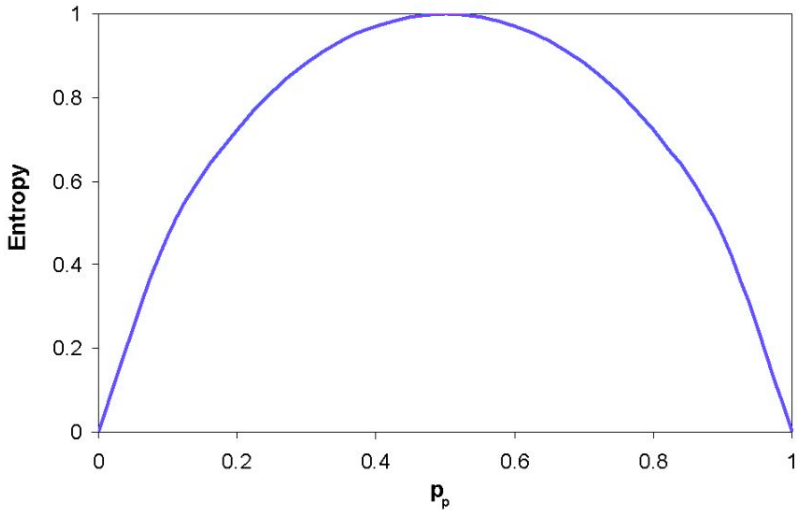



**Figure 2.19: Entropy is plotted in relation to the probability of a positive sample selected. As entropy increases, the variability of the sample decreases. When entropy is equal to one, there is an equal number of positive and negative samples [33].**

Entropy is used to determine information gain, which is a measure of the feature's effectiveness in classification [33]. Information gain is used to narrow down the feature selection process throughout the decision tree methodology, increasing efficiency [64]. The definition of information gain is

$$I(Y; X) = H(Y) - H(Y \mid X) \tag{2.21}$$

where $H(Y)$ is the entropy of $Y$ and $H(Y|X)$ is the conditional entropy of $Y$ given $X$, and where,

$$H(Y \mid X) = \sum_{\text{v: values of X}} P(X = v)\, H(Y \mid X = v) \tag{2.22}$$

where $Y$ is a label value, $X$ is a feature, and $v$ is an answer to a question about the feature [64].

## 2.6 Classification Training and Output

The classifiers are trained and tested with cross-validation in WEKA Explorer. Once a model is built, the "supplied test set" option in WEKA is implemented with the validation set. Cross-validation, using either five or ten folds, depending on the size of the dataset, separates the entire supplied train/test set into five or ten buckets and alternates training and testing across the different buckets. For example, if buckets one through four are used to train the dataset, the fifth bucket is used for testing; the next "fold" would train the dataset on buckets one through three and bucket five, and test on the fourth bucket. This continues until the training model is tested against every bucket.

Explorer outputs three important items following classification: numbers of correct and incorrect classification, percent accuracy, and a receiver operating characteristic (ROC) curve. The numbers of correct and incorrect classification form a confusion matrix, illustrated as a general example in Table 2.1. The confusion matrix provides more knowledge on how the classifier performs with a specific dataset. The false positive and false negative boxes indicate how many samples are misclassified. Ideally, the true positive and negative boxes contain the highest values because these represent correct classification. A ROC curve is determined in WEKA using data from confusion matrices. Data from one confusion matrix represents one point on a ROC curve. Each point is calculated and plotted as follows,

$$\frac{TP}{TP+FN} \text{ vs. } \frac{FP}{FP+TN}, \tag{2.23}$$

where TP means true positive, FN means false negative, FP means false positive, and TN means true negative [88]. The overall curve is built using predictions made by the classifier for each sample [34, 86, 87]. The predictions are sorted in descending order according

to the likelihood of the positive class [34, 86, 87]. WEKA generates the curve by going through the list and counting the number of TPs and FPs up to that point [34, 86, 87]. Therefore, the threshold value at each point on the curve is the probability of the positive class at that location in the list [34, 86, 87].

The percent classification accuracy may sometimes be an inaccurate quantifier if one class has more samples and is easier to distinguish than the other class. Table 2.2 gives an example. In this situation, a ROC curve is an appropriate evaluator of accuracy. The area under the curve (AUC) is calculated based off the ROC curve for an accuracy measurement.

**Table 2.1: Confusion Matrix Description.**

| Classified as: | Class1 | Class2 | Total |
|---|---|---|---|
| Class1 | True Positive | False Positive | # Class1 |
| Class2 | False Negative | True Negative | # Class2 |
| Total: | classified Class1 | classified Class2 | # correct |

**Table 2.2: Confusion matrix for naive Bayes classification results on Subject 5 validation dataset with features from NASAFS-IDF1. The overall accuracy for this model is 78.57%. For this case, the classifier misidentified 24 samples overall, but because the Non-Stress class is more than double the size of the Stress class, and all of the Non-Stress samples are correctly identified, the percent accuracy is skewed. The ROC returned an AUC value of 0.6830, which is a better indicator of the model's accuracy.**

| | Stress | Non-Stress | |
|---|---|---|---|
| S | 5 | 24 | 29 |
| NS | 0 | 83 | 83 |
| | 5 | 107 | 88 |
| Accuracy: 78.57% | | | |

## 2.7 Feature Generation

Feature generation [46, 47] processes features, creating new features. Features can be generated using statistical measures, e.g. mean, median, and mode, or a transformation, e.g. Fourier coefficients [49]. Once features are generated, they can be applied to feature selection and classification algorithms.

Variance is a statistical feature that can be generated from a dataset. It is the amount of spread in a set of numbers and is equal to the square of the standard deviation [48]. There are two different calculations of variance: biased and unbiased [48]. A biased variance estimator represents the variance of a sample mean and an unbiased variance estimator is the sample variance [48]. This thesis implements the default variance calculated in MATLAB®, which uses an unbiased estimator. Eq. (2.24) and (2.25) represent biased and unbiased variance, respectively [48],

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \tag{2.24}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \tag{2.25}$$

where $n$ is the number of samples, $x_i$ is a sample, and $\bar{x}$ is the sample mean. The sample mean is calculated as [48]:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{2.26}$$

## 2.8 Summary

HSI plays an important role in specific object recognition, where it is often difficult to detect targets among a cluttered background. For example, in a search-and-rescue scenario, the ocean or desert creates a difficult recovery environment in regards to target detection. This type of scenario proves to be very difficult due to vastness of the background. HSI is proving to be a common solution to this problem. HSI has been and is currently used for detection of specific targets in a cluttered background [43]. By implementing these techniques, an image highlighting the desired target is produced, making the objective more attainable. In aspects of stress detection, HSI could be used to assist pilots, air traffic controllers, deep-sea divers, and emergency medical personnel. To this extent, un-intrusive means of detecting stress is essential in such applications. With previous methods that

utilized contact probes, mobility and agility of body movement are limited for such tasks. By implementing non-contact means of stress detection, the dismount can continue their task with no interference or distraction, while oversight is provided, allowing actions to be taken to ensure mission success.

There are different ways to carry out stress detection: thermal imaging, PPG imaging, and HSI. HSI shows great potential due to its wide range of characteristics produced. Specific features of the skin that change at the onset of stress, particularly hemoglobin, are used to identify and classify stress. By implementing the feature selection algorithms, ReliefF, SVM AE, and NASAFS-IDF, discriminating wavelengths are extracted and applied to the datasets. These optimal datasets are processed through three classification algorithms: naive Bayes, SVM, and a decision tree. Feature generation is also applied by calculating the variance of each class. These new features are also processed through the feature selection and classification algorithms. All of these methods are examined in this thesis to determine their viability to produce accurate stress detection.

# III. Methodology

The evaluation of hyperspectral data as a practical means to detect stress is accomplished with several different techniques. Using a hyperspectral camera can eliminate the burden of intrusive equipment, which allows an individual to be unencumbered during data collection. This thesis uses hyperspectral data to detect stress by means of feature selection and classification algorithms.

Data collection, which is discussed in Section 3.1, details the experimental procedures and the data captured. Section 3.2 addresses the preprocessing of the data. Feature selection and classification algorithms applied to the data are discussed in Sections 3.3 and 3.4. Section 3.5 presents the results of the feature selection and classification algorithms.

## 3.1 Data Collection Stage

Hyperspectral data is collected at nominal and accelerated heart rate levels, producing a two-class problem. Section 3.1.1 explains how to characterize the emotional states of non-stress and stress. Sections 3.1.2 and 3.1.3 address the application of hyperspectral imaging (HSI) and electrocardiogram (ECG) recording. The procedures followed to accomplish data collection are discussed in Section 3.1.4.

### *3.1.1 Characterizing "Stress" and "Non-Stress".*

Data is collected on subjects under two different emotional states: stress and non-stress. These states are characterized based off heart rate (HR) and heart rate variability (HRV). HR is affected by age, general activity level, and breathing pattern [72]. Yuen et al. [4] noted that part of the body's physiological response to stress includes an elevated HR, though there is not a medically determined range of beats per minute (bpm) to characterize stress. Therefore, an increased HR in comparison to the baseline reading is one technique that is used in this thesis to characterize a state of increased stress. The optimum baseline

HR is collected when the subject is completely relaxed. Relaxation procedures are used to obtain a resting HR prior to a subject beginning the stress-inducing activity, as discussed in 3.1.4. HRV is proven to decrease as a subject's work load increases [66, 67, 71]. An ECG is attached to the torso to continuously record HR and HRV throughout the experiment. By comparing the baseline HR to the HR during the "stress" portion of the experiment, combined with comparing the overall HRV for each session allows the experimenters to validate whether a subject is experiencing stress.

### *3.1.2 Hyperspectral Data Collection.*

There are two available ways to collect hyperspectral data with an Analytic Spectral Devices (ASD) FieldSpec3® Pro spectroradiometer [12]: contact and non-contact (NC) fore optic. Both probes are used to record reflectance of the skin in the area above the carotid artery; the location of this artery is referenced in Fig. 3.1.

The carotid artery is a large artery on the side of the neck. The carotid is one of the larger arteries in the body, with a diameter of 6.10 ± 0.80mm in women and 6.52 ± 0.98mm in men [79]. Collecting data in the area of an artery is useful since arteries are the largest blood vessels and are responsible for transporting clean, oxygen-rich blood to the rest of the body [78]. This thesis considers the physiological changes that occur as a result of stress, specifically the change in the hemoglobin oxygen level (HOL), therefore, it is necessary to image oxygen-rich blood.

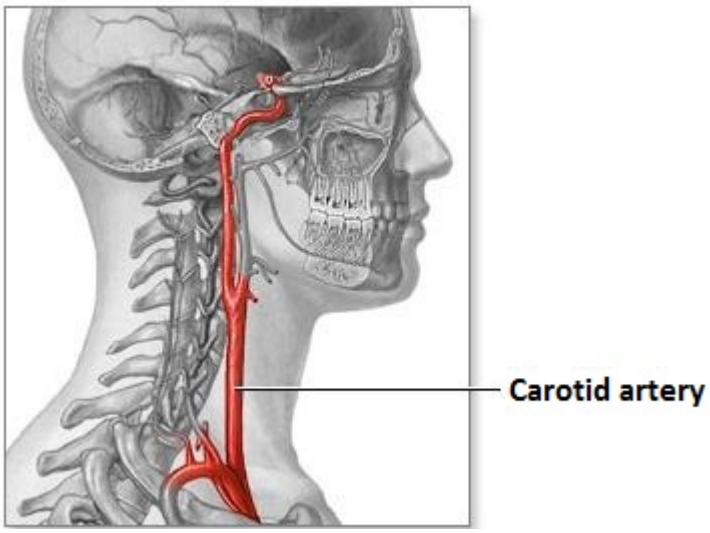



**Figure 3.1: This image of the carotid artery [73] is one of the major arteries in the human body. This location is a common place to measure pulse because the artery is near the skin surface and the side of the neck offers a wide, flat plane to place a sensor. The carotid, one of the larger arteries, is generally greater than 10mm in diameter, compared to smaller arteries, which range from 0.1-10mm [78].**

The contact and NC probes are both used to train the learning algorithms. The contact probe provides a clean, noiseless collection of the reflectance signature. The NC optic is necessary to provide the conditions of a real-world scenario by introducing atmospheric noise with a non-invasive stress detection technique. The NC data collected is used two different ways: as "real-world" validation of models trained with contact data and to train and build models on noisy data.

The contact probe contains an internal light source that illuminates the target area with visible through infrared energy. The contact probe is placed on the skin a total of four times per subject per data collect, with each collection lasting 20 seconds or less. To discover the exact positioning of the contact probe, the subject is asked to locate and identify the location of their carotid artery on the side of their neck, directly under their jaw line.

Collecting data with the NC optic creates the challenge of ensuring only the desired surface is recorded. This issue is mitigated by calculating a field-of-view (FOV) of the fore optic to include only the region-of-interest (ROI). The FOV is determined by the lens viewing angle of the sensor and the distance between the probe and the surface of the ROI.

Equations (3.1) and (3.2) calculate the FOV in squared inches, as:

$$r = h * \tan\left[\left(\frac{\alpha}{2}\right) * \frac{\pi}{180}\right] \tag{3.1}$$

$$FOV = \pi * r^2, \tag{3.2}$$

where $h$ is the distance between the probe lens and the surface of the ROI in inches and $\alpha$ is the viewing angle in degrees [12]. Figure 3.2 shows a visualization of the required variables.

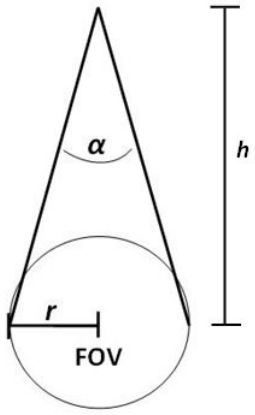



**Figure 3.2: This is an illustration of the variables that are used to calculate the FOV as described in Eq. (3.1) and (3.2). The FOV is a product of the radius, $r$, height, $h$, and viewing angle, $\alpha$.**

The ASD provides a pistol grip NC probe with a bare fiber optic cable that has a 25 degree viewing angle. For this thesis, a 1 degree viewing lens is attached to provide an approximate FOV radius of 0.5 inches at a 12-inch distance.

### *3.1.3 Electrocardiogram Data Collection.*

An ECG is a medical device used to monitor the heart's electrical activity [80]. Electrical signals travel through the heart, causing the muscle to expand and contract, and causing the heart to circulate blood throughout the body [80]. A 3-lead ECG is used to provide the basic information pertaining to the heart: the HR and HRV [80]. The leads corresponding to the ECG machine are placed across the torso, as in Fig. 3.3. The lead

locations are positioned to capture the electric current as it crosses the heart, one on the left shoulder, one on the right shoulder, and the last on the left side, directly below the lead on the left shoulder, and in line with the umbilicus.

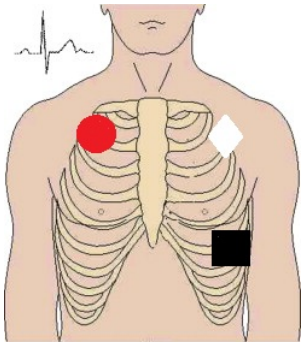


**Figure 3.3: The ECG implements a 3-lead configuration, as displayed here. The leads are positioned such that they can capture the electrical signal passing across the heart: white (diamond) on the upper left of the chest, red (circle) on the upper right, and black (square) on the lower right torso [74].**

Throughout the experiment, the ECG collects information on the heart. The ECG shows a continuous time waveform of the subject's heartbeat, as in Fig. 3.4 [80]. The HR and HRV values are processed and viewed after the collection. HR is output as a list of bpm with a timestamp and HRV is given as a value for each experimental session. An example of the HR output is in Appendix A.


**Figure 3.4: The ECG continuously records the HR waveform. The software then computes a HR that corresponds to each pulse, which is output as a list of bpm and the associated time of each recording [80].**

### *3.1.4 Data Collection.*

Six subjects are used for testing in this thesis. The experiment roster can be found in Appendix B. Skin reflectance curves are relatively the same shape, however, with a range of melanin concentrations, the reflectance amplitudes will vary at certain wavelengths, as shown in Fig. 2.9. The subject population consists of both male and female 23- to 26-years-old.

The subject is outfitted with the ECG leads and allowed to relax to a restful state. The ECG begins a constant recording and the ASD contact probe and NC fore optic are used to collect hyperspectral reflectance, recording non-stress and stress data.

The non-stressed state consists of a subject sitting in a chair, relaxed. To achieve the optimum state of relaxation, each test subject is provided calming images to visualize as they focus on slow, controlled breathing. The test subject applies the relaxing techniques for 5 minutes after the ECG leads are attached. While the subject is still focused on relaxing, the experimenters start recording data with the ECG and ASD contact and NC probes.

To bring about emotional stress, subjects engage in the interactive computer program, Air Force Mutli-Attribute Test Battery (AF_MATB) [5]. This program offers a method to introduce different levels of mental workload with varying task requirements. The software implemented in this thesis is based off the original MATB software, developed in 1992, which has become a foundation for psychological and psychophysiological research on cognitive workload [5].

AF_MATB operates on a standard laptop, where the test subject uses the keyboard and a USB joystick to perform certain tasks. The program has three pre-determined difficulty levels, low, medium, and high, to change the mental workload levels, but this thesis will only involve imaging during the high level. AF_MATB's viewing screen, as seen in Fig. 3.5, includes four tasks [5]. The tasks, which are the two leftmost windows and two middle windows, consist of System Monitoring, Resource Management, Communications, and

Tracking. System Monitoring, located in the upper-left corner, involves monitoring four gauges and two lights. Once a gauge or light experiences a malfunction, the user provides corrective action via the keyboard appropriately. The bottom-middle window is Resource Management. The goal of this task is to maintain and balance the fuel supply in two consumption tanks (Tank A and Tank B).

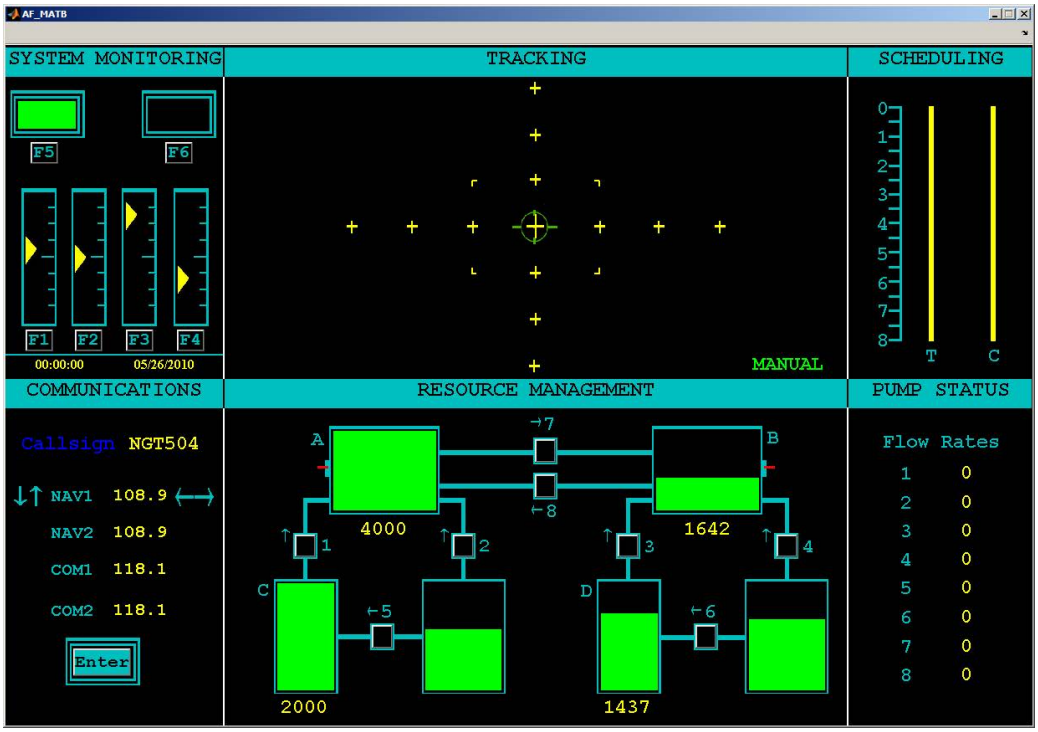



**Figure 3.5: The viewing screen of the AF_MATB computer software. The program consists of four tasks, which are represented by the two windows on the left and two middle windows. The windows from left-to-right top-to-bottom are System Monitoring, Tracking, Scheduling, Communications, Resource Management, and Pump Status [5].**

This is accomplished by opening and closing eight different pumps. To adjust these pumps, the user presses the corresponding number on the keyboard. The window directly to the right of the Resource Management window is the Pump Status window. Pump Status indicates the current flow rates of all the pumps in the Resource Management. This information can be referenced by the user to improve the overall performance of resource allocation. The Tracking task, located in the upper-middle window, requires the user to control the joystick to keep the unstable crosshairs within the rectangular box and as close

to the center crosshairs as possible. Lastly, Communications, in the bottom-left corner, involves listening for the appropriate call sign and then making changes based on what is heard. A radio call will come in first stating a call sign, which may or may not refer to the test subject, which is followed by directions to change to a certain frequency, which the test subject accomplishes with the up, down, left, and right arrows. The last window is Scheduling, which involves information for future developments in the areas of Tracking ("T" line) and Communications ("C" line). However, this control window is turned off for the experiment so that the user does not have any knowledge about the upcoming levels of difficulty. The combination of these controls running concurrently simulates tasks analogous to a flight crewmember.

Each subject is introduced to the AF_MATB program and allowed a ten-minute session to familiarize themselves with the software on a low workload setting. At this time, they are encouraged to ask questions about the operation of the software. After the training is completed and upon confirmation from the subject that they are comfortable with AF_MATB, the program officially begins. The low workload level is only used for training; the high level is used for testing. The subject accomplishes a five-minute session at the high workload level. The ECG is set to constantly monitor and record heartbeat during the entire experiment. While the subject accomplishes each level, the contact probe is applied to the skin above the carotid artery to collect HSI data. The subject then repeats the entire process so reflectance using the NC optic can be recorded. For more information and details on the experimental procedures, especially referencing the ECG collection and analysis, see Capt. Splawn's thesis [82].

### 3.2 Data Pre-processing

The data pre-processing consists of two steps, pre-processing accomplished with the ASD before and during data collection and the pre-processing accomplished in MATLAB®

post-collection. After data is recorded, the $RS^3$ software is used to convert raw radiance to reflectance data [12].

Prior to every data collect, two actions are accomplished to adjust the hyperspectral camera's sensitivity to light: optimization and white reference (WR) collection. Optimization is necessary to ensure the detectors do not saturate due to changing downwelling irradiance levels [75]. Downwelling irradiance is the diffuse and direct radiant energy emanating downwards [75]. A WR collection calibrates the spectroradiometer to register 100% reflection from surfaces that are nearly 100% reflectant [12]. The calibration is needed due to the differences in the light sources and their effects on the collected radiance values [12]. Taking an independent reading of the light source's illumination on a known reference material provides a means to attain relative the reflectance of the sample alone [12]. Such a material is required to have 95-99% reflectance across the entire spectrum and is called a WR panel or WR standard [12]. Spectralon from Labsphere is a type of WR standard that is characteristic of being nearly 100% reflective across the visible-to-near-infrared (VNIR) and shortwave infrared (SWIR) spectral ranges [75]. The material is made of polytetraflouroethylene and cintered halon [12].

There are three detectors within the ASD spectroradiometer: one VNIR and two SWIR detectors [12]. The VNIR (wavelengths 350-1000nm) detector converts received photons to electrons [12]. This electric current is continually converted to a voltage and digitized by a 16-big analog-to-digital (A/D) converter at regular intervals. The digitized data is transferred to the device controller for processing and analysis [12]. Unlike the VNIR spectrometer, which holds an array of 512 detectors and scans in parallel, the SWIR has two detectors, scanning from wavelengths 1000-1830nm (SWIR1) and 1830-2500nm (SWIR2) [12]. Thus, these detectors gather wavelength data sequentially. The SWIR detectors follow the same conversion path as a VNIR detector after the data is collected.

Another important difference between the detectors is the aspect of dark current (DC). DC is the amount of electrical current produced by electrons within the spectroradiometer when there are no other external photons present at the detector. This additional electrical signal is a type of additive noise that must be corrected. The two SWIR detectors automatically correct for dark current, but the VNIR detector requires a frequent DC measurement update, which is set in the software options.

One issue that emerges when using the NC optic is atmospheric noise due to water absorption bands [12]. These bands are located approximately between 1350-1400nm and 1810-1940nm [12]. The energy in these zones drops to zero, or nearly zero, in a typical outdoor setting [12]. Since the collection is held indoors, the resulting noise should not result in this degree of change assuming the humidity level is controlled to a low value. Therefore, the energy should remain constant with little resulting noise. One option is to discard the reflectance values between these two bands. A second option is to consider the noise negligible and continue analysis using all wavelengths. Choosing to discard certain groups of wavelengths would result in the loss of data that could prove important for analysis, so the low levels of noise generated are ignored.

The $RS^3$ software, used in conjunction with the ASD spectroradiometer, converts the raw output from the spectroradiometer recording to reflectance data in the form of text files that can be imported into MATLAB®. The ASD spectroradiometer collects the electromagnetic reflectance for wavelengths 350-2500nm with a sampling interval of 1nm, which equals 2,150 features, for each sample it records [12]. Figure 3.6 is an example of wavelength versus reflectance.

The ECG outputs a continuous HR waveform, as in Fig. 3.4, bpm, and HRV. The bpm are analyzed to determine whether significant changes in HR occurred during each experimental session and how the changes are correlated with the reflectance data collected from the spectroradiometer. An example of the output bpm is in Appendix A.

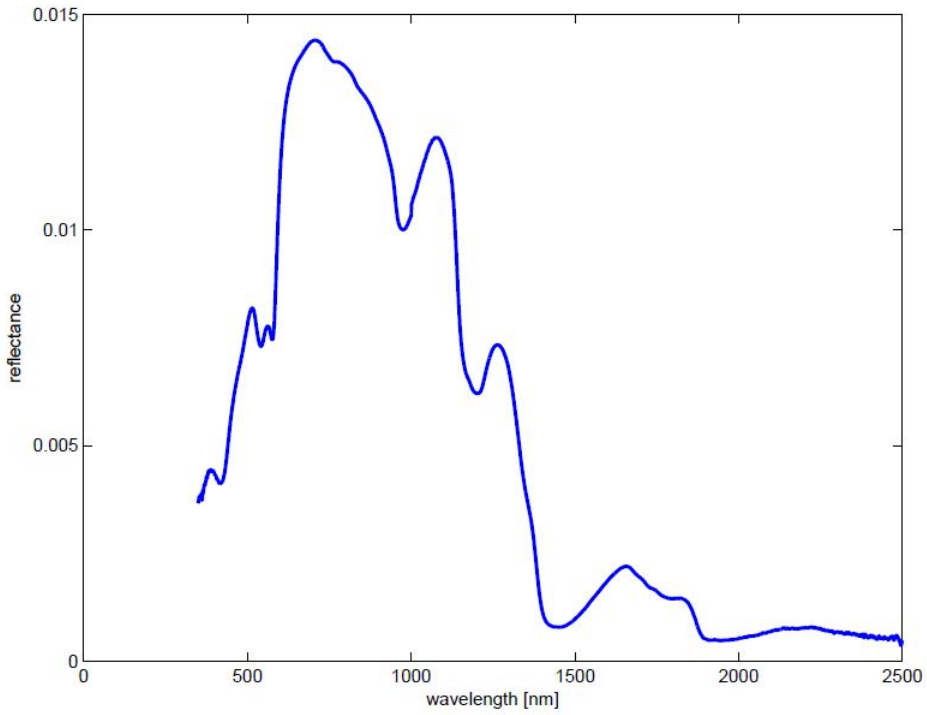



**Figure 3.6: The RS$^3$ software is used to process raw radiance data from the spectroradiometer into reflectance. This data is output as text files that can be imported into MATLAB®. Each text file is a sample, which consists of electromagnetic reflectance values for wavelengths 350-2500nm, sampled at 1nm [12]. The result is a signature spectral response.**

The HR is used to give a general idea of the level of stress experienced by the subject, but is not used to confirm that stress is induced. The HRV numbers are used to indicate that stress has occurred. HRV is a better indicator of stress than HR based on extensive research in this field [66, 67, 71]. Therefore, if a subject's HRV for a "stress" response is less than that of their baseline, this will be annotated as a state of "stress," otherwise the sample is discarded.

MATLAB® is implemented to normalize the data and change the format to comma-separated value (CSV). Normalization is necessary to ensure all the samples in the dataset are proportioned between [0,1] for consistency. The dataset, $S$, is normalized via the Euclidean distance with the equation,

$$S_{norm} = \frac{S}{||S||}, \tag{3.3}$$

where

$$\|S\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2},$$

for

$$\{x_i\}_{i=1}^{n},$$

where $x_i$ is a sample and $n$ is the total number of samples.

The contact probe is used to obtain the most accurate reflectance data. Since the contact probe is in contact with the skin, the atmospheric attenuation is negligible. The feature selection and classification algorithms are trained on both contact and NC data, and the NC data is also used for validating a "real-world" situation with models trained on contact data.

The normalized reflectance data files, CSV files, are imported into the data mining software Waikato Environment for Knowledge Analysis (WEKA). WEKA contains an assortment of different machine learning algorithms for data pre-processing, classification, regression, clustering, association rules, and visualization [34]. WEKA requires the data to be in a matrix format such that each row is a sample and each column a feature. The features are reflectances at the wavelengths 350-2500nm, thus each data matrix has 2,151 columns, where the last column is the class designator. A total of 16 datasets resulted from the data collection, where eight sets were determined from the contact probe (C) and eight sets determined from the NC probe. The datasets consist of:

- Subject 1-6: each subject's reflectance dataset, (C) and (NC),
- Combo: a combination of all subject's data, (C) and (NC),
- Var: the variance of all subjects, (C) and (NC),

The datasets listed above are broken down into training/testing and validation sets for both contact and NC collections for two different cases. Figures 3.7 and 3.8 outline the

process of the two cases used in this thesis. The figures annotate where feature selection and classification algorithms are introduced and applied to each dataset. In Fig. 3.7 (case I), two original datasets are used: contact and NC. The contact is used to build the model (training/testing with 5-fold cross-validation) and to validate with the holdout set of the data. The NC portion is also used to validate the models based on a "real-world" scenario. Figure 3.8 illustrates case II, the progression of the original NC dataset, which includes building a model with 10-fold cross-validation and validating with a holdout set.

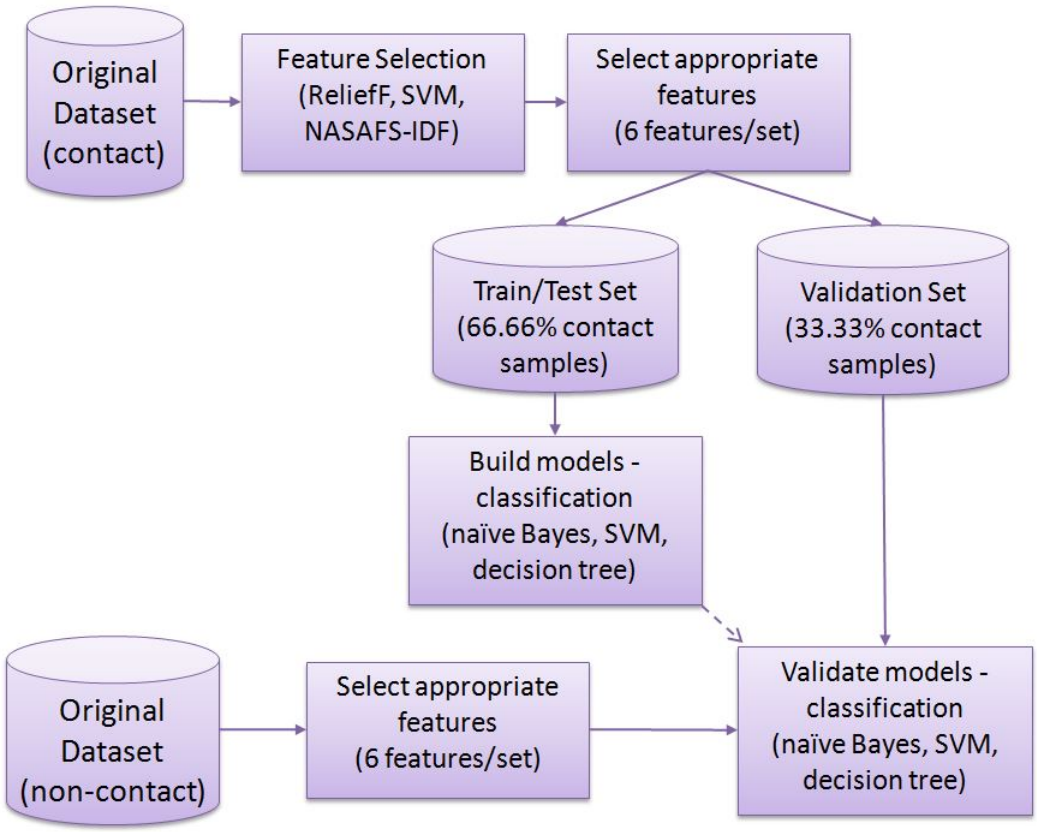



**Figure 3.7: This flowchart represents the progression of the contact test/train, validation, and "real-world" validation datasets. Models are trained, built, and tested with two-thirds of the data and validated with the remaining one-third of the contact data. Data collected with the NC probe is used as a "real-world" validation.**

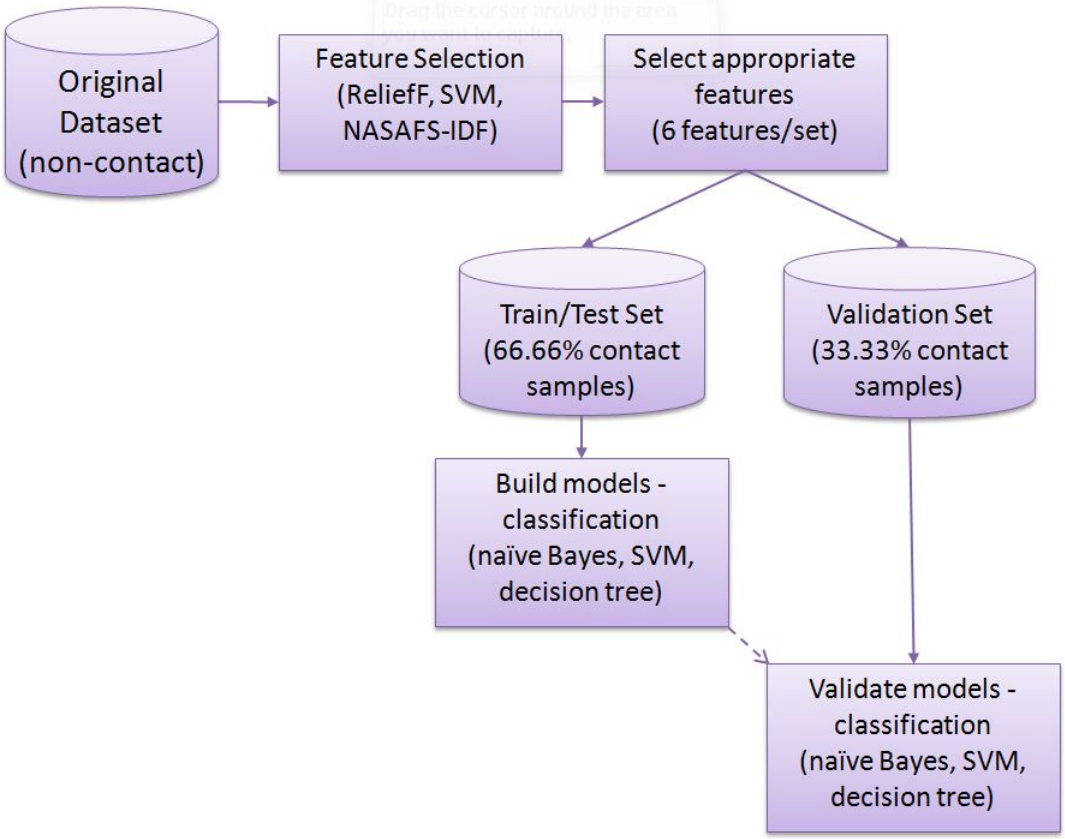



**Figure 3.8: This flowchart represents the progression of the NC test/train and validation datasets. Models are trained, built, and tested with two-thirds of the data and validated with the remaining one-third.**

For completeness, Fig. 3.9 displays the feature set selection/classification process for all available datasets. A model is created for each of the feature set/classification combinations in the figure (nine total). Figure. 3.10 further illustrates the breakdown of the datasets from the collected data. This equals 144 total models that are trained/tested and validated.

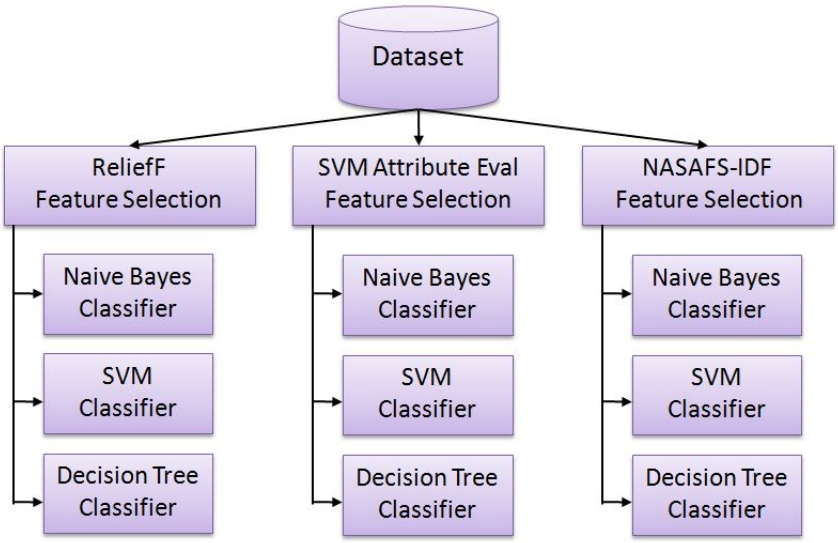



**Figure 3.9: This chart organizes and shows all combinations of feature selection and classification algorithms that are applied in this thesis. A model is built for each combination, both contact and NC datasets. There are a total of nine pairings of feature selection and classification algorithm per dataset.**

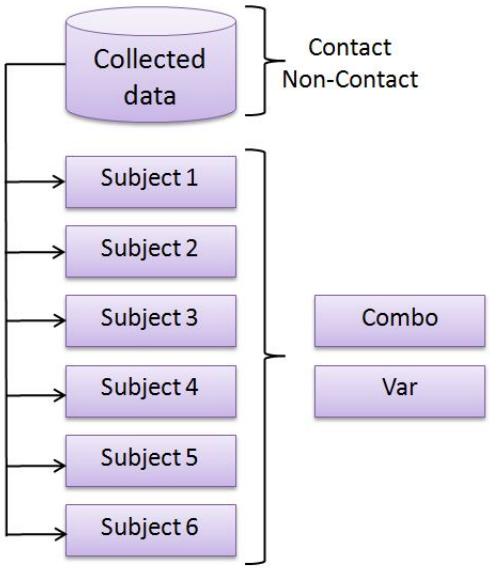



**Figure 3.10: This chart shows how the collected data is organized as specific datasets: Subject 1-6, Combo, and Var. Where the rectangular container is considered a different dataset that is individually applied to Fig. 3.9 for feature selection and classification. Each set consists of both contact and NC collections.**

## 3.3 Feature Selection Stage

Each hyperspectral sample contains 2,150 features; however, not all features are necessary to distinguish between the two classes. Feature selection algorithms are applied to the datasets to indicate the features that best separate the classes [76]. The feature selection algorithms implemented in this thesis are ReliefF [18–20], Support Vector Machine Attribute Evaluator (SVM AE) [21, 22, 34, 45], and Non-Correlated Aided Simulated Annealing Feature Selection–Integrated Distribution Function (NASAFS-IDF) [35, 77]. These algorithms are chosen in order to span the feature selection taxonomy presented by Blum and Langley, who delineate algorithms into three categories: filters, wrappers, or embedded [81]. ReliefF is identified as a filter method, SVM AE is in the embedded category [81], and NASAFS-IDF possesses a hybrid methodology [35, 81]. NASAFS-IDF utilizes simulated annealing, which is a type of wrapper according to [81], but the algorithm also contains a heuristic, which falls under filter methodologies [35, 81]. The algorithms are implemented on the datasets listed in Section 3.2. The combination datasets provide a global feature set to distinguish stress from non-stress. Feature selection is performed on data from both contact and NC collections.

ReliefF [18–20] and SVM AE [21, 22, 34, 45] are both implemented in WEKA and NASAFS-IDF [35, 77] is implemented in MATLAB®. ReliefF is a multi-class feature selection algorithm that calculates the distance between classes to determine feature rank [19]. In this thesis, a two-class methodology is used. The farther apart the two classes are for a particular feature, the higher the rank assigned to that feature. The issue that often happens with ReliefF is that the features with the highest ranks are often spatially located close to each other [19]. This may be acceptable, but often the result is a feature set containing similar features, which may be limiting. SVM AE is based off the support vector machine (SVM) classification algorithm. The SVM AE searches for features that contribute the most to the support vectors that produce the largest margin separating the two

classes [21]. Similarly to ReliefF, the features that result in the farthest distance between classes receive the higher rank. NASAFS-IDF is a feature selection algorithm that selects highly discriminating features that are non-redundant. This algorithm selects a group of features at random across the entire dataset and evaluates the group with a simulated annealing process that optimizes a heuristic. The output of NASAFS-IDF is a feature set per class based on a one-versus-all method for determining class separation [35]. NASAFS-IDF compares each individual class to all other classes and determines the best features to distinguish that class from the others. In this thesis, there are two classes, thus NASAFS-IDF produced two feature sets. The two feature sets are evaluated as Class1-versus-Class2 (feature set 1) and Class2-versus-Class1 (feature set 2). Since NASAFS-IDF is a stochastic process, the two feature sets are not equivalent, though they are similar. Either of the feature sets are appropriate to apply to the data for classification because each set is evaluated based on its ability to separate two classes.

In WEKA, both ReliefF and SVM AE are applied to the dataset and WEKA outputs features in rank order. The two feature selection methods are applied to each group of data: sets containing each individual subject's reflectance signature (labeled Subjects 1-6), sets of all subjects' reflectance (labeled Combo), a set of all subjects' reflectance variance (labeled Var), and the same sets, but with data collected using a NC probe (with "NC" attached to the label). Figures 3.11-3.13 show examples of spectral responses from each dataset. Each dataset contains the entire available spectrum of wavelengths, 350-2500nm. The top six features from ReliefF, SVM AE, and NASAFS-IDF are displayed in Table 3.1.

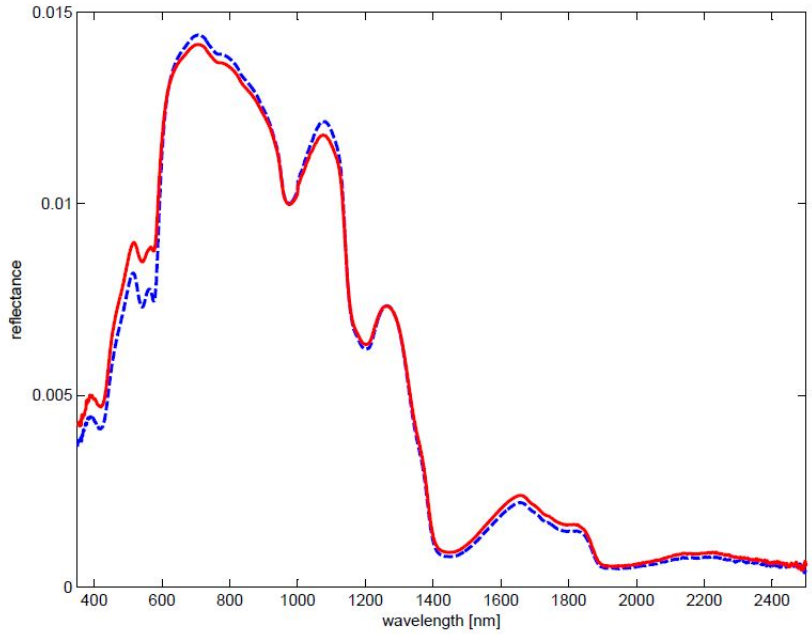



**Figure 3.11: Feature selection and classification algorithms are applied to three types of datasets. This is one sample from a subject's skin reflectance signature showing "stress" (red solid line) and "non-stress" (blue dashed line). There are six subjects, resulting in six datasets that process through feature selection and classification algorithms.**

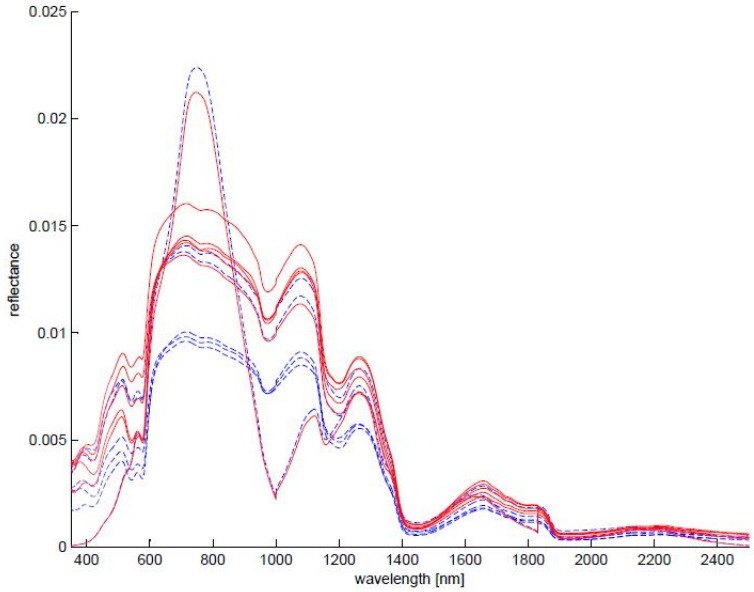



**Figure 3.12: Feature selection and classification algorithms are applied to three types of datasets. This shows the averaged combination of all six subject's reflectance response in "stress" (red solid line) and "non-stress" (blue dashed line). Though this shows the average, all samples from all subjects' reflectance results are processed with the feature selection and classification algorithms.**

**Table 3.1: Feature selection results for (a) Subject 1, (b) Combo, and (c) Var datasets. The remaining five subjects' feature selection results and NC feature selection results are located in Appendix C, Each dataset is collected using a contact probe and is processed through the feature selection algorithms ReliefF, SVM AE, and NASAFS-IDF to achieve a feature set of six features.**

(a) Subject 1 dataset

| | ReliefF | SVM | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| wavelength [nm] | 1211 | 1211 | 1205 | 585 |
| | 1210 | 1210 | 585 | 1335 |
| | 1212 | 1209 | 415 | 1635 |
| | 1209 | 1208 | 1625 | 415 |
| | 1215 | 1212 | 1945 | 385 |
| | 1214 | 1207 | 1385 | 1165 |

(b) Combo dataset

| | ReliefF | SVM | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| wavelength [nm] | 557 | 1534 | 575 | 565 |
| | 558 | 1531 | 1315 | 1315 |
| | 556 | 2496 | 1025 | 1145 |
| | 559 | 1521 | 1745 | 975 |
| | 560 | 1535 | 395 | 395 |
| | 555 | 2455 | 385 | 2495 |

(c) Variance dataset

| | ReliefF | SVM | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| wavelength [nm] | 2491 | 582 | 575 | 575 |
| | 2492 | 583 | 1105 | 1115 |
| | 2489 | 603 | 2495 | 925 |
| | 2490 | 581 | 1315 | 2495 |
| | 586 | 584 | 2485 | 405 |
| | 585 | 2489 | 1725 | 1585 |

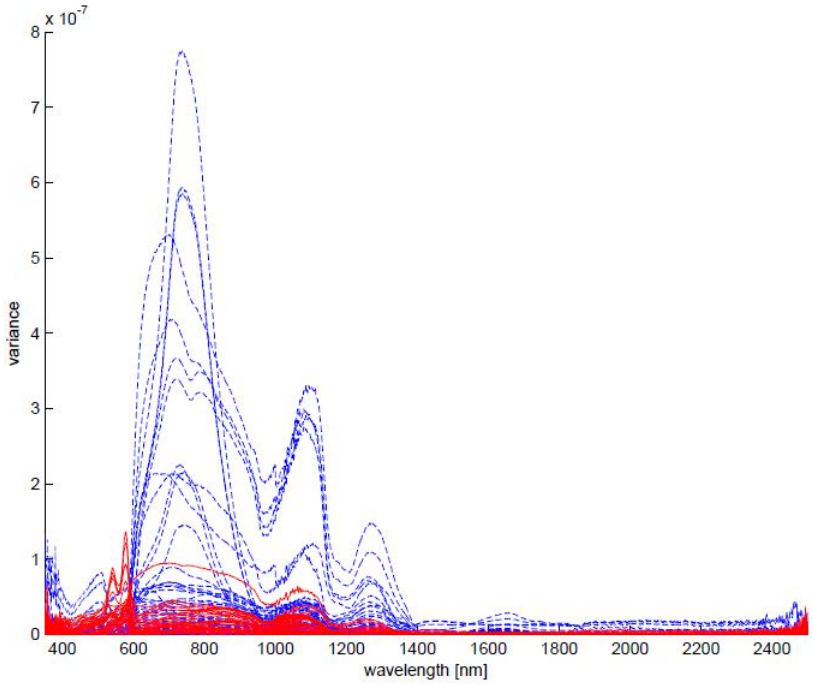



**Figure 3.13: Feature selection and classification algorithms are applied to three types of datasets. This displays the averaged variance of "stress" (solid red line) and "non-stress" (dashed blue line) for all subjects. Though this shows the average, all samples from all subjects' variance results are processed with the feature selection and classification algorithms.**

After the feature sets are sent through classification algorithms, the sets that return the highest accuracy and area under the curve (AUC) are noted as wavelengths of discrimination. One of the objectives in applying feature selection algorithms is to discover wavelengths that indicate universal distinction between the two classes, stress and non-stress.

## 3.4 Classification Stage

Features from ReliefF, SVM AE, and NASAFS-IDF are selected from the datasets listed in Section 3.2 and processed through the classification algorithms, naive Bayes [23–25], SVM [26, 28, 29, 45], and a decision tree [30–33]. Table 3.2 displays all dataset combinations used for classification. The holdout method is used on the contact dataset and is labeled: validation set. The train/test set consists of about 66.66% of the data and is used to build and test the models. The remaining 33.33% of the data is the validation set and is

used to evaluate the accuracy of the models built. Validation sets provide a way to evaluate classifier accuracy on new, previously unseen contact data. NC data is implemented for a second form of validation: "real-world" validation. Once models are validated, the "real-world" validation datasets are used to determine the generalization of the models for data containing atmospheric noise. The NC data is also used for training/testing purposes. Training/testing a classifier on data collected using a NC fore optic allows another option for modeling. The NC datasets contain many more samples than the contact datasets. This is because recording time is limited with a contact probe due to heat produced by the internal light source. The NC collection utilizes artificial lights, but the heat produced by these lights is negligible, thus the spectroradiometer can record for longer periods of time.

**Table 3.2: Datasets used for classification. Each dataset consists of a variety of samples collected with either a contact probe or a NC fore optic. Each dataset is comprised of six features selected from the feature selection algorithms ReliefF, SVM AE (SVM AE), and NASAFS-IDF (NAS1/2).**

| Dataset: | Sub1−6 | All Subjects | Var |
|---|---|---|---|
| ReliefF | $Sub_i$−RF | Combo−RF | Var−RF |
| SVM AE | $Sub_i$−SVM | Combo−SVM | Var−SVM |
| NASAFS-IDF1 | $Sub_i$−NAS1 | Combo−NAS1 | Var−NAS1 |
| NASAFS-IDF2 | $Sub_i$−NAS2 | Combo−NAS2 | Var−NAS2 |
| Dataset: | Sub1−6NC | All SubsNC | VarNC |
| ReliefF | $Sub_i$NC−RF | ComboNC−RF | VarNC−RF |
| SVM AE | $Sub_i$NC−SVM | ComboNC−SVM | VarNC−SVM |
| NASAFS-IDF1 | $Sub_i$NC−NAS1 | ComboNC−NAS1 | VarNC−NAS1 |
| NASAFS-IDF2 | $Sub_i$NC−NAS2 | ComboNC−NAS2 | VarNC−NAS2 |

## 3.5 Results

The features listed in Section 3.3 are selected and processed with the classification algorithms discussed. To create the models, the datasets, Subject 1-6, Combo, Var, Subject 1-6NC, ComboNC, and VarNC, are divided into four datasets according to each feature set, ReliefF, SVM, NASAFS-IDF1, and NASAFS-IDF2, as outlined in Table 3.2. It is pertinent to note here that one of Subject 3's experiment sessions returned HRV values that do not indicate stress; the baseline HRV is lower than the stress trial. Therefore, this

portion of Subject 3's dataset (non-contact data) is discarded. Each feature set is used with each of the three classifiers for training. An example of the results is shown in Table 3.3. This represents the classification accuracies for training models on the datasets Subject 1, Combo, and Var. Table 3.4 displays a sampling from different datasets of the confusion matrices calculated with the annotated feature selection and classification methods.

**Table 3.3: Percent accuracy on train/test sets Subject 1, Combo, and Var. The sets are comprised of two-thirds of the samples in each dataset. The sets include six features selected using the feature selection algorithms, ReliefF, SVM AE, and NASAFS-IDF and are evaluated using the classifiers, naive Bayes, SVM, and decision tree.**

| Subject1 | ReliefF | SVM | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| Naive Bayes | 100 | 100 | 100 | 100 |
| SVM | 100 | 100 | 100 | 100 |
| Decision Tree | 95.00 | 95.00 | 95.00 | 95.00 |
| Combo | | | | |
| Naive Bayes | 65.95 | 82.97 | 78.72 | 78.72 |
| SVM | 65.95 | 80.85 | 76.59 | 78.72 |
| Decision Tree | 80.85 | 90.42 | 88.29 | 89.36 |
| Var | | | | |
| Naive Bayes | 76.05 | 78.87 | 54.92 | 50.70 |
| SVM | 73.23 | 78.87 | 66.19 | 66.19 |
| Decision Tree | 71.83 | 73.23 | 76.05 | 80.28 |

**Table 3.4: Selected confusion matrices for classification on train/test sets. The sets include (a) Subject 1 with ReliefF features and a naive Bayes classifier, (b) Combo with SVM AE and a SVM classifier, (c) Var with NASAFS-IDF1 features and a decision tree classifier, and (d) ComboNC with NASAFS-IDF2 features and a naive Bayes classifier. The train/test sets are comprised of 66.66% of the contact data.**

(a) Classification train/test results on Subject 1 dataset using ReliefF features and naive Bayes classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 14 | 0 | 14 |
| S | 0 | 6 | 6 |
| | 14 | 6 | 20 |

(b) Classification train/test results on Combo dataset using SVM Attribute Evaluator features and SVM classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 46 | 12 | 58 |
| S | 4 | 32 | 36 |
| | 50 | 44 | 78 |

(c) Classification train/test results on Var dataset using NASAFS-IDF1 features and a decision tree classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 17 | 29 | 46 |
| S | 3 | 22 | 25 |
| | 20 | 51 | 39 |

(d) Classification train/test results on ComboNC dataset using NASAFS-IDF2 features and naive Bayes classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 99 | 0 | 99 |
| S | 11 | 290 | 311 |
| | 110 | 290 | 389 |

# IV. Results and Analysis

This chapter covers the results and analysis of the experimental procedures conducted in this thesis. The experiment is executed according to Section 3.1.4 with results from training and testing outlined in Section 3.5. Features are chosen from the feature selection algorithms ReliefF, Support Vector Machine Attribute Evaluator (SVM AE), and Non-Correlated Aided Simulated Annealing Feature Selection–Integrated Distribution Function (NASAFS-IDF). The classification algorithms, naive Bayes, support vector machine (SVM), and decision tree, are trained and tested on two-thirds of the samples in the datasets outlined in Table 3.2. These datasets come from the processed reflectance of hyperspectral imaging (HSI) using both the contact and non-contact (NC) probes. A contact probe is used for training because the data does not contain atmospheric noise that occurs with a NC fore optic. The NC data is implemented for training purposes to provide other potential models. The NC collection is also used for a "real-world" validation of models trained on contact data. The validation sets are comprised of numerous samples at six different wavelengths. Validation results from the contact data are discussed in Section 4.1 and "real-world" validation results using the NC collections in Section 4.1.1. Section 4.3 begins analysis of the results by comparing the different feature selection and classification algorithms' results.

## 4.1 Contact Data

Validation sets evaluate the ability of models to classify previously unseen data. While two-thirds of the samples in the datasets defined by Table 3.2 are used for training/testing classifiers, the remaining one-third are implemented for validation.

A visual representation of a dataset is seen in Fig. 4.1. This represents the normalized reflectance of the Combo contact validation set, which includes six features and 46 samples.

The samples are comprised of the "stress" (red circles) and "non-stress" (blue x's) response collected with a contact probe. The features in this specific validation set are from the ReliefF feature selection algorithm.

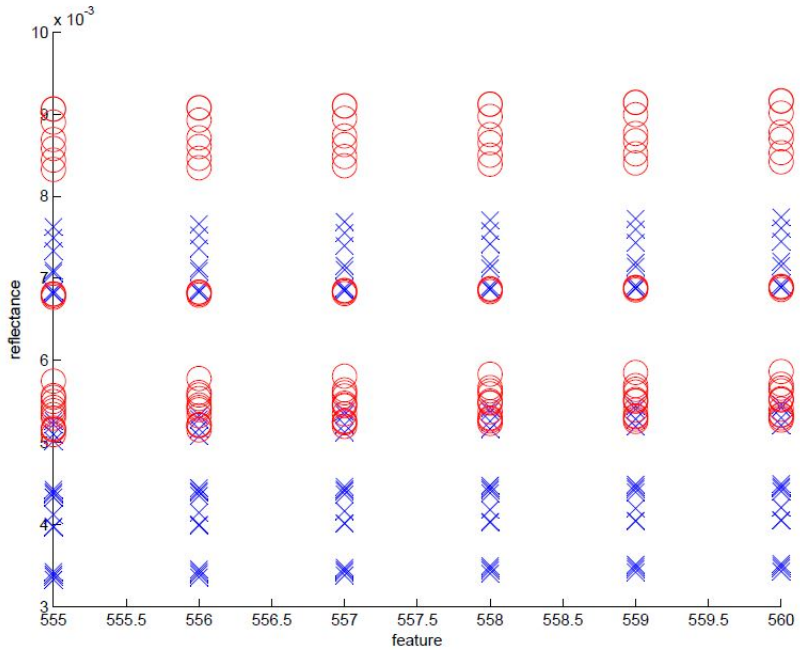



**Figure 4.1: This set represents six features and 46 samples from the Combo contact validation set, which includes the normalized skin reflectance of all subjects. The set consists of "stress" (red circles) and "non-stress" (blue x's) that denote each sample. These particular features are from the ReliefF feature selection algorithm.**

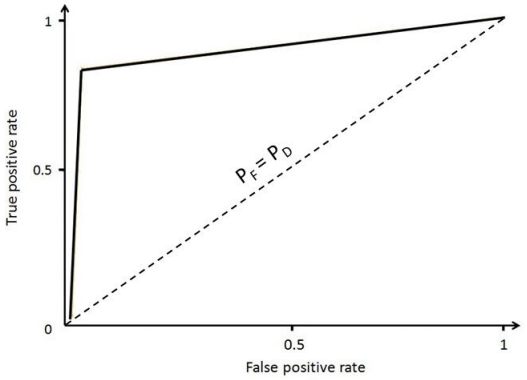
By examining the separation between classes from a particular contact feature set, wavelengths of importance are identified. For Subjects 1-6 datasets, there is sufficient class separation and consistency that led to high validation results. Because of this, there was not a particular feature set that indicated better class separation than another. Out of all the Combo and Var feature sets, the NASAFS-IDF sets provided features indicating the best classification. These wavelengths are noted in Table 4.1.

**Table 4.1: Wavelengths of maximum discrimination between classes for Combo and Var contact datasets from NASAFS-IDF2 feature sets.**

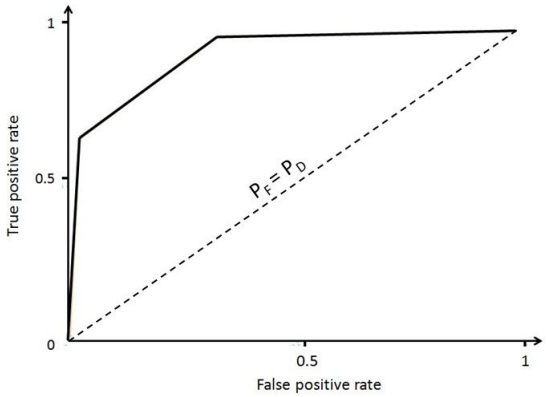
| | Combo | | Var |
|---|---|---|---|
| wavelength [nm] | 575 | 565 | 575 |
| | 1315 | 1315 | 1115 |
| | 1025 | 1145 | 925 |
| | 1745 | 975 | 2495 |
| | 395 | 395 | 405 |
| | 385 | 2495 | 1585 |

Literature indicates there is peak absorption of oxygenated hemoglobin in the 400, 550, and 575nm ranges and peaks in deoxygenated hemoglobin around 400 and 580nm [85]. These numbers do not directly reflect the wavelengths values discovered with the feature selection algorithms, but some of the numbers are located within a similar range. Also, the peak absorption of water is located at 1025nm, which was selected as a discriminating feature [83]. Literature does not indicate prominent wavelengths in the regions of 1100, 1300, 1500, and 1700nm, but these may provide additional opportunities for differentiation.

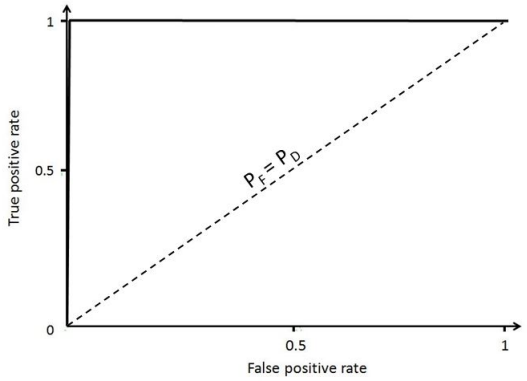
For each feature selection/classification algorithm pair, the accuracies for contact validation results on individual subjects are consistently 100% with a few exceptions, noted in Table 4.2. Notice that Subject 5 shows an accuracy of 83.33%, but an area under the curve (AUC) of 1.000. This resulted because the validation dataset is very small (six samples), so one misclassification leads to a misleading accuracy value. The corresponding receiver operating characteristic (ROC) curves are shown in Fig. 4.2. These curves are calculated in Waikato Environment for Knowledge Analysis (WEKA) and are associated with AUC values in Table 4.2.



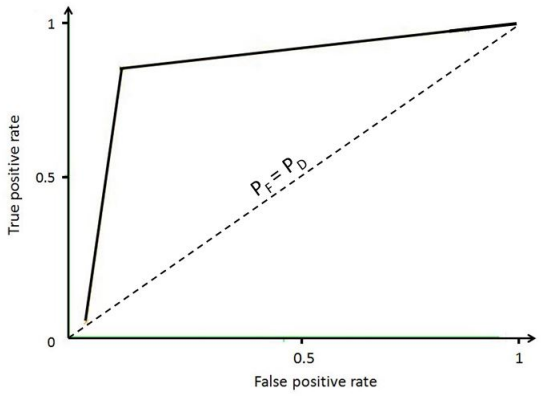

(a) Subject 1, ReliefF, Decision Tree




(b) Subject 3, NASAFS-IDF1, Decision Tree




(c) Subject 5, NASAFS-IDF2, Naive Bayes




(d) Subject 6, ReliefF, Decision Tree


(e) Subject 6, SVM, Decision Tree

**Figure 4.2: Selected ROC curve results on subject contact validation sets that correspond to accuracy and AUC in Table 4.2. (a) is Subject 1 with ReliefF features and a decision tree classifier; (b) is Subject 3 with NASAFS-IDF1 features and a decision tree classifier; (c) is Subject 5 with NASAFS-IDF2 features and a naive Bayes classifier; (d) is Subject 6 with ReliefF features and a decision tree classifier; and (e) is Subject 6 with SVM AE features and a decision tree classifier.**

**Table 4.2: Percent accuracy and AUC for Subject contact validation sets with less than 100% accuracy. The AUC is calculated as the area under the ROC curve. The associated ROC curves can be found in Fig. 4.2. The sets include six features selected using the feature selection (FS) algorithms, ReliefF (RF), SVM AE, and NASAFS-IDF (NAS1/2) and are evaluated using the classifiers, naive Bayes (NB), SVM, and a decision tree (DT).**

| Dataset | FS | Classifier | Accuracy | AUC |
|---|---|---|---|---|
| Subject 1 | RF | DT | 88.88 | 0.9167 |
| Subject 3 | NAS1 | DT | 83.33 | 0.9444 |
| Subject 5 | NAS2 | NB | 83.33 | 1.000 |
| Subject 6 | RF | DT | 88.88 | 0.9167 |
| Subject 6 | SVM AE | DT | 88.88 | 0.9167 |

On the average, the Combo contact validation set has a decreased accuracy in comparison to contact validation results on individual subjects. Across the combination sets, most accuracies are above 80%. The validation accuracies of datasets corresponding to those train/test datasets selected in Table 3.3 are shown in Table 4.3. To allow direct comparison between training/testing and validation, the accuracies and confusion matrices for the same four datasets listed in Table 4.3 are shown throughout this chapter. Table 4.4 displays the appropriate confusion matrices for the four selected datasets.

**Table 4.3: Classification results on selected validation sets that correspond to the selected sets in Table 3.3: (a) Subject 1 with ReliefF features and a naive Bayes classifier, (b) Combo with SVM AE and SVM classifier, (c) Var with NASAFS-IDF1 features and a decision tree classifier, and (d) ComboNC with NASAFS-IDF2 features and a naive Bayes classifier.**

| Dataset | Accuracy | AUC |
|---|---|---|
| (a) | 100.00 | 1.000 |
| (b) | 84.78 | 0.865 |
| (c) | 51.42 | 0.833 |
| (d) | 97.00 | 0.998 |

**Table 4.4: Selected confusion matrices for classification on validation sets that correspond to the selected sets in Table 3.3. The sets include (a) Subject 1 with ReliefF features and a naive Bayes classifier, (b) Combo with SVM AE and SVM classifier, (c) Var with NASAFS-IDF1 features and a decision tree classifier, and (d) ComboNC with NASAFS-IDF2 features and a naive Bayes classifier. The test sets are comprised of 33.33% of the contact data. Each set has different numbers of samples and six features.**

(a) Classification contact validation results on Subject 1 dataset using ReliefF features and naive Bayes classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 6 | 0 | 6 |
| S | 0 | 3 | 3 |
| | 6 | 3 | 9 |

(b) Classification contact validation results on Combo dataset using SVM Attribute Evaluator features and SVM classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 22 | 6 | 28 |
| S | 1 | 17 | 18 |
| | 23 | 23 | 38 |

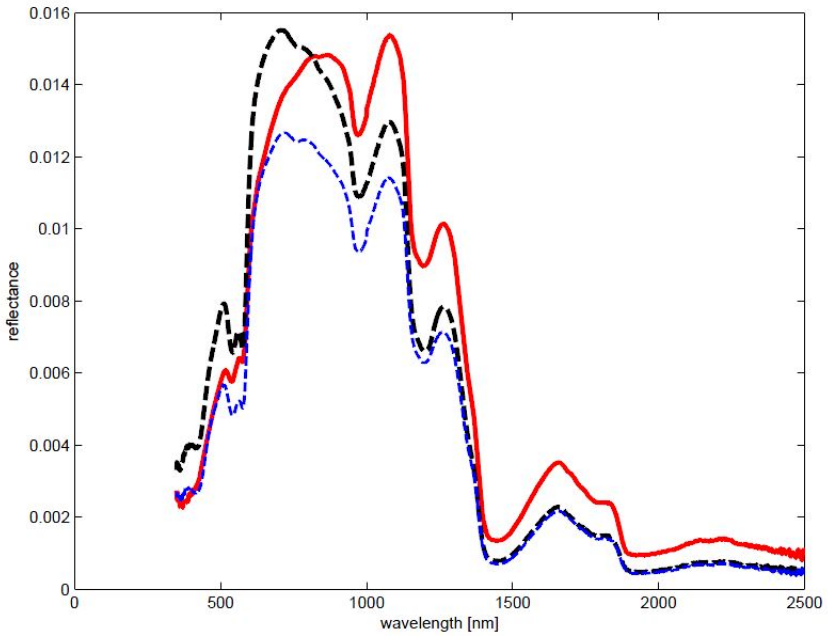
(c) Classification contact validation results on Var dataset using NASAFS-IDF1 features and a decision tree classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 7 | 16 | 23 |
| S | 1 | 11 | 12 |
| | 8 | 27 | 18 |

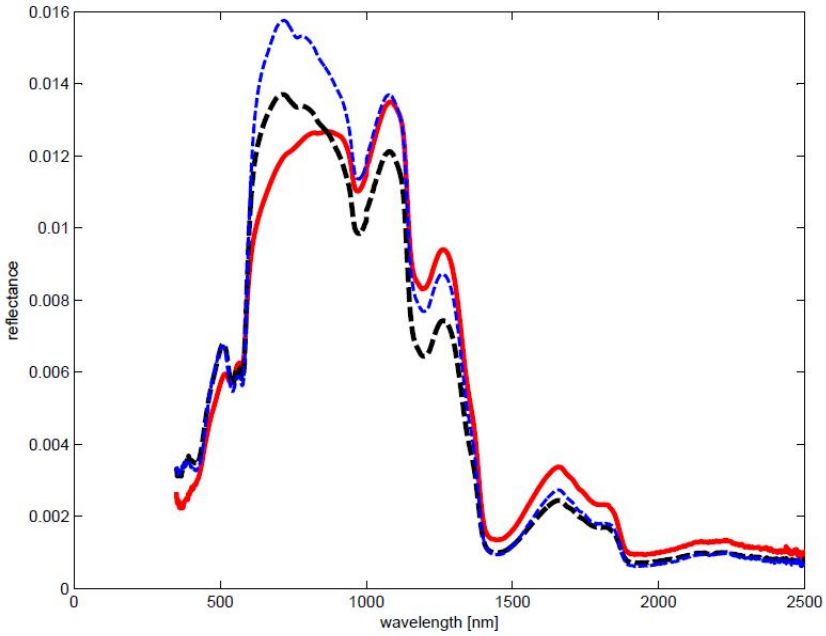
(d) Classification NC validation results on ComboNC dataset using NASAFS-IDF2 features and naive Bayes classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 49 | 0 | 49 |
| S | 6 | 145 | 151 |
| | 55 | 145 | 194 |

On average, the accuracy of individual subject's contact validation results across the different feature sets and classifiers is higher than that of the Combo contact validation. This is because the amplitude of the reflectance of skin varies between subjects. Figure 4.3 shows three averaged "stress" and "non-stress" skin signatures for three different subjects. These figures show that even the "stress" and "non-stress" spectral responses differ across the subjects, which makes group classification difficult. For completeness, group classification is still performed. The combination set is plotted in Fig. 3.12. This set shows the average "stress" and average "non-stress" signatures for Subjects 1-6; however, for classification, all individual samples are used.




(a) Three different subject's stress skin signatures.




(b) Three different subject's non-stress skin signatures.

**Figure 4.3: Three different subject's spectral responses to show that reflectance of both stress and non-stress has inconsistent amplitude. (a) shows the stress skin signature of three different subjects (Subject 1 solid red, Subject 2 dashed black, Subject 3 dotted blue) and (b) shows the non-stress skin signature of the same three subjects. Because the amplitudes vary, group classification is difficult and the most accurate results occur when detecting stress on an individual basis.**

The accuracy, confusion matrix, and ROC curve for the top performing Combo contact validation feature set and classifier pair are in Table 4.5 and Fig. 4.4. The results include 95.65% accuracy with an AUC of 0.9620 for the NASAFS-IDF2 feature set and a decision tree classifier. The second-best performing feature selection algorithm and classifier combination is NASAFS-IDF1 features with a decision tree classifier. This returned an accuracy of 95.65% also, but the AUC is 0.9610.

**Table 4.5: Accuracy, AUC, and confusion matrix for the top performing feature set and classifier on the Combo contact validation set: NASAFS-IDF2 features and a decision tree classifier.**

| | Stress | Non-stress | |
|---|---|---|---|
| S | 27 | 1 | 28 |
| NS | 1 | 17 | 18 |
| | 28 | 18 | 44 |
| Accuracy: 95.65% | | | |
| AUC: 0.962 | | | |

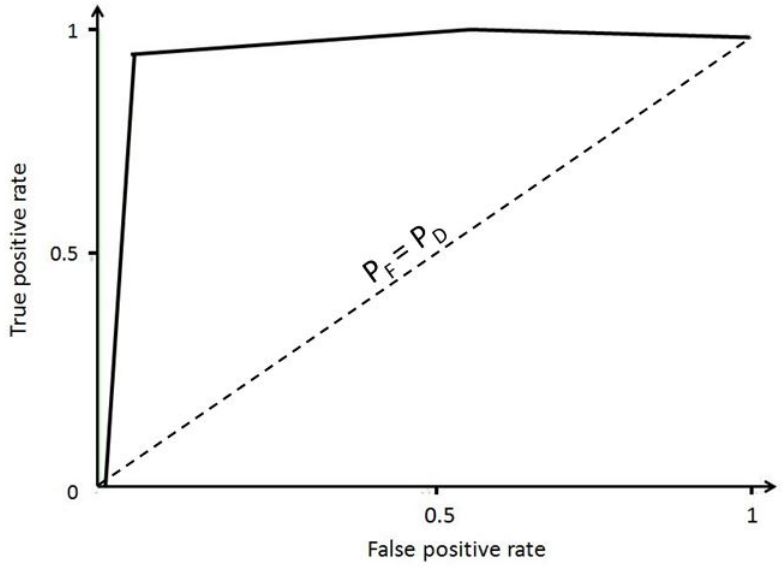



**Figure 4.4: ROC curve from the top performing contact validation feature selection and classifier pair on the Combo contact validation set: NASAFS-IDF2 features and a decision tree classifier. The Combo dataset is validated with one-third of the contact data used to build a model. The set includes all subjects' normalized reflectance, comprising 46 samples of "stress" and "non-stress" and six features.**

Normalized reflectance data is processed in MATLAB® to obtain the variance of skin reflectance under "stress" and "non-stress" conditions. All subjects' variances are combined into one dataset for evaluation. Figure 4.5 shows the variance of all subjects' data collected using a contact probe. On average, the variance of "stress" (red solid lines) tends to be lower than that of "non-stress" (blue dashed lines) because as heart rate (HR) increases, heart rate variability (HRV) decreases [66, 71]. Like the normalized reflectance data, two-thirds of this data is separated and used for training/testing, with the remaining one-third saved for validation. The variance data is collected with both the contact and NC probe.

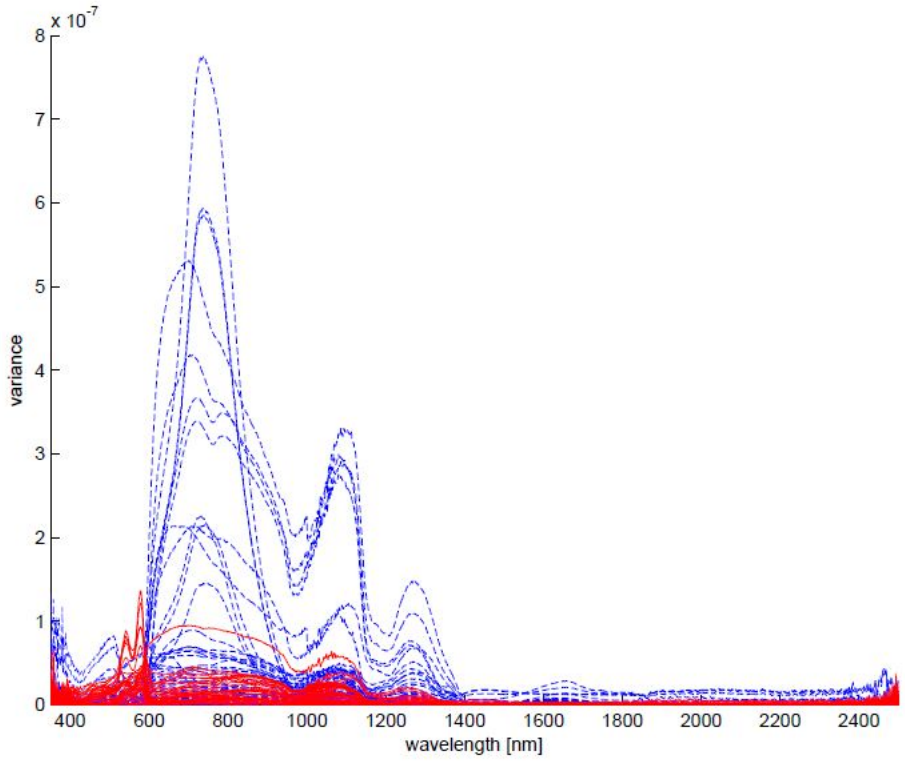



**Figure 4.5: All subjects' variances of reflectance collected with a contact probe. The variance of "stress" (red solid lines) is lower on average than "non-stress" (blue dashed lines) because the HRV decreases as stress increases [66, 71]. Two-thirds of these samples are used for training models and the remaining one-third is used for testing the models built.**

Table 4.6 shows the results from the top performing feature set (NASAFS-IDF2) and classifier (decision tree) on the Var contact validation sets. Figure 4.6 shows the associated ROC curve. Similar results were obtained using ReliefF features with a decision tree and SVM AE features with naive Bayes. These pairs both returned 85.71% accuracy and an AUC of 0.9350.

**Table 4.6: Accuracy, AUC, and confusion matrix for the top performing feature set and classifier on the Var contact validation set. The validation set is classified with a decision tree and is comprised of 35 samples of "stress" and "non-stress" and six features from NASAFS-IDF2.**

| | Stress | Non-stress | |
|---|---|---|---|
| S | 22 | 1 | 23 |
| NS | 0 | 12 | 12 |
| | 22 | 13 | 34 |
| Accuracy: 97.14% | | | |
| AUC: 0.996 | | | |

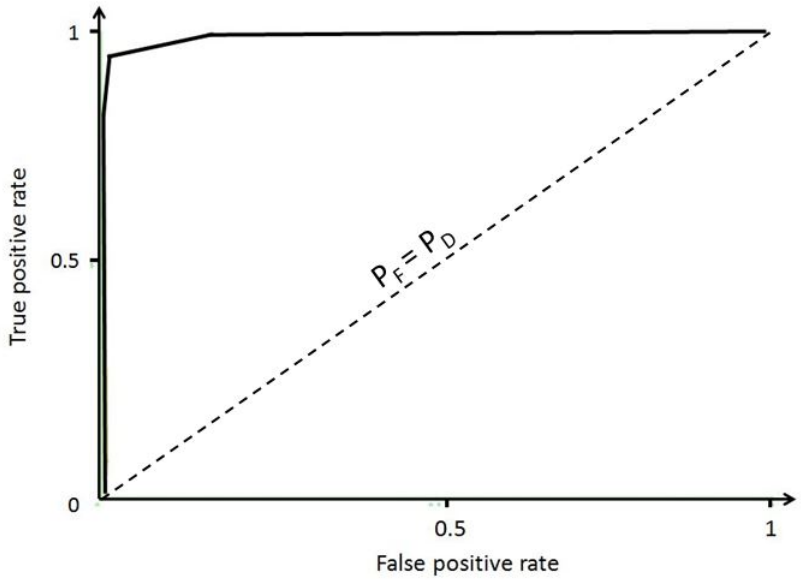



**Figure 4.6: ROC curve from the top performing feature selection algorithm (NASAFS-IDF2) and classifier (decision tree) on the Var contact validation set. The Var dataset is validated with one-third of the contact data. The set includes all subjects' normalized reflectance variance, comprising 35 samples of "stress" and "non-stress" and six features from NASAFS-IDF2 feature set.**

### *4.1.1 Validation on Contact Models with "Real-World" Non-Contact Data.*

The NC validation results provide a means of judging the applicability and accuracy of the models using real-world data. Validation datasets are comprised of reflectance recorded with the NC fore optic, which introduces atmospheric noise and is less consistent. All sets are evaluated using the models built from the naive Bayes, SVM, and decision tree classifiers. Each set consists of six features selected from ReliefF, SVM AE, and NASAFS-IDF.

When using a NC optic, the potential for inaccurate readings increases. The positioning of the probe relies heavily on the subject's posture while performing tasks. Since the data collection occurs while the subject performs the computer program, Air Force Multi-Attribute Test Battery (AF_MATB), movements could cause the camera's field-of-view (FOV) to deviate from the target location. The target location, about an inch diameter of the skin in the area of the carotid artery, can be held consistent with a contact probe, but is difficult to ensure exactness with the NC fore optic.

**Table 4.7: Percent accuracy on Subject 1, Combo, and Var NC validation datasets The sets are comprised of previously unseen data collected using a NC fore optic on six different subjects. The sets include six features selected using the feature selection algorithms, ReliefF, SVM AE, and NASAFS-IDF and are evaluated using the classifiers, naive Bayes, SVM, and decision tree.**

| Subject 1 | ReliefF | SVM | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| Naive Bayes | 20.68 | 20.68 | 20.68 | 20.68 |
| SVM | 20.68 | 20.68 | 20.68 | 20.68 |
| Decision Tree | 20.68 | 20.68 | 20.68 | 20.68 |
| Combo | | | | |
| Naive Bayes | 24.66 | 24.66 | 24.66 | 24.50 |
| SVM | 24.66 | 24.66 | 24.66 | 24.66 |
| Decision Tree | 21.66 | 24.66 | 43.00 | 24.66 |
| Var | | | | |
| Naive Bayes | 65.09 | 34.52 | 36.30 | 41.50 |
| SVM | 50.75 | 32.07 | 20.56 | 21.50 |
| Decision Tree | 69.24 | 27.35 | 21.69 | 22.64 |

The "real-world" NC validation results decrease significantly in accuracy compared to the contact validation results, as observed in Table 4.7. One reason for this is the additional

noise created using a NC probe, as well as numerous situational opportunities that are introduced with a NC recording. As discussed, excess movement from the subject causes changes in reflectance and, since the "stress" and "non-stress" collections are not recorded at the same time and the light source is turned off in between collections, the lighting could be inconsistent between collects.

Table 4.8 shows the confusion matrices of "real-world" validation sets corresponding to the datasets selected in Table 4.3 (Subject 1, Combo, Var, ComboNC), which displays results from contact validation. The same datasets chosen for training/testing in Table 3.4 and contact validation in Table 4.4 are chosen to allow comparison between training/testing, contact validation, and "real-world" NC validation results.

**Table 4.8: Selected confusion matrices for classification on "real-world" validation sets that correspond to Tables 3.4 and 4.4 to allow direct comparison. The sets include (a) Subject 1 with ReliefF features and a naive Bayes classifier, (b) Combo with SVM AE, (c) Var with NASAFS-IDF1 features and a decision tree classifier, and (d) ComboNC with NASAFS-IDF2 features and a naive Bayes classifier. The validation sets are comprised of data collected with a NC fore optic. Each set has different numbers of samples and six features.**

(a) ReliefF features, naive Bayes classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 30 | 0 | 30 |
| S | 115 | 0 | 115 |
| | 145 | 0 | 30 |

(b) SVM Attribute Evaluator features, SVM classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 148 | 0 | 148 |
| S | 452 | 0 | 452 |
| | 600 | 0 | 148 |

(c) NASAFS-IDF1 features, decision tree classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 108 | 5 | 112 |
| S | 410 | 7 | 417 |
| | 518 | 12 | 115 |

(d) NASAFS-IDF2 features, naive Bayes classifier.

| | Non-stress | Stress | |
|---|---|---|---|
| NS | 49 | 0 | 49 |
| S | 6 | 145 | 151 |
| | 55 | 145 | 194 |

Table 4.9 contains the confusion matrices for the top results in each category of the "real-world" validation sets (Subject 1-6NC, ComboNC, VarNC). These sets are from models trained on contact data that are validated again with NC data. The first validation was accomplished using the holdout method. NASAFS-IDF2 features provided the highest

NC validation accuracy on the normalized reflectance data (Subject 1-6NC and ComboNC), while the ReliefF feature set had the highest NC validation accuracy on the VarNC set. The decision tree classifier was most successful for both the ComboNC and VarNC NC validation sets, while naive Bayes returned the highest accuracy on the SubjectNC NC validation set. Figures 4.7-4.9 display the corresponding ROC curves.

**Table 4.9: Confusion matrix for the top performing feature sets and classifiers on (a) Subject 5NC, (b) ComboNC, and (c) VarNC "real-world" NC validation set. The models are trained/tested on contact data, validated using the holdout method with contact data, and then validated again, with results shown here, using NC data.**

(a) Confusion matrix for Subject 5NC "real-world" validation set of the top performing feature set (NASAFS-IDF1) and classifier (naive Bayes).

| | Stress | Non-stress | |
|---|---|---|---|
| S | 5 | 24 | 29 |
| NS | 0 | 83 | 83 |
| | 5 | 107 | 88 |
| Accuracy: 78.57% | | | |
| AUC: 0.683 | | | |

(b) Confusion matrix for ComboNC "real-world" validation set of the top performing feature set (NASAFS-IDF1) and classifier (decision tree).

| | Stress | Non-stress | |
|---|---|---|---|
| S | 119 | 29 | 148 |
| S | 313 | 139 | 452 |
| | 432 | 168 | 258 |
| Accuracy: 43.00% | | | |
| AUC: 0.556 | | | |

(c) Confusion matrix for VarNC "real-world" validation set of the top performing feature set (ReliefF) and classifier (decision tree).

| | Stress | Non-stress | |
|---|---|---|---|
| S | 19 | 94 | 113 |
| NS | 69 | 348 | 417 |
| | 88 | 442 | 367 |
| Accuracy: 69.24% | | | |
| AUC: 0.452 | | | |

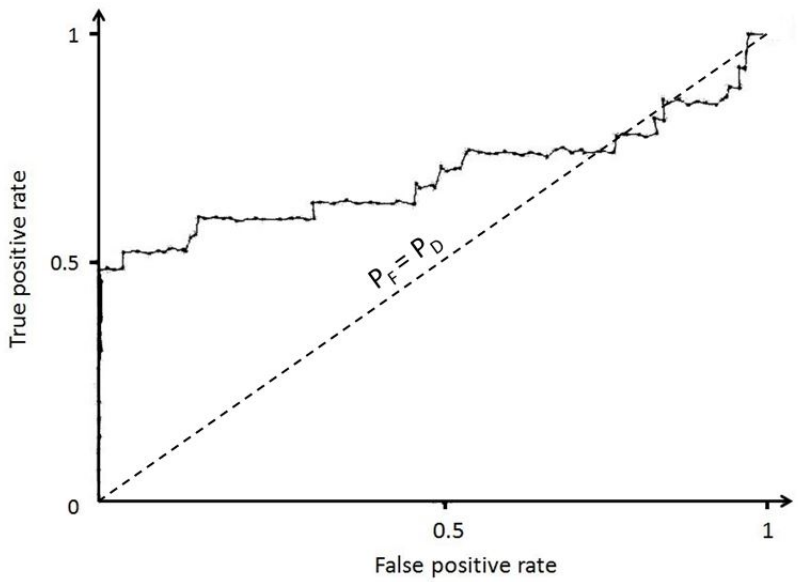



**Figure 4.7: ROC curve for the most successful validation result (Subject 5) among the subjects' validation sets, which results from NASAFS-IDF1 features and a naive Bayes classifier. Subject 5 dataset is validated with data collected with a NC fore optic. The set is comprised of 600 samples of "stress" and "non-stress" and six features.**

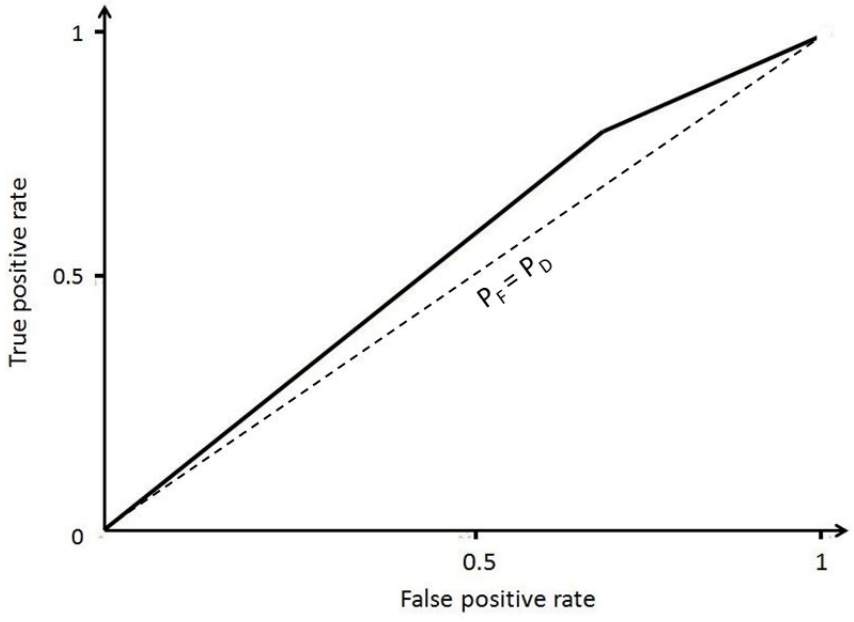



**Figure 4.8: ROC curve for the most successful validation result among the Combo validation sets, which results from NASAFS-IDF1 features and a decision tree classifier. The Combo dataset is validated with data collected with a NC fore optic. The set is comprised of 112 samples of "stress" and "non-stress" and six features.**

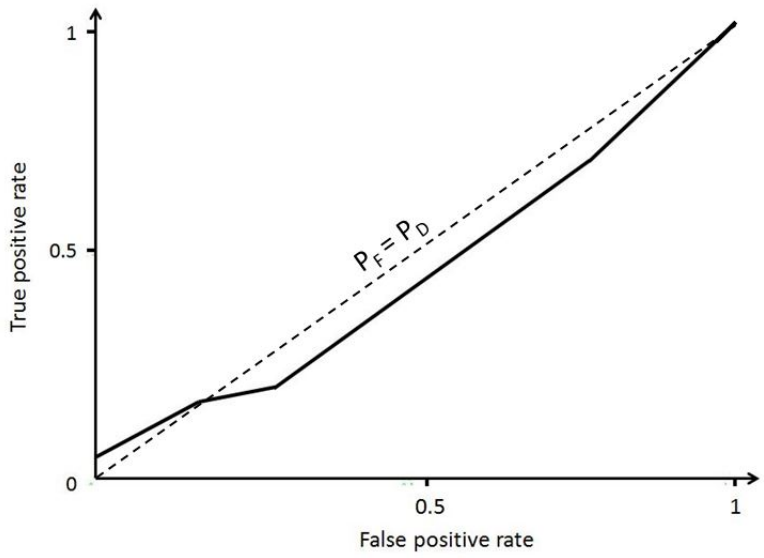
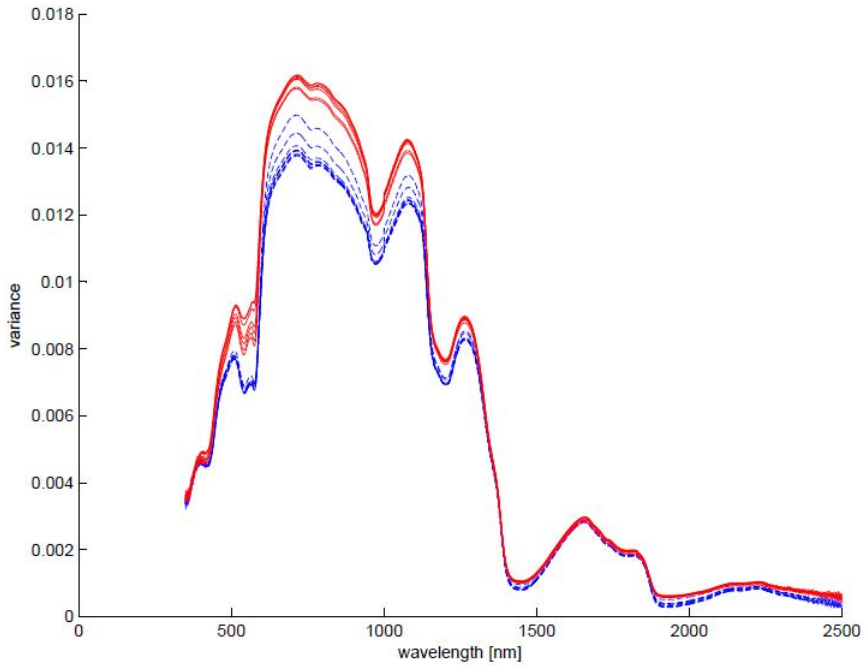


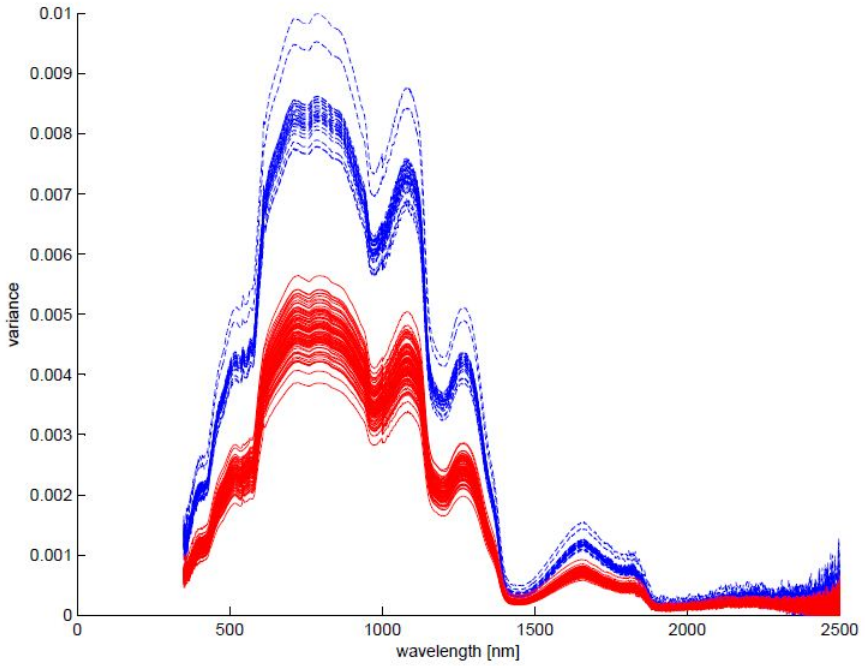

**Figure 4.9: ROC curve for the most successful validation result among the Var validation sets, which results from ReliefF features and a decision tree classifier. The Var dataset is validated with data collected with a NC fore optic. The set is comprised of 530 samples of "stress" and "non-stress" and six features.**

## 4.2 Validation on Non-Contact Data

The data recorded using the NC fore optic is used for two functions: to validate models built from contact data and to build new models. Because the NC collections resulted in different responses compared to the contact collections, new models are trained and tested. Figure 4.10 shows the comparison of one subject's contact and NC reflectance spectral responses. Notice that the "non-stress" (dashed blue) reflectance values are lower than "stress" (solid red) for a contact recording and higher for the NC recording. This could be an artifact of the test setup and collection and should be investigated in the future. The NC collection contains many more samples than the contact dataset; for example, the contact Combo set has 19 samples and NC has 600 samples. This is because the amount of time the contact probe rests on a subject's skin is limited due to warmth felt from the probe's illumination source, while the NC optic can record for longer time periods.




(a) Normalized reflectance response recorded with a contact probe for one subject.




(b) Normalized reflectance response recorded with a NC fore optic for one subject.

**Figure 4.10: (a) is a normalized reflectance response recorded with a contact probe on one subject (b) is the normalized reflectance response of the same subject with the data recorded using a NC fore optic. The red solid lines are "stress" and blue dashed lines "non-stress." This shows the difference that can exist for one subject between a contact and NC data collection. Because of this difference, models are trained and built for both datasets.**

Similarly to the contact data feature sets, all the SubjectsNC 1-6 feature sets provided class separation, leading to high classification results across all twenty sets (five subjects because one subject was thrown out and four feature sets for each subject). The three feature selection algorithms all returned high classification results for the ComboNC dataset. These wavelengths are listed in Table 4.10 along with the wavelengths of the most discrimination between classes for the VarNC dataset, from the SVM AE feature set.

**Table 4.10: Wavelengths of maximum discrimination between classes for ComboNC and VarNC datasets.**

| | ComboNC | | | | VarNC |
|---|---|---|---|---|---|
| | ReliefF | SVM AE | NASAFS-IDF1 | NASAFS-IDF2 | SVM AE |
| wavelength [nm] | 1203 | 830 | 605 | 985 | 2382 |
| | 1206 | 833 | 995 | 595 | 1974 |
| | 1202 | 1201 | 365 | 365 | 2108 |
| | 1207 | 922 | 355 | 1615 | 2327 |
| | 1201 | 1202 | 1385 | 1655 | 2381 |
| | 1204 | 831 | 1615 | 1285 | 1960 |

Based on prior literature, oxygenated hemoglobin is known to have a lower reflectance than deoxygenated hemoglobin at 830nm, which appears in Table 4.10 [84]. Interestingly, the literature also indicates that there is a dip in oxygenated hemoglobin with respect to deoxygenated hemoglobin at 690nm [84]; however, this wavelength was not identified in any of the feature sets.

Overall, the results from NC validation on models trained with NC data returned high accuracies for all datasets, with 100% accuracy for most individual SubjectNC reflectance. This is because there is good separation between classes on an individual subject basis. For the ComboNC and VarNC validation sets, the highest accuracy and AUC results across the three feature selection algorithms are displayed in Table 4.11. The ComboNC validation set has several feature set/classifier pairs that give 100% accuracy. The highest returning ComboNC validation results are mostly 100% because the features selected offer wide class

separation, as can be seen in Fig. 4.11. One of these pairs is chosen (SVM and naive Bayes) and the confusion matrix is given in Table 4.12.

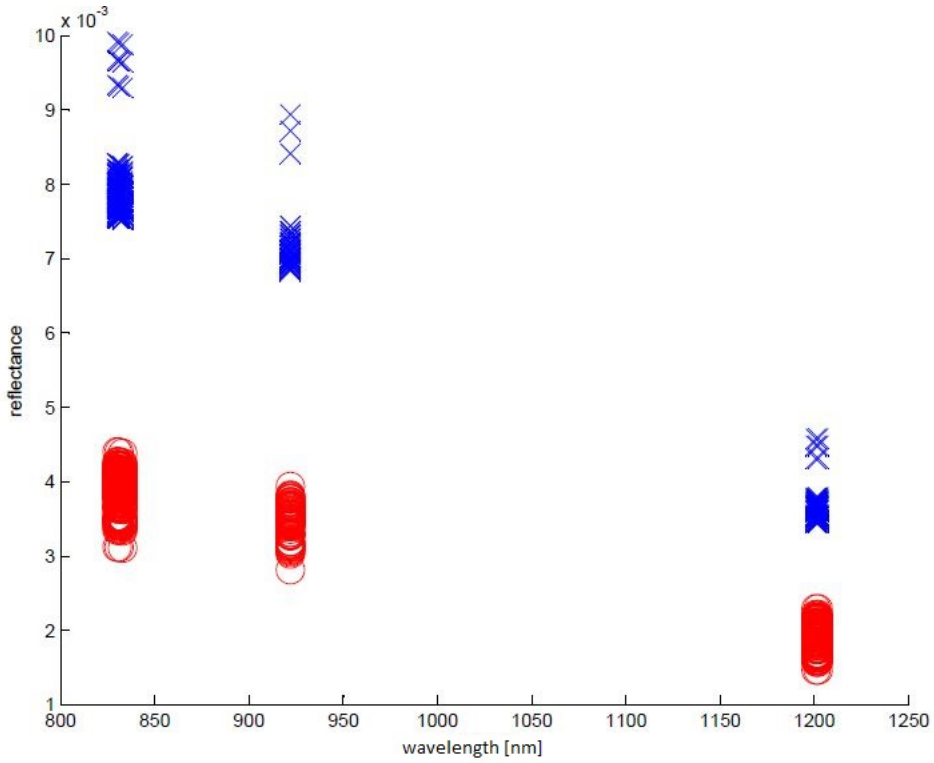



**Figure 4.11: This is the ComboNC validation set with six features from SVM AE. There is a wide margin between classes, thus making 100% accuracy achievable, as displayed in Table 4.11 The features include the wavelengths (in nm): 830, 833, 1201, 922, 1202, 831.**

**Table 4.11: Validation percent accuracy and AUC on ComboNC and VarNC models built using NC data. The sets with results displayed returned the highest accuracy for each feature selection algorithm, across the three classifiers. The sets include six features selected using the feature selection (FS) algorithms, ReliefF (RF), SVM AE (SVM), and NASAFS-IDF (NAS1/2) and are evaluated using the classifiers, naive Bayes (NB), SVM, and a decision tree (DT).**

| Dataset | FS | Classifier | Accuracy | AUC |
|---|---|---|---|---|
| ComboNC | RF | NB | 98.00 | 1.000 |
| ComboNC | SVM AE | NB | 100.00 | 1.000 |
| ComboNC | SVM AE | SVM | 100.00 | 1.000 |
| ComboNC | NAS1 | SVM | 100.00 | 1.000 |
| ComboNC | NAS2 | SVM | 100.00 | 1.000 |
| VarNC | RF | NB | 64.77 | 0.870 |
| VarNC | SVM AE | SVM | 90.90 | 0.968 |
| VarNC | NAS1 | DT | 88.06 | 0.806 |
| VarNC | NAS2 | DT | 89.77 | 0.903 |

**Table 4.12: A selected confusion matrix for one of the top performing feature set (SVM AE) and classifier (naive Bayes) on the ComboNC validation set. Several other pairs (ReliefF/naive Bayes, SVM AE/SVM, NASAFS-IDF1/SVM, NASAFS-IDF2/SVM) also returned an accuracy and/or AUC of 100% and 1.000.**

| | Stress | Non-stress | |
|---|---|---|---|
| S | 49 | 0 | 49 |
| NS | 0 | 151 | 151 |
| | 49 | 151 | 200 |
| Accuracy: 100.00% | | | |
| AUC: 1.000 | | | |

The VarNC validation set results varied based on feature set. The SVM AE feature set returned the highest results, averaging 94.31% accuracy with an average AUC of 0.954. The NASAFS-IDF feature sets returned accuracies in the mid-to-high eighties with an average AUC of 0.765. The SVM AE with a SVM classifier has the highest accuracy, at 97.15%, but the same feature set combined with a naive Bayes classifier has a higher AUC at 0.968, compared to 0.952 for the SVM classifier. Table 4.13 summarizes results and Fig. 4.12 provides the ROC curve for the VarNC test set.

**Table 4.13: Confusion matrix for the most successful feature set (SVM AE) and classifier (naive Bayes) on the VarNC validation set. The validation set is comprised of 176 samples of "stress" and "non-stress" and six features.**

| | Stress | Non-stress | |
|---|---|---|---|
| S | 32 | 5 | 37 |
| NS | 11 | 128 | 139 |
| | 43 | 133 | 160 |
| Accuracy: 90.90% | | | |
| AUC: 0.968 | | | |

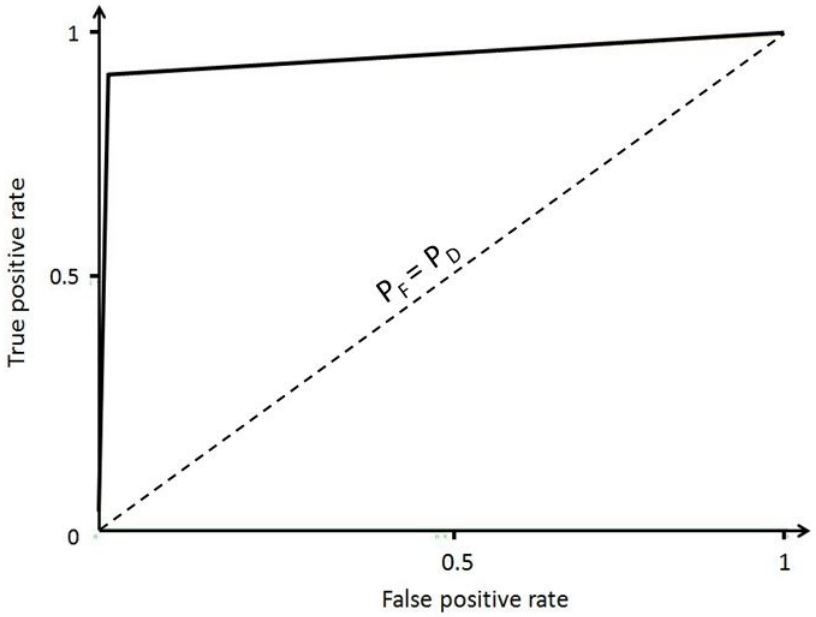



**Figure 4.12: ROC curve from most successful validation result on the VarNC validation set, which is SVM AE features and a naive Bayes classifier. The set includes all subjects' normalized reflectance variance, comprising 176 samples of "stress" and "non-stress" and six features.**

## 4.3 Analysis

The datasets described in Table 3.2 are first processed through the three classifiers to build, train, and test models. Following this, an unseen portion of the contact dataset is used for validation purposes. For a comparison to a real-world scenario, the models are validated again using data from a NC collection. The train/test and contact validation results returned an average accuracy of above 90% on all datasets. The NC validation results are low overall,

with a few exceptions. The next step is to analyze the outcome of each algorithm in order to introduce a potential model for stress detection.

There are two aspects of the stress detection process that need analysis: the feature selection algorithms and classification algorithms. A total of sixteen datasets are processed through three types of feature selection algorithms. This results in four feature sets (because NASAFS-IDF provides two feature sets) of six features for each dataset. Table 4.14 displays the average accuracies for contact validation (using one-third of the contact dataset) and NC validation on NC data (using one-third of the NC dataset). Table 4.15 displays the average AUC for each validation dataset.

**Table 4.14: Percent accuracy for each respective feature selection and classification algorithm as they apply to each contact and NC validation dataset. The datasets include the average results of Subject 1-6(C/NC), Combo(C/NC), and Var(C/NC) datasets. The contact data is validated on one-third of the contact dataset and the NC data is validated on one-third of the NC dataset.**

| Dataset: | Sub1−6 | Combo | Var | SubNC | ComboNC | VarNC | Average |
|---|---|---|---|---|---|---|---|
| ReliefF | 98.76 | 74.36 | 80.95 | 100.00 | 99.00 | 41.10 | 82.36 |
| SVM FS | 98.76 | 83.33 | 89.52 | 100.00 | 99.83 | 94.31 | 94.29 |
| NASAFS-IDF1 | 99.07 | 86.95 | 71.42 | 99.69 | 98.67 | 86.36 | 90.36 |
| NASAFS-IDF2 | 93.52 | 85.50 | 73.33 | 99.69 | 98.67 | 88.70 | 89.90 |
| Naive Bayes | 97.92 | 77.51 | 66.42 | 100 | 97.88 | 81.15 | 86.81 |
| SVM | 98.61 | 79.35 | 78.57 | 100 | 99.88 | 74.75 | 88.53 |
| Decision Tree | 96.06 | 90.76 | 91.42 | 99.51 | 99.38 | 78.07 | 92.53 |

**Table 4.15: AUC for each respective feature selection and classification algorithm as they apply to each validation dataset. The datasets include the average results of Subject 1-6(C/NC), Combo(C/NC), and Var(C/NC) datasets. The contact data is validated on one-third of the contact dataset and the NC data is validated on one-third of the NC dataset.**

| Dataset: | Sub1−6 | Combo | Var | SubNC | ComboNC | VarNC | Average |
|---|---|---|---|---|---|---|---|
| ReliefF | 0.9907 | 0.8041 | 0.8250 | 1.0000 | 0.9980 | 0.8843 | 0.9170 |
| SVM FS | 0.9907 | 0.8887 | 0.8950 | 1.0000 | 0.9990 | 0.9536 | 0.9545 |
| NASAFS-IDF1 | 0.9969 | 0.8944 | 0.7637 | 0.9979 | 0.9976 | 0.7673 | 0.9030 |
| NASAFS-IDF2 | 0.9167 | 0.8850 | 0.8040 | 0.9979 | 0.9970 | 0.7626 | 0.8939 |
| Naive Bayes | 0.9791 | 0.8719 | 0.8533 | 1.0000 | 0.9985 | 0.7682 | 0.9254 |
| SVM | 0.9773 | 0.7908 | 0.6923 | 1.0000 | 0.9992 | 0.7682 | 0.8713 |
| Decision Tree | 0.9630 | 0.9415 | 0.9203 | 0.9969 | 0.9960 | 0.9082 | 0.9543 |

A feature selection algorithm that returns features maximizing class separation is essential because the classification algorithms applied rely on a distinction between classes. By averaging the accuracies and AUC values from each feature set on contact and NC validation datasets, the SVM AE returns the highest results at 94.29% and an AUC of 0.9545. By averaging the accuracies and AUC values from each classifier on contact and NC validation datasets, the decision tree classifier returned the highest results at 92.53% accuracy and an AUC of 0.9543. The six algorithms are further analyzed.

ReliefF returns six wavelengths that are located within close proximity to each other on the electromagnetic spectrum. A small grouping of features can be functional if the majority of the dataset does not provide class separation or if such features selected result in significantly wider class separation. ReliefF could pose the issue of overcompensating by selecting all similar features. In such a case, one feature may accomplish the same level of accuracy as the entire selected feature set. Then, it would be more efficient to select features across the dataset to take advantage of the contributions of other features, which is a specific goal in the NASAFS-IDF feature selection algorithm. In general, the ReliefF algorithm returned results similar to the other feature selection algorithms. This is because skin has a distinct reflectance signature that results in consistent separation between classes at certain points for each subject. When comparing results of individual subjects (Subject 1-6) using ReliefF features to results on the Combo datasets, the ReliefF features do not return as high accuracies. ReliefF features average the lowest accuracy across the three classifiers on the Combo dataset for both contact and "real-world" validation. The average accuracy for classifying variance with ReliefF feature set is second highest for validation and highest for "real-world" validation. The results from models trained/tested on NC data are high using ReliefF features from individual subject's reflectance responses and the Combo validation set, but the accuracy of the NC variance validation set is as low as 41.10%, which

is significantly lower than that of the other two feature selection algorithms. Table 4.16 summarizes the classification results for each dataset and classification algorithm.

**Table 4.16: This table summarizes the classification results on contact validation sets using the ReliefF feature set. The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| ReliefF | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 66.56 | 0.830 |
| | Var | 77.14 | 0.812 |
| | Average | 92.96 | 0.955 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 67.39 | 0.643 |
| | Var | 80.00 | 0.728 |
| | Average | 93.42 | 0.921 |
| Decision Tree | Sub1 | 88.88 | 0.917 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 88.88 | 0.917 |
| | Combo | 89.13 | 0.939 |
| | Var | 85.71 | 0.935 |
| | Average | 94.08 | 0.964 |

The SVM AE selects features based off an area of widest separation between classes. The feature sets can be categorized in between the results of ReliefF and NASAFS-IDF because the algorithm selects groups of features in close proximity, but also selects some wavelengths across the dataset. The SVM AE's features returned similar results to that of

ReliefF and NASAFS-IDF for the individual subject's reflectance classification (Subject 1-6), as well as the Combo dataset. SVM AE returned higher accuracies than NASAFS-IDF when classifying variance contact validation sets, but did not succeed ReliefF's variance contact validation results. SVM AE features are the most accurate on all NC validation sets from models built using the NC data. Table 4.17 outlines the accuracies and AUC values for contact validation using SVM AE features.

**Table 4.17: This table summarizes the classification results on contact validation sets using the SVM AE feature set (FS). The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| SVM Attribute Evaluator | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 82.60 | 0.897 |
| | Var | 85.71 | 0.935 |
| | Average | 96.04 | 0.979 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 84.78 | 0.865 |
| | Var | 91.42 | 0.875 |
| | Average | 97.03 | 0.968 |
| Decision Tree | Sub1 | 88.88 | 0.917 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 88.88 | 0.917 |
| | Combo | 82.60 | 0.904 |
| | Var | 91.42 | 0.875 |
| | Average | 93.97 | 0.952 |

NASAFS-IDF provides the widest range of features across the electromagnetic spectrum. When classifying the combination of all subjects' reflectance, NASAFS-IDF features result in an improvement in classification accuracy from ReliefF and SVM AE. For Subject 1-6 classification, NASAFS-IDF does not differ significantly compared to the other two feature selection algorithms. The features selected for classifying variance data do not return as of high accuracies compared to ReliefF and SVM AE. This could be due to the fact that across the electromagnetic spectrum, there is not significant separation between classes for all samples, as can be observed in Fig. 4.5. ReliefF and SVM AE select features grouped together that lead to the greatest separation between classes. NASAFS-IDF omits the possibility of selecting features grouped together, so the algorithm forces the selection of a wider range of features. This affected classifying variance more so than classifying the normalized reflectance of subjects because the normalized reflectance data has greater separation between classes across the entire dataset. NASAFS-IDF's feature sets gave high accuracies on all sets validated on models built with NC data. The "real-world" validation results for Subjects 1-6 datasets and the Combo dataset are similar to the results using ReliefF and SVM AE. Features from NASAFS-IDF evaluated on the NC variance data returned results second highest at 88.27% accuracy compared to 94.31% accuracy of SVM AE.

Both feature sets from NASAFS-IDF perform similarly overall. Since NASAFS-IDF compares the two classes to each other, the feature sets should return similar results. The algorithm is stochastic, so there are differences in the selected features. Table 4.18 and 4.19 outline the classification results for the NASAFS-IDF feature sets.

**Table 4.18: This table summarizes the classification results on contact validation sets using the NASAFS-IDF1 feature set. The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| NASAFS-IDF1 | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 80.43 | 0.877 |
| | Var | 51.42 | 0.833 |
| | Average | 91.48 | 0.964 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 84.78 | 0.845 |
| | Var | 71.42 | 0.583 |
| | Average | 94.53 | 0.929 |
| Decision Tree | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 83.33 | 0.944 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 95.65 | 0.961 |
| | Var | 91.42 | 0.875 |
| | Average | 96.30 | 0.973 |

**Table 4.19: This table summarizes the classification results on contact validation sets using the NASAFS-IDF2 feature set. The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| NASAFS-IDF2 | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 83.33 | 1.000 |
| | Sub6 | 66.66 | 0.500 |
| | Combo | 80.43 | 0.883 |
| | Var | 51.42 | 0.833 |
| | Average | 85.23 | 0.902 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 66.66 | 0.500 |
| | Combo | 80.43 | 0.810 |
| | Var | 71.42 | 0.583 |
| | Average | 89.81 | 0.862 |
| Decision Tree | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 66.66 | 0.500 |
| | Combo | 95.65 | 0.962 |
| | Var | 97.14 | 0.996 |
| | Average | 94.93 | 0.932 |

The three classification algorithms each perform well on certain datasets and poorly on others. Overall, the classifiers returned high accuracies on contact validation sets and validation sets from models trained on NC data, but not on "real-world" validation sets, which are from models trained on contact data and validated with NC data.

For the contact validation sets, Subjects 1-6 classification had similar results across all three classifiers in the range of above 95%, with the SVM classifier as the highest average accuracy at 98.61%. The Combo contact validation set accuracies ranged from an average

of 77.51% for the naive Bayes classifier to 90.76% for the decision tree classifier. Similarly for variance contact validation sets, naive Bayes gave the lowest accuracy on average and decision tree the highest. For the contact validation sets, the top performers were the naive Bayes classifier, with SVM AE features.

Classification on the models trained/tested with NC data gave high accuracies on Subjects 1-6 and Combo validation sets, with an average of over 99% across the three classifiers. The variance processed and tested from the NC collection did not return as high of accuracies, though still gave an average of 81.15% with the naive Bayes classifier as the highest. The models that are trained and tested on NC data are used to report overall results for the most accurate feature selection and classification algorithm. This is because this type of data collection represents a real-world scenario: recording data using a NC probe and testing that data against unseen data recording with a NC probe. Results for each feature selection and classifier pair are in Tables 4.20-4.23.

**Table 4.20: This table summarizes the classification results on contact validation sets using the ReliefF feature set. The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| ReliefF | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 98.00 | 1.000 |
| | Var | 64.77 | 0.870 |
| | Average | 95.35 | 0.984 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 99.50 | 0.997 |
| | Var | 30.68 | 0.800 |
| | Average | 91.27 | 0.975 |
| Decision Tree | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 99.50 | 0.997 |
| | Var | 27.84 | 0.983 |
| | Average | 90.92 | 0.998 |

**Table 4.21: This table summarizes the classification results on contact validation sets using the SVM AE feature set (FS). The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| SVM Attribute Evaluator | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 100.00 | 1.000 |
| | Var | 90.90 | 0.968 |
| | Average | 98.86 | 0.996 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 100.00 | 1.000 |
| | Var | 97.15 | 0.952 |
| | Average | 99.64 | 0.994 |
| Decision Tree | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 99.50 | 0.997 |
| | Var | 94.88 | 0.941 |
| | Average | 99.30 | 0.992 |

**Table 4.22: This table summarizes the classification results on contact validation sets using the NASAFS-IDF1 feature set. The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| NASAFS-IDF1 | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 96.50 | 0.996 |
| | Var | 85.22 | 0.834 |
| | Average | 97.71 | 0.958 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 100.00 | 1.000 |
| | Var | 85.79 | 0.662 |
| | Average | 98.22 | 0.958 |
| Decision Tree | Sub1 | 97.91 | 0.987 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 83.33 | 0.944 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 0.982 |
| | Combo | 99.50 | 0.997 |
| | Var | 88.06 | 0.806 |
| | Average | 98.45 | 0.975 |

**Table 4.23: This table summarizes the classification results on contact validation sets using the NASAFS-IDF2 feature set. The results include the accuracy and AUC for each dataset and all accuracies and AUC values averaged.**

| NASAFS-IDF2 | | | |
|---|---|---|---|
| Classifier | Dataset | Accuracy | AUC |
| Naive Bayes | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 97.00 | 0.982 |
| | Var | 84.65 | 0.726 |
| | Average | 97.71 | 0.964 |
| SVM | Sub1 | 100.00 | 1.000 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 100.00 | 1.000 |
| | Combo | 100.00 | 1.000 |
| | Var | 85.22 | 0.659 |
| | Average | 98.15 | 0.957 |
| Decision Tree | Sub1 | 97.91 | 0.987 |
| | Sub2 | 100.00 | 1.000 |
| | Sub3 | 100.00 | 1.000 |
| | Sub4 | 100.00 | 1.000 |
| | Sub5 | 100.00 | 1.000 |
| | Sub6 | 97.43 | 0.982 |
| | Combo | 99.00 | 0.993 |
| | Var | 89.77 | 0.903 |
| | Average | 98.60 | 0.987 |

To summarize findings, Table 4.24 outlines the most accurate feature selection and classification algorithms based off the average percent accuracies and average AUC for NC validation sets.

## 4.4 Summary

This chapter detailed the results and analysis of the experiment and initial training completed in Chapter 3. Numerous datasets were created according to Table 3.2 using data collected from a contact probe and a NC fore optic and features selected using ReliefF,

SVM AE, and NASAFS-IDF. These datasets were trained and models were built, with initial results presented in Chapter 3. The models were validated using contact data and with data collected using a NC probe. The NC data is also utilized for training/testing models to allow a real-world scenario for stress detection.

Based off results from validating models built on the NC data, the SVM AE features combined with a SVM classifier returned the highest accuracy, though ReliefF features with a decision tree classifier gave the highest AUC. This feature selection algorithm and classification algorithm did not consistently provide the highest classification results for every dataset. When examining the results from models built and validated with contact data, SVM AE and a naive Bayes classifier gave the highest accuracy and the NASAFS-IDF1 feature set with a decision tree gave the highest AUC. These differences are most likely a result from the inconsistencies that arise using a NC probe. Using the contact models, Subjects 1-6 classification accuracies remained in the high 90's for all feature selection and classification algorithms, Combo accuracies in the high 80's for NASAFS-IDF features and decision tree classifier, and Var accuracies in the low 90's for SVM AE features and SVM and decision tree classifiers. The NC models also returned high results for all feature selection and classification algorithms on all datasets except the VarNC, where only the SVM AE feature set provided good results with the three classifiers. Table 4.24 summarizes the top performing feature selection and classification algorithms for both contact and NC models. The winners are determined based on the average accuracy and AUC over all datasets (Subject 1-6, Combo, Var).

**Table 4.24: A conclusion of the classification results to reveal the top performing feature selection (FS) and classification algorithms. The algorithms are chosen from validation results on models trained/tested with contact data and models trained/tested with NC data, which best simulates a real-world scenario. The results are based off the average percent accuracies and average AUC. The feature selection algorithms evaluated are ReliefF, SVM AE (SVM), and NASAFS-IDF1/2 (NAS1/2). The classification algorithms evaluated are naive Bayes (NB), SVM, and decision tree (DT).**

| | Contact | | | |
|---|---|---|---|---|
| | FS | Classifier | Accuracy | AUC |
| Highest accuracy: | NAS1 | DT | 96.30 | 0.973 |
| Highest AUC: | SVM AE | NB | 96.04 | 0.979 |
| | Non-Contact | | | |
| Highest accuracy: | SVM AE | SVM | 99.64 | 0.994 |
| Highest AUC: | ReliefF | DT | 90.92 | 0.998 |

# V. Conclusion

This chapter summarizes the methods, results, and important points of stress detection by hyperspectral imaging (HSI) and proposes ideas for future work. Section 5.1 summarizes the method followed and the results discovered in this thesis. Future research in the areas of stress detection using hyperspectral data is discussed in Section 5.2 and contributions of this work are discussed in Section 5.3.

## 5.1 Summary of Methodology and Conclusions

This thesis presents a method of stress detection using HSI by analyzing normalized reflectance and reflectance variance curves. The research explores three feature selection algorithms and three classifiers to distinguish the reflectance signatures of stressed from non-stressed subjects. The feature selection algorithms include, ReliefF, Support Vector Machine Attribute Evaluator (SVM AE), and Non-Correlated Aided Simulated Annealing Feature Selection–Integrated Distribution Function (NASAFS-IDF). The classification algorithms include, naive Bayes, SVM, and decision tree. Each feature selection algorithm and classifier processes data differently, so three of each are chosen and evaluated. The feature selection algorithms and classifiers are compared to each other based on their accuracy and receiver operating characteristic (ROC) curve results, to include the area under the curve (AUC).

First, skin reflectance was recorded on six different subjects as hyperspectral data in the area of the carotid artery. The carotid was imaged because it is one of the largest blood vessels in the body and its' location is easily accessible for contact and non-contact (NC) imaging purposes. The carotid is responsible for delivering oxygen rich blood, which is relevant because previous research confirms a change in the hemoglobin oxygen saturation ($HbO_2$) at the onset of stress [44]. One of the main focuses of this thesis was based on

the change of the amount of oxygen in the blood, resulting in changes in the reflectance signature of skin between stress and non-stress states. A contact probe was used to collect clean, noiseless data that was used for training models and a NC fore optic collection was used for validation of the models built, as well as training NC models. The data was processed in MATLAB® to acquire datasets of normalized reflectance and reflectance variance and Waikato Environment for Knowledge Analysis (WEKA) was used to select features and build and test models. After training classifiers, testing was accomplished using an unused portion of the contact datasets. Then the validation sets were applied to the models. Validation sets included data recorded with the NC fore optic.

The accuracies and AUC were calculated for each feature selection and classification algorithm pair. Over all contact validation sets, the NASAFS-IDF1 feature set with a decision tree classifier gave the highest accuracy at 96.30% and the SVM AE feature set with a naive Bayes classifier gave the highest AUC at 0.979. The NC validation results had a high accuracy of 99.64% with SVM features and a SVM classifier. The highest AUC for NC validation sets was 0.998 with ReliefF features and a decision tree classifier.

## 5.2 Future Work

The following introduces potential topics of future work to expand on the research covered in this thesis. First, increasing the size of the database used for training and testing would provide more robust models. The feature selection and classification algorithms that returned the highest success rates should be retrained and evaluated with new data. Second, the aspect of skin tone diversity was not emphasized in this research, but could play a role in stress detection and should be examined. As discussed in previous literature [38], the concentration of melanin in the skin alters skin reflectance. Also, aspects of consistency throughout the data collection and further applications of reflectance variance are discussed in Sections 5.2.1 and 5.2.2, respectively. The overall contributions from this thesis to stress detection and analysis are given in Section 5.3.

### *5.2.1 Non-Contact Data Collection.*

The results from contact data testing and validating have high accuracies and AUC, offering potential for stress detection. However, one of the goals of this thesis was to design a model that can differentiate between individuals experiencing stress from those that are not via non-invasive methods. Certain aspects of the NC data collection can be improved for more accurate results, such as consistency and imaging location on the skin.

Consistency of probe placement for each subject is essential to acquire data relevant to across-subject classification. Areas for improved consistency for a NC collection include lighting, placement of the probe, and timing of the collection. It is important to keep lighting consistent throughout collections because scene illumination has a direct impact on reflectance. A recording taken with poor lighting results in an attenuated skin signature that affects classification results. For example, if a "non-stress" recording has proper illumination, but for the second collection, the "stress" case, the illumination is different, the resulting data is not completely accurate; the two datasets do not have comparable reflectance values.

The NC fore optic is in the shape of a pistol that is mounted on a tripod. The probe has a laser sight to assist in alignment for data collection. Even with the ability to align the camera with a specific region-of-interest (ROI), if the subject being imaged moves, the field-of-view (FOV) is disrupted. In this regard, there is potential for research to investigate different areas of the skin to image, with a focus on increasing the ROI to allow subject movement. One area that may provide improved results is the cheek, which contains numerous capillaries at the surface of the skin and is a large and relatively flat surface.

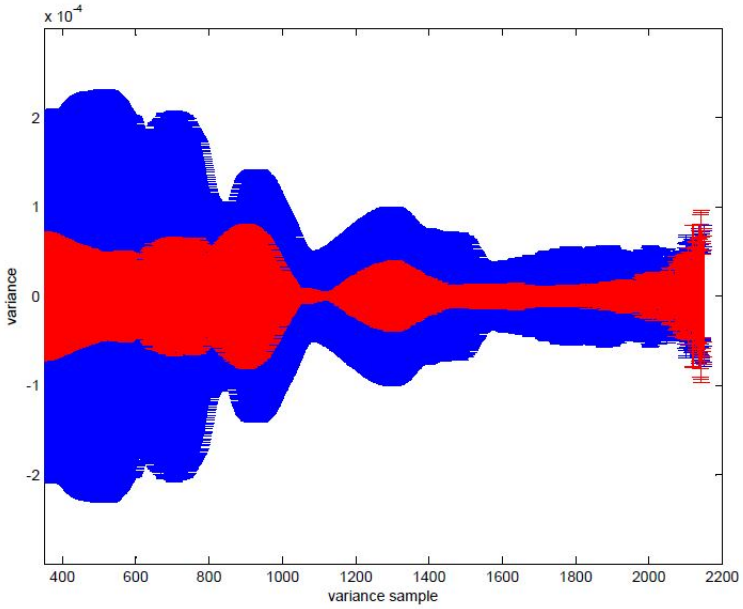
The "stress" data is collected while the subject accomplishes a level of Air Force Multi-Attribute Test Battery (AF_MATB), which lasts five minutes. In this study, the spectroradiometer began recording with the NC fore optic shortly after one minute elapsed and stopped recording near the end of the level. Possible areas of research are to examine

the results based on when recordings begin and end and to analyze the length of recording necessary to identify a detection.

### *5.2.2 Variance.*

There are more potential applications of reflectance variance that can be explored to aid in stress detection. In this thesis, a sliding window approach was used to calculate variance. The window ranged in number of samples from five to ten, depending on the dataset size. Future implementation of a stress detection method using variance may require more or less samples and/or may utilize an improved method for calculating variance instead of a sliding window.

The variance graphs presented display calculated variance across all features. In examining variance graphed with error bars, the shape suggests the likeness of a speech waveform, as in Fig. 5.1. If these waveforms can be processed through voice analysis, this response has the potential to provide an audio alert when someone reaches a state of stress. This type of detection would work well if the variance results are consistent across multiple subjects.




(a)

(b)

**Figure 5.1: (a) is the variance waveform with error bars. This plot shows "stress" variance (red) and "non-stress" variance (blue). (b) is an example of a speech waveform. The similarities indicate the potential for variance to be used as an audio indicator of acute stress.**

Lastly, though there has been an emphasis on stress detection in the workplace environment, heart rate variability (HRV) may provide some insight to detect a subject falling asleep, which is also related to workplace productivity. It is known that HRV is high when a subject is awake and in a rested state and that HRV is low when a subject experiences a high workload [71]. When one falls asleep, the heart rate (HR) decreases as the body enters a new state of rest, where, potentially, the HRV also decreases. Initial testing has been accomplished on this method [67], where there was a gradual increase

in HRV when fatigue decreased a subject's attention. Further research on the HRV during sleep and the reflectance variance during sleep can be accomplished.

## 5.3 Contributions

This thesis provided results and analysis on different feature selection and classification algorithms applied to "stress" and "non-stress" hyperspectral data. Testing a range of algorithms allowed the ability to compare results to lead to the best model. This research showed that a SVM AE feature set and naive Bayes classifier built the most successful model for contact validation and a ReliefF feature set and decision tree classifier built the most successful model for NC validation on NC models. The NC data provides a real-world scenario of data collection and testing that can be used in further research in the area of stress detection. The variance of reflectance introduced another probable avenue of stress detection that has not been demonstrated prior to this work.

## Appendix A: Sample heart rate (HR) collected with the electrocardiogram (ECG).

**Table A.1: After processing the HR waveform produced by the ECG, time-stamped beats per minute (bpm) are output. Below is an example of an ECG recording lasting approximately 10 seconds. The cycle number indicates each QRS pulse.**

| Cycle | Time | Heart Rate |
|---|---|---|
| 1 | 0.817 | 60.181 |
| 2 | 1.814 | 67.189 |
| 3 | 2.707 | 57.526 |
| 4 | 3.750 | 60.667 |
| 5 | 4.739 | 61.038 |
| 6 | 5.722 | 61.038 |
| 7 | 6.705 | 61.350 |
| 8 | 7.683 | 69.124 |
| 9 | 8.551 | 68.886 |
| 10 | 9.422 | 66.152 |

## Appendix B: Roster of subjects who participated in the experiment.

**Table B.1: This list encompasses all subjects that participated in the experiment of this thesis.**

| Subject # | Age | Gender | Skin Tone |
|---|---|---|---|
| 1 | 26 | M | White |
| 2 | 23 | F | White |
| 3 | 24 | F | White |
| 4 | 22 | F | White |
| 5 | 24 | M | White |
| 6 | 23 | M | White |

## Appendix C: Feature selection results for Subjects 2-6, ComboNC, and VarNC.

### C.1 Feature selection results for Subjects 2-6.

**Table C.1: Feature selection results on the datasets of Subjects 2-6. Each dataset contains samples of "stress" and "non-stress" collected using a contact probe and is processed through the feature selection algorithms ReliefF, SVM AE, and NASAFS-IDF to achieve a feature set of six features.**

| Dataset | ReliefF | SVM AE | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| Sub2 | 593 | 593 | 545 | 575 |
| wavelength [nm] | 592 | 592 | 1315 | 1315 |
| | 594 | 1121 | 865 | 2495 |
| | 591 | 594 | 375 | 1685 |
| | 595 | 591 | 2495 | 1925 |
| | 1123 | 590 | 2035 | 845 |
| Sub3 | 1610 | 1706 | 575 | 575 |
| wavelength [nm] | 1611 | 1712 | 1285 | 1625 |
| | 1609 | 1711 | 995 | 985 |
| | 1593 | 1713 | 2375 | 375 |
| | 1600 | 1287 | 2045 | 2495 |
| | 1614 | 1730 | 1885 | 2325 |
| Sub4 | 1213 | 1213 | 575 | 575 |
| wavelength [nm] | 1212 | 1212 | 1305 | 1305 |
| | 1214 | 1211 | 2265 | 365 |
| | 1215 | 560 | 395 | 1745 |
| | 1211 | 1210 | 365 | 1665 |
| | 1210 | 1214 | 1925 | 1135 |
| Sub5 | 2095 | 2079 | 1895 | 1895 |
| | 2096 | 2080 | 445 | 445 |

Continued on Next Page...

Table C.1 – Continued

| Dataset | ReliefF | SVM AE | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| wavelength [nm] | 2099 | 1211 | 2245 | 2065 |
| | 2098 | 357 | 2075 | 2245 |
| | 2138 | 1212 | 1015 | 1575 |
| | 1936 | 381 | 1345 | 1385 |
| Sub6 | 1006 | 1006 | 585 | 575 |
| wavelength [nm] | 1007 | 1004 | 1365 | 1335 |
| | 1005 | 1005 | 355 | 355 |
| | 1017 | 1007 | 1665 | 1655 |
| | 1016 | 1003 | 1835 | 2475 |
| | 1018 | 1013 | 1175 | 2025 |

## C.2 Feature selection results for Subjects 1-6NC.

**Table C.2: Feature selection results on the datasets of Subjects 2-6NC. Each dataset contains samples of ”stress” and ”non-stress” from data collected with a stand-off fore optic. The datasets are processed through the feature selection algorithms ReliefF, SVM AE, and NASAFS-IDF to achieve a feature set of six features.**

| Dataset | ReliefF | SVM AE | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| Sub1NC | 369 | 535 | 375 | 375 |
| wavelength [nm] | 497 | 544 | 1635 | 1495 |
| | 489 | 545 | 1365 | 1365 |
| | 488 | 542 | 2495 | 1655 |
| | 496 | 546 | 1895 | 565 |
| | 490 | 586 | 545 | 1995 |
| Sub2NC | 544 | 544 | 2495 | 2415 |

Continued on Next Page...

Table C.2 – Continued

| Dataset | ReliefF | SVM AE | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| wavelength [nm] | 543 | 543 | 2035 | 1995 |
| | 545 | 545 | 2305 | 2215 |
| | 542 | 542 | 1285 | 1295 |
| | 546 | 546 | 1265 | 1185 |
| | 547 | 547 | 915 | 1755 |
| Sub4NC | 611 | 611 | 2475 | 2475 |
| wavelength [nm] | 612 | 612 | 2075 | 2065 |
| | 610 | 610 | 635 | 615 |
| | 613 | 613 | 1065 | 935 |
| | 614 | 614 | 915 | 2295 |
| | 615 | 544 | 1895 | 1895 |
| Sub5NC | 544 | 545 | 545 | 465 |
| wavelength [nm] | 545 | 544 | 1065 | 1555 |
| | 543 | 543 | 1305 | 1195 |
| | 542 | 611 | 1295 | 975 |
| | 546 | 612 | 895 | 2045 |
| | 541 | 546 | 1675 | 1955 |
| Sub6NC | 886 | 821 | 2485 | 2475 |
| wavelength [nm] | 911 | 820 | 2145 | 355 |
| | 892 | 830 | 355 | 2095 |
| | 890 | 829 | 1675 | 2265 |
| | 912 | 822 | 1445 | 1765 |
| | 876 | 819 | 2315 | 1575 |

**Table C.3: Feature selection results on the datasets of ComboNC and VarNC. Each dataset contains samples of ”stress” and ”non-stress” from data collected with a stand-off fore optic. The datasets are processed through the feature selection algorithms ReliefF, SVM AE, and NASAFS-IDF to achieve a feature set of six features.**

| Dataset | ReliefF | SVM AE | NASAFS-IDF1 | NASAFS-IDF2 |
|---|---|---|---|---|
| ComboNC | 1203 | 830 | 605 | 985 |
| wavelength [nm] | 1206 | 833 | 995 | 595 |
| | 1202 | 1201 | 365 | 365 |
| | 1207 | 922 | 355 | 1615 |
| | 1201 | 1202 | 1385 | 1655 |
| | 1204 | 831 | 1615 | 1285 |
| VarNC | 2385 | 2382 | 2485 | 2485 |
| wavelength [nm] | 2386 | 1974 | 615 | 615 |
| | 2382 | 2108 | 1115 | 1115 |
| | 2368 | 2327 | 795 | 855 |
| | 2490 | 2381 | 1305 | 2045 |
| | 2383 | 1960 | 2085 | 1815 |

## Bibliography


[1] Huynh, C. and A. Robles-Kelly, "Hyperspectral imaging for skin recognition and biometrics," in *IEEE 17th Int. Conf. Image Process.*, Hong Kong, China, Sept. 2010, pp. 2325-2338.

[2] Kamshilin, A., S. Miridonov, V. Teplov, R. Saarenheimo, and E. Nippolainen, "Photoplethysmographic imaging of high spatial resolution," *Biomedical Optics Express*, vol. 2, no. 4, pp. 996-1006, Apr. 2011.

[3] Yuen, P., T. Chen, K. Hong, A. Tsitiridis, F. Kam, J. Jackman, D. James, M. Richardson, L. Williams, W. Oxford, J. Piper, F. Thomas, and S. Lightman, "Remote detection of stress using hyperspectral imaging technique," *ICDP Presentation Crime Detection and Prevention*, Dec. 2009.

[4] Yuen, P., K. Hong, T. Chen, A. Tsitiridis, F. Kam, J. Jackman, D. James, M. Richardson, L. Williams, and W. Oxford, "Emotional and physical stress detection and classification using thermal imaging technique," *ICDP Presentation Crime Detection and Prevention*, Sept. 2009.

[5] *The U.S. Air Force-Developed Adaptation of the Multi-Attribute Task Battery for the Assessment of Human Operator Workload and Strategic Behavior*, AFRL, WPAFB, OH, 2010.

[6] Gabriel, P. "Mental health in the workplace: situation analyses, United States," *GLADNET Collection*, vol. 228, Jan. 2000.

[7] Nunez, A. and M. J. Mendenhall, "Detection of human skin in near infrared hyperspectral imagery," *IGARSS*, pp. 621-624, 2008.

[8] Fiedler, K., J. Schmid, and T. Stahl, "What is the current truth about polygraph lie detection?" *Basic and Applied Social Psychology*, vol. 24, no. 4, pp. 313-324, 2002.

[9] Schaefer, K. and H. Alvis, "The effect of inhalation of low oxygen concentration (10.5% $O_2$ in $N_2$) over a period of 33 minutes on respiration, pulse rate, arterial oxygen saturation (oximeter) and oxygen uptake," *Medical Research Laboratory Report*, vol. 10, pp. 76-97, 1951.

[10] Owens, J., C. Stoney, and K. Matthews, "Menopausal status influences ambulatory blood pressure levels and blood pressure changes during mental stress," *Circulation*, vol. 88, pp. 2794-2802, 1993.

[11] Pan, Z., G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552-1560, Dec. 2003.

[12] *FieldSpec ®User Manual*, ASD Inc., CO, 2010.

[13] Socolinsky, D., L. Wolff, J. Neuheisel, and C. Eveland, "Illumination invariant face recognition using thermal infrared imagery," in *Proc. 2001 Comp. Society Conf. Comp. Visual Pattern Recog.*, vol. 1, 2001.

[14] Mansor, S., A. Cracknell, B. Shilin, and V. Gornyi, "Monitoring of underground coal fires using thermal infrared data," *International Journal of Remote Sensing*, vol. 15, no. 8, pp. 1675-1685, 1994.

[15] Kimura, M. and K. Udagawa, "Thermal infrared sensors, imaging devices, and manufacturing methods for such sensors," U.S. Patent 6 034 374, March 7, 2000.

[16] Tardy, Y., J. Meister, F. Perret, H. Brunner, and M. Arditi, "Non-invasive estimate of the mechanical properties of peripheral arteries from ultrasonic and photoplethysmo-

graphic measurements," *Clinical Physics and Physiological Measurement*, vol. 12, no. 1, pp. 39, 2000.

[17] Allen, J. "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, pp. R1, 2007.

[18] Robnik-Sikonja, M. and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vo. 53, no. 1-2, pp. 23-69, Oct. 2003.

[19] Huang, Y., P. McCullagh, and N. Black, "Feature selection via supervised model construction," in *Proc. 4th IEEE Int. Conf. Data Mining*, Nov. 2004.

[20] Dash, M. and H. Liu, "Feature selection for classification," *Intelligent Data Analysis I*, pp. 131-156, 1997.

[21] W. Hsu, C. Chang, and J. Lin, "A practical guide to support vector classification," National Taiwan University, Taiwan, Tech. Rep., 2003.

[22] Pavlidis, P., I. Wapinski, and W. Noble, "Support vector machine classification on the web," *Bioinformatics*, vol. 20, no. 4, pp. 586-587, Oct. 2003.

[23] Rish, I. "An empirical study of the naive Bayes classifier," *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3, no. 22, pp. 41-46, 2001.

[24] Fomby, T., "Naive Bayes classifier," Apr. 2008, unpublished.

[25] Zhang, H., "The optimality of naive Bayes," *Authorship Attribution*, vol. 1, no. 2, 2004.

[26] Scholkopf, B. and A. Smola, "Support vector classification," in *Learning with Kernels*, Boston: MIT, 2002, ch. 1, sec. 1.6, pp. 15-17.

[27] Scholkopf, B. and A. Smola, "Examples and properties of kernels," in *Learning with Kernels*, Boston: MIT, 2002, ch. 2, sec. 2.3, pp. 45-48.

[28] Scholkopf, B. and A. Smola, "Pattern recognition," in *Learning with Kernels*, Boston: MIT, 2002, ch. 7, sec. 7.1-7.5, pp. 189-210.

[29] Cristianini, N. and J. Shawe-Taylor, "Support vector classification," in *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge, UK: Cambridge University Press, 2000, ch. 6, sec. 6.1, pp. 93-112.

[30] Friedl, M. and C. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 399-409, 1997.

[31] Rastogi, R. and K. Shim, "Public: A decision tree classifier that integrates building and pruning," *Data Minding and Discovery*, vol. 4, 2000, pp. 315-344.

[32] Kohavi, R. "Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid," *Proc. 2nd Int. Conf. Knowl. Disc. Data Mining*, 1996.

[33] Safavian, S. "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man. Cybern.*, vol. 21, no. 3, pp. 660-674, 1991.

[34] Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The WEKA Data Mining Software: An Update; SIGKDD Explorations, vol. 11, no. 1, 2009.

[35] Clark, J. "Distributed spacing stochastic feature selection and its application to textile classification," Ph.D. dissertation, AFIT, Dayton, OH, 2005.

[36] Cui, W., L. Ostrander, and B. Lee, "In vivo reflectance of blood and tissue as a function of light wavelength," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 6, pp. 632-639, 1990.

[37] Corral, L., G. Paez, and M. Strojnik, "Optimal wavelength selection for non-contact reflection photoplethysmography," *Proc. of SPIE*, vol. 8011, 2011.

[38] Beisley, A. “Spectral detection of human skin in VIS-SWIR hyperspectral imagery without radiometric calibration,” M.S. thesis, AFIT, Dayton, OH, 2012.

[39] Akbari, H., Y. Kosugi, K. Kojima, and N. Tanaka, “Blood vessel detection and artery-vein differentiation using hyperspectral imaging,” *IEEE Conf. Eng. Med. Biol. Soc.*, 2009.

[40] Di, W., L. Zhang, D. Zhang, and Q. Pan, “Studies on hyperspectral face recognition in visible spectrum with feature band selection,” *IEEE Trans. Syst., Man Cybern.*, vol. 40, no. 6, pp. 1354-1361, Nov. 2010.

[41] Zhang, D. and Z. Zhou, “$(2D)^2$PCA: 2-directional 2-dimensional PCA for efficient face representation and recognition,” *Neurocomputing*, vol. 69, no. 1-3, pp. 224-231, 2005.

[42] Freeman, J., F. Downs, L. Marcucci, E. Lewis, B. Blume, and J. Rish, “Multispectral and hyperspectral imaging: Applications for medical and surgical diagnostics,” *IEEE Conf. Eng. Med. Biol. Soc.*, vol. 2, pp. 700-701, Oct. 1997.

[43] Yuen, P. and M. Richardson, “An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition,” *Imaging Science Journal*, vol. 58, no. 5, pp. 241-253, 2010.

[44] Zuzak, K., M. Gladwin, R. Cannon III, and I. Levin, “Imaging hemoglobin oxygen saturation in sickle cell disease patients using noninvasive visible reflectance hyperspectral techniques: Effects of nitric oxide,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 285, no. 3, pp. H1183-H1189, 2003.

[45] Haykin, S. “Support Vector Machines,” in *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River: Pearson Ed., Inc., 2009, ch. 6, pp. 268-283.

[46] Gabrilovich, E. and S. Markovitch,"Feature generation for text categorization using world knowledge," *Int. Joint Conf. on Artificial Intell.*, vol. 19, 2005.

[47] Getoor, L. and B. Taskar, eds., "Dynamic Feature Generation," in *Introduction to Statistical Relational Learning*, Boston: MIT press, 2007, ch. 16, sec. 3, pp. 469-471.

[48] Miller, I., J. Freund, and R. Johnson, "Measures of Variation (or Spread)," in *Probability and Statistics for Engineers*, 4th ed., Englewood Cliffs, NJ: Prentice-Hall, 1965, ch. 1, sec. 4.3, pp 29-32.

[49] Clark, J. "Feature generation, feature selection, and classification," class notes for CSCE 623, Department of Computer Science and Engineering, Air Force Institute of Technology, Jun. 2012.

[50] O'Kane, M., J. Gillis, P. Rose, and M. Wagner, "Deciphering speech waveforms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 11, pp. 227-2230, Apr. 1986.

[51] Creates, M. and J. Grayson, "The effect of adrenaline and noradrenaline on coronary vascular reserve in the dog, *Journal of Physiology*, vol. 309, no. 1, pp. 557-568, 1980.

[52] Kim, M., Y. Chen, and P. Mehl, "Hyperspectral reflectance and fluorescence imaging system for food quality and safety, *Trans. ASAE*, vol. 44, no. 3, May/Jun. 2001.

[53] Gowen, A., C. ODonnell, P. Cullen, G. Downey, and J. Frias, "Hyperspectral imaging-an emerging process analytical tool for food quality and safety control, *Trends in Food Science & Technology*, vol. 18, no. 12, pp. 590-598, 2007.

[54] Ariana, D., R. Lu, and D. Guyer, "Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers, *Computers & Electronics in Agriculture*, vol. 53, no. 1, pp. 60-70, 2006.

[55] Qiao, J., M. Ngadi, N. Wang, C. Gariepy, and S. Prasher, "Pork quality and marbling level assessment using a hyperspectral imaging system, *Journal of Food Eng.*, vol. 83, no. 1, pp. 10-16, 2007.

[56] Khoobehi, B., J. Beach, and H. Kawano, "Hyperspectral imaging for measurement of oxygen saturation in the optic nerve head, *Investigative Ophthalmology & Visual Science*, vol. 45, no. 5, pp. 1464-1472, 2004.

[57] Beach, J., J. Ning, and B. Khoobehi, "Oxygen saturation in optic nerve head structures by hyperspectral image analysis,*Current Eye Research*, vol. 32, no. 2, pp. 161-170, 2007.

[58] Vo-Dinh, T., F. Yan, and M. Wabuyele, "Surface-enhanced Raman scattering for medical diagnostics and biological imaging, *Journal of Raman Spectroscopy*, vol. 36, no. 6-7, pp. 640-647, 2005.

[59] Panasyuk, S., S. Yang, D. Faller, D. Ngo, R. Lew, J. Freeman, and A. Rogers, "Medical hyperspectral imaging to facilitate residual tumor identification during surgery, *Cancer Biology & Therapy*, vol. 6, no. 3, pp. 439-446, 2007.

[60] Akbari, H., Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging, *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 2011-2017, 2010.

[61] Randeberg, L., E. Larsen, and L. Svaasand, "Characterization of vascular structures and skin bruises using hyperspectral imaging, image analysis and diffusion theory, *Journal of Biophotonics*, vol. 3, no. 1-2, pp. 53-65, 2010.

[62] Lewis, E. and J. Freeman, "Multispectral/hyperspectral medical instrument, U.S. Patent 6 937 885, Aug. 30, 2005.

[63] Wang, Q. and C. Suen, "Analysis and design of a decision tree based on entropy reduction and its application to large character set recognition, *IEEE Trans. Pattern Analy. Mach. Intell.*, vol. 4, pp. 406-417, 1984.

[64] Shao, L., A. Hero, W. Rogers, and N. Clinthorne, "Information gain from count corrections in SPECT image reconstruction and classification, *IEEE Trans. Nucl. Sci.*, vol. 37, no. 2, pp. 652-657, 1990.

[65] McKinney, M., M. Miner, H. Ruddel, H. Mcllvain, H. Witte, J. Buell, R. Eliot, and L. Grant, "The standardized mental stress test protocol: Test-retest reliability and comparison with ambulatory blood pressure monitoring," *Psychophysiology*, vol. 22, no. 4, pp. 453-463, 1985.

[66] Nickel, P. and F. Nachreiner, "Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 4, pp. 575-590, 2003.

[67] Egelund, N., "Spectral analysis of heart rate variability as an indicator of driver fatigue," *Ergonomics*, vol. 25, no. 7, pp. 663-672, 1982.

[68] Malik, M., J. Bigger, A. Camm, R. Kleiger, and A. Malliani, "Heart rate variability," *Circulation*, vol. 93, no. 5, pp. 1043-1065, 1996.

[69] Stauss, H., "Heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 285, no. 5, pp. R927-R931, 2003.

[70] Hirsch, J. and B. Bishop, "Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 241, no. 4, pp. H620-H629, 1981.

[71] Meshkati, N., "Heart rate variability and mental workload assessment," *Advances in Psychology*, vol. 52, pp. 101-115, 1988.

[72] Romito, K. and B. D. O'Brien. (2010, Apr. 5). *Pulse Measurement* [Online]. Available: http://www.webmd.com

[73] Dietzek, C. *Vein Problems* [Online]. Available: http://www.veinvascular.com/vvi-artery-problems.html

[74] Carden, K., "Recording sleep: the electrodes, 1020 recording system, and leep system specifications," *Sleep Medicine Clinics*, vol. 4, no. 3, pp. 333-341, Sept. 2009.

[75] *Field Guide for the ASD FieldSpec Pro-White Reference Mode*, 2nd ed., 2007.

[76] Jain, A. and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153-158, Feb. 1997.

[77] Clark, J., M. Mendenhall, and G. Peterson, "Stochastic feature selection with distributed spacing for hyperspectral data," *IEEE 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1-4, June 2010.

[78] Saladin, K., "The circulatory system: blood vessels and circulation," in *Anatomy & Physiology*, 4th ed. New York: McGraw-Hill, 2007, ch. 20, pp. 754-800.

[79] Krejza, J., M. Arkuszewski, and S. Kasner, "Carotid artery diameter in men and women and the relation to body and neck size," *Stroke*, vol. 27, pp. 1103-1105, 2006.

[80] *BIOPAC Systems #AH-109: 3-, 6-, and 12-Lead ECG*, BIOPAC Syst., Inc., CA, 2005.

[81] Blum, A. and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1, pp. 245-271, 1997.

[82] Splawn, J., "Applying Hyperspectral Imaging to Heart Rate Estimation for Adaptive Automation," M.S. thesis, AFIT, Dayton, OH, 2013.

[83] Friebel, M. and M. Meinke, "Model function to calculate the refractive index of native hemoglobin in the wavelength range of 250-1100 nm dependent on concentration," *Applied optics*, vol. 45, no. 12, pp. 2838-2842, 2006.

[84] Gussakovsky, E., O. Jilkina, Y. Yang, and V. Kupriyanov, "Hemoglobin plus myoglobin concentrations and near infrared light pathlength in phantom and pig hearts determined by diffuse reflectance spectroscopy," *Analytical biochemistry*, vol. 382, no. 2, pp. 107-115, June 2008.

[85] Kostenich, G., S. Kimel, S. Peled, and A. Orenstein, "Monitoring PDT-induced damage using spectrally resolved reflectance imaging if tissue oxygenation," *Cancer letters*, vol. 219, no. 2, pp. 169-175, Mar. 2005.

[86] Bouckaert, R., E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, "WEKA Manual for Version 3-7-4," 2011.

[87] Kirkby, R., E. Frank, and P. Reutemann, "WEKA Explorer User Guide for Version 3-5-5," *Relation* 10, no. 1.35, 2007.

[88] McNeil, B. and J. Hanley, "Statistical approaches to the analysis of receiver operating characteristic (ROC) curves," *Medical decision making: an international journal of the Society for Medical Decision Making*, vol. 4, no. 2, pp. 137-150, 1984.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704–0188*



| 1. REPORT DATE *(DD–MM–YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From — To)* |
|---|---|---|
| 21–03–2013 | Master's Thesis | Oct 2012–Mar 2013 |

**4. TITLE AND SUBTITLE**

Spectral Detection of Acute Mental Stress with VIS-SWIR Hyperspectral Imagery

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Norvell, Elizabeth K., Second Lieutenant, USAF

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

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

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

**8. PERFORMING ORGANIZATION REPORT NUMBER**

AFIT-ENG-13-M-38

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

Justin Estepp 711 HPW/RHCPA BLDG 840, W200 2510 Fifth Street
Wright-Patterson AFB, OH 45433-7951 Phone: (937) 938-3602 / DSN: 798-3602
Fax: (937) 656-6894 e-mail: justin.estepp@wpafb.af.mil

**10. SPONSOR/MONITOR'S ACRONYM(S)**

711 HPW/RHCPA

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

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

**13. SUPPLEMENTARY NOTES**

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

**14. ABSTRACT**

THE ability to identify a stressed person is becoming an important aspect across different work environments. Especially in higher-stress career fields, such as first responders and air traffic controllers, mental stress can inhibit a person's ability to accomplish their job. A person's efficiency and psychological state in the work environment can be impeded due to poor mental health. Stress can result in harmful effects on the body, both physically and mentally, including depression, lack of sleep, and fatigue, which can lead to reduced work productivity.

Research is being conducted to detect stress in workload-intensive environments. This thesis implements an imaging approach that utilizes hyperspectral data across the visible through shortwave infrared electromagnetic spectrum. The data is applied to the feature selection algorithms ReliefF, Support Vector Machine Attribute Evaluator (SVM AE), and Non-Correlated Aided Simulated Annealing Feature Selection-Integrated Distribution Function (NASAFS-IDF) to obtain features that discriminate between the classes, "stress" and "non-stress." This data is classified using naive Bayes, Support Vector Machine (SVM), and decision tree methodologies.

The feature set and classifier that produce the highest classification results are calculated using percent accuracy and area under the curve (AUC). The reported results are divided into contact and non-contact (NC) validation sets. The contact validation returned a high accuracy of 96.30% and high AUC of 0.979. Validation on NC models returned a high accuracy of 99.64% and high AUC of 0.998.

**15. SUBJECT TERMS**

hyperspectral, mental stress, stress detection, non-invasive

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | (ENG) |
| U | U | U | UU | 149 | 19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 ext. 4614 |