

## **Chapter 5:**

# **Electronic Nose Breathprint Discrimination in Collegiate Long Distance Runners**

(Manuscript is being prepared for submission to Biological Engineering Transactions)

**THIS PAGE IS FOR INDEXING PURPOSES AND WILL NOT BE PRINTED.**

For additional details please see the [ASABE Guide for Authors](#).

**Author 1**

First Name or initial	Middle Name or initial	Surname	Role (ASABE member, professor, etc.)	E-mail (and phone for contact author)	Contact author? yes or no
Christan		Whysong	ASABE Member, Graduate Research Assistant	<a href="mailto:cwhysong@vt.edu">cwhysong@vt.edu</a> (540/231-6615)	Yes

**Affiliation**

Organization	Address	Country	URL or other info.
Biological Systems Engineering, Virginia Tech	Department of Biological Systems Engineering (MC0303) Seitz Hall, RM 200, Virginia Tech 155 Ag Quad Lane Blacksburg, VA 24061	USA	<a href="http://www.bse.vt.edu/">http://www.bse.vt.edu/</a>

**Author 2**

First Name or initial	Middle Name or initial	Surname	Role (ASABE member, professor, etc.)	E-mail (and phone for contact author)	Contact author? yes or no
Parameswarakumar		Mallikarjunan	ASABE Member, Associate Professor	<a href="mailto:kumar@vt.edu">kumar@vt.edu</a>	No

**Affiliation**

Organization	Address	Country	URL or other info.
Biological Systems Engineering, Virginia Tech	Department of Biological Systems Engineering (MC0303) Seitz Hall, RM 308, Virginia Tech 155 Ag Quad Lane Blacksburg, VA 24061	USA	<a href="http://www.bse.vt.edu/">http://www.bse.vt.edu/</a>

**Author 3**

First Name or initial	Middle Name or initial	Surname	Role (ASABE member, professor, etc.)	E-mail (and phone for contact author)	Contact author? yes or no
Per	Gunnar	Brolinson	Associate Dean for Clinical Research	<a href="mailto:pbrolins@vcom.vt.edu">pbrolins@vcom.vt.edu</a>	No

**Affiliation**

Organization	Address	Country	URL or other info.
Edward Via College of Osteopathic Medicine	Edward Via College of Osteopathic Medicine	USA	<a href="http://www.vcom.vt.edu/index.html">http://www.vcom.vt.edu/index.html</a>

	309 N Knollwood Drive Blacksburg, VA 24060		
--	---	--	--

#### Author 4

First Name or initial	Middle Name or initial	Surname	Role (ASABE member, professor, etc.)	E-mail (and phone for contact author)	Contact author? yes or no
Hara	P	Misra	Professor Emeritus	<a href="mailto:hmisra@vcom.vt.edu">hmisra@vcom.vt.edu</a>	No

#### Affiliation

Organization	Address	Country	URL or other info.
Edward Via College of Osteopathic Medicine	Edward Via College of Osteopathic Medicine 2265 Kraft Drive Blacksburg, VA 24060	USA	<a href="http://www.vcom.vt.edu/index.html">http://www.vcom.vt.edu/index.html</a>

#### Author 5

First Name or initial	Middle Name or initial	Surname	Role (ASABE member, professor, etc.)	E-mail (and phone for contact author)	Contact author? yes or no
Zhenquan		Jia	Assistant Professor	<a href="mailto:z_jia@uncg.edu">z_jia@uncg.edu</a>	No

#### Affiliation

Organization	Address	Country	URL or other info.
Department of Biology	Department of Biology, College of Arts and Science The University of North Carolina at Greensboro 321 McIver Street, 312 Eberhart Building Greensboro, NC 27412	USA	<a href="http://biology.uncg.edu/faculty/Zhenquan_Jia/">http://biology.uncg.edu/faculty/Zhenquan_Jia/</a>

# ELECTRONIC NOSE BREATHPRINT DISCRIMINATION IN COLLEGIATE LONG DISTANCE RUNNERS

C. Whysong, P. Mallikarjunan, P. G. Brolinson, H. P. Misra, Z. Jia

The authors are **Christan Whysong, ASABE Member**, Graduate Student Research Assistant, Biological Systems Engineering Department, Virginia Tech, Blacksburg, Virginia, USA; **Parameswarakumar Mallikarjunan, ASABE Member**, Associate Professor, Biological Systems Engineering Department, Virginia Tech, Blacksburg, Virginia, USA; **Per Gunnar Brolinson**, Associate Dean for Clinical Research, Edward Via College of Osteopathic Medicine, Blacksburg, Virginia, USA; **Hara P. Misra**, Professor Emeritus, Edward Via College of Osteopathic Medicine, Blacksburg, Virginia, USA; and **Zhenquan Jia**, Assistant Professor, Department of Biology, University of North Carolina at Greensboro, Greensboro, North Carolina. **Corresponding author:** Christan Whysong, Department of Biological Systems Engineering (MC0303), Seitz Hall, RM 200, Virginia Tech, 155 Ag Quad Lane, Blacksburg, Virginia, 24061; 540-231-6615; cwhysong@vt.edu.

## ABSTRACT.

*Competitive athletes strive to improve performance through a balance of training loads and rest periods. Overtraining occurs when this balance favors intense training loads and can lead to early retirement. A Cyranose<sup>®</sup> 320, a conducting polymer-based electronic nose (enose,) was proposed as a rapid, noninvasive, point-of-care device for clinically diagnosing the onset of overtraining. Athletes actively involved in a collegiate long distance running program were observed throughout a competitive training season. The effects of two acute (one low intensity and one high intensity) and cumulative training loads on the volatile organic compounds in the breath, or breathprints, of the athletes were observed through the enose. The results from thirty-one data standardization and preprocessing methods and linear discriminant analyses were compared. It was concluded the enose can discriminate between breathprints of collegiate long distance runners after cumulative and acute training loads. An external statistical model standardizing data with the baseline (Pre-Study) breath samples allowed the Cyranose<sup>®</sup> 320 to correctly classify 96% of the breathprints. Future work should focus on developing a larger breathprint library to improve the prediction model.*

**Keywords.** *electronic nose, linear discriminant analysis, overtraining, runners, standardization*

## INTRODUCTION

Ancient medical practitioners used smells of body secretions to select treatments for the infirmed (Pavlou & Turner, 2000). Sweet-smelling sweat has been linked to diphtheria, (Rock et al., 2008) skin that smells like freshly baked brown bread has been associated with typhoid (Rock et al., 2008), and breath smelling like acetone has been found in diabetic patients (Wang et al., 1997). Smell had largely been lost in modern medicine because the physician's nose lacks precision and subjectivity and is not able to quantify data (Hockstein et al., 2004; Thaler, 2002).

Electronic nose (enose) technology was developed to mimic the human nose and has recently brought the use of smell back to medicine. The main component of an electronic nose is an array of sensors, each with a different specificity to a range of volatile molecules (Nagle et al., 1998). To analyze a sample, a vacuum pulls the sample air consisting of volatiles (headspace) into the enose, the headspace passes over the sensor array, each sensor reacts to the volatiles, each sensors' response is recorded, and the headspace gas is purged from the unit (Nagle et al., 1998). Unlike other analytical methods that identify specific volatiles in samples, enoses use the pattern of each sensors' response, or smellprint, is to identify samples in their entirety.

Enoses have previously detected colorectal cancer in fecal gas (de Meij et al., 2014), infected root canals (Yamada et al., 2007), renal failure in skin (Voss et al., 2005), *S. aureus* infections in ear, nose, and throat swabs (Dutta et al., 2005), and urinary tract infections in urine (Kodogiannis et al., 2008). Many metabolic processes produce volatile organic compounds (VOC) that can be found in breath (Fens et al., 2013). Enoses have been able to detect these breath VOCs in humans with asthma (Dragonieri et al., 2007), uremia (Lin et al., 2001), pneumonia (Hockstein et al., 2004), lung cancer (Di Natale et al., 2003), asbestos-related diseases (Chapman et al., 2012), and diabetes (Wang et al., 1997).

The production of Reactive oxygen species (ROS) increases with increased oxygen metabolism such as that which occurs with exercise. The negative effects of ROS are mitigated by an elaborate antioxidant system. Oxidative stress occurs when the antioxidant system cannot maintain homeostasis within the body and has been suggested as a cause of overtraining in athletes (Finaud et al., 2006). Oxidative stress leads to lipid peroxidation which can be detected through pentane, hexane, and ethane in expired air (Finaud et al., 2006).

Overtraining describes an athlete-specific syndrome characterized as a long term persistent inability to perform at expected optimums even though intense training is still incurred (Armstrong & VanHeest, 2002; Morgan et al., 1987). It can lead to increased illnesses (Lakier Smith, 2003), lingering fatigue (Booth et al., 2006), mood disturbances (Anglem et al., 2008), and tissue inflammation (Margonis et al., 2007). Recovery consists of complete rest from the sport until complete recovery which can take months to years (Halsen & Jeukendrup, 2004) and in some cases forces an early retirement (Peluso & Guerra de Andrade, 2005).

After reviewing 152 overtraining and overreaching studies in endurance-based exercise, Bell and Ingle (2013) found many overtraining biomarkers have been proposed but no one single clinically-specific diagnostic marker has been identified. Many of the proposed markers require invasive testing and lengthy result turnaround times, leaving decreased athletic performance as the only reliable overtraining predictor (Margonis et al., 2007). However, even the most experienced coaches admit they are not able to predict which athletes will overtrain (Armstrong & VanHeest, 2002). Athletic staffs (athletes, coaches, trainers, and physicians) would benefit from a clear, quick, and simple method for clinically diagnosing overtraining.

It is hypothesized distinct levels of physical training loads will provide distinct patterns of VOCs in the breath, or breathprints, of athletes which are detectable with enose technology. A pilot study found a conducting polymer-based enose is able to detect differences in breathprints after training loads but improvements in the prediction model are required (Whysong et al., 2014). This paper will focus on data manipulation, or standardization, techniques to improve the prediction model.

## **MATERIALS AND METHODS**

### **STUDY OVERSIGHT**

Institutional Review Boards for the Edward Via College of Osteopathic Medicine (VCOM) and Virginia Tech, both located in Blacksburg, Virginia, approved (Appendix C) the study protocol (Appendix D). Subjects were informed of the study's purpose and potential risks before providing written consent (Appendix E). Dr. Brolinson, the lead physician, ensured subjects' eligibility prior to participation. For confidentiality, biological samples were collected into containers marked only with subjects' assigned study numbers. Subjects were not compensated for their participation in the study.

### **SUBJECTS AND TRAINING LOADS**

Nine athletes actively involved in a collegiate long distance running program served as subjects in this preliminary study. Each athlete was observed multiple times throughout one of three participation terms: Summer 2008, Spring 2010, and Fall 2010. Each subject completed a training questionnaire (Appendix G) prior to providing any other data to help ensure study eligibility and better understand health and training history during the 30 days prior to enrolling in the study. Summer 2008 subjects had an average age of 33.5 (minimum 28; maximum 39) years, average height of 172.7 (minimum 165; maximum 180) cm, and average weight of 63.6 (minimum 52; maximum 75) kg. Spring 2010 and Fall 2010 subjects had an average age of 20.1 (minimum 18; maximum 21) years, average height of 175.6 (minimum 160; maximum 188) cm, and average weight of 65.8 (minimum 57; maximum 74) kg. All subjects indicated they never smoke. Subjects indicated they drink alcohol between never and occasionally (three to five times per week) in and out of the training season. Eight subjects

identified as Caucasian while one subject identified as “Caucasian/Asian.” Some subjects indicated they had taken allergy, birth control, and pain relief medication and nutritional supplements such as iron, multivitamins, Omega-3, and Vitamin C. No subjects reported an injury but three subjects reported having a minor cold within the 30 days prior to the study. Subjects indicated they ran between 72 and 97 km per week and completed two to four strength training sessions per week prior to the study. Other training activities consisted of biking and swimming. Appendix H provides more details on subject health and training activities 30 days prior to enrolling in the study.

A common problem with longitudinal studies is subject dropout for all or part of the study and this study was no different. Reasons included aversion to having blood drawn, injury, forgetting to return for sample collection, and schedule conflicts. Table 1 summarizes subject participation and collected samples.

The effects of cumulative training loads over a season were observed by collecting baseline breath samples at the beginning of the training season (Pre-Study) and at the end of the training season (Post-Study). Immediate effects of training were observed for two training loads, one low intensity (LI) and one high intensity (HI). Athletes were observed during a real training setting. Training loads were not directly manipulated for experimental purposes. The research team worked closely with the athletes and their coaches to identify sample collection times that coincided with the desired training loads. Example LI training loads included a short run of 5 km (30 min), a LI bike workout (20 min), and a LI run (50 min). Examples of HI training loads included a 13 km hard run and a long 15 km run (80 min).

Subjects came to the clinic to provide a breath sample before the LI short run (BSR) and HI long run (BLR), completed the training session, immediately provided a second breath sample in the field (ASRField; ALRField), and then returned to the clinic to provide a third breath sample (ASRClinic; ALRClinic). The minimum time between the Field and Clinic collections was 10 min and the maximum was 25 min. Pre-Study and Post-Study samples were not collected during the Summer 2008 participation term so these subjects’ BSR samples served as their baseline. Each subject served as his or her own control.

**Table 1. Nine athletes provided breath samples during one of three participation terms.**

Term	Subject	Average Age (years)	Pre-Study	BSR	ASRField	ASRClinic	BLR	ALRField	ALRClinic	Post-Study
Summer 2008	Male-1	33.5	None	Breath	Breath	Breath	None			
	Female-1		None	Breath	Breath	Breath	Breath	Breath	Breath	None
Spring 2010	Male-2	20	Breath	None						
	Female-2		Breath	None						
Fall 2010	Male-3	20.2	Breath	None						
	Male-4		Breath	Breath	Breath	Breath	Breath	Breath	None	Breath
	Female-3		Breath	Breath	Breath	Breath	Breath	None	Breath	Breath
	Female-4		Breath	Breath	Breath	Breath	None			
	Female-5		Breath	Breath	Breath	Breath	Breath	None	Breath	Breath

## BREATH COLLECTION AND ANALYSIS

Breath samples were collected into an alveolar air collection device (GaSampler System, QuinTron Instrument Company, Inc., Milwaukee, WI). Summer 2008 samples were analyzed directly from the collection bags within 10 h of collection. The remaining samples were transferred from the collection bags to storage bags through a sample flush drying tube. Transfer occurred between 30 min and 2.5 h after collection. Samples were analyzed from the storage bags between five and 13 days after collection. All samples in a class (e.g. Pre-Study) were analyzed before moving to the next one.

A Cyranose<sup>®</sup> 320 (C320) enose with an array of 32 conducting polymer-based sensors (Cyranose<sup>®</sup> 320, Sensigent, Baldwin Park, California) analyzed the breath samples. PCnose<sup>®</sup>, (PCnose<sup>®</sup>, Sensigent, Baldwin Park, California) C320's software package, was utilized to optimize unit settings, access datasets, and monitor sensor response during sample analysis. Prior to analysis, a series of checks (Appendix B) were performed on the C320 to ensure proper function. During breath analysis, the snout of the C320 was inserted into a stopcock in the luer port of the sample bag (fig. 1a). The stopcock was opened immediately prior to the C320 drawing in the breath sample.

During the Fall 2010 participation term double the breath samples were collected for the BLR, ALRField, ALRClinic, and Post-Study collections. These extra samples were used to test a second method of introducing the sample to the enose. A series of tubes connected the stopcock inserted into the luer port to a tightfitting tube which fit snugly over the enose snout (fig. 1b).



Figure 1. The C320 analyzed breath samples by inserting the snout directly into the storage bag stopcock (a) or by connecting the snout to the storage bag through a series of tubes (b).

## CARBON DIOXIDE COLLECTION AND ANALYSIS

A significant component of exhaled breath is carbon dioxide (Fens et al., 2013) and was explored as a potential method for standardizing breathprint data collected with the C320. Small pieces of dry ice (Penguin Brand™, Airgas, Inc.) were placed into small plastic containers that had two small holes in the lid: one for the enose snout and one to provide ventilation to prevent extreme gas buildup. The dry ice remained in the containers for at least ten minutes before the snout of the C320 was inserted into the lid and the headspace was analyzed (fig. 2).

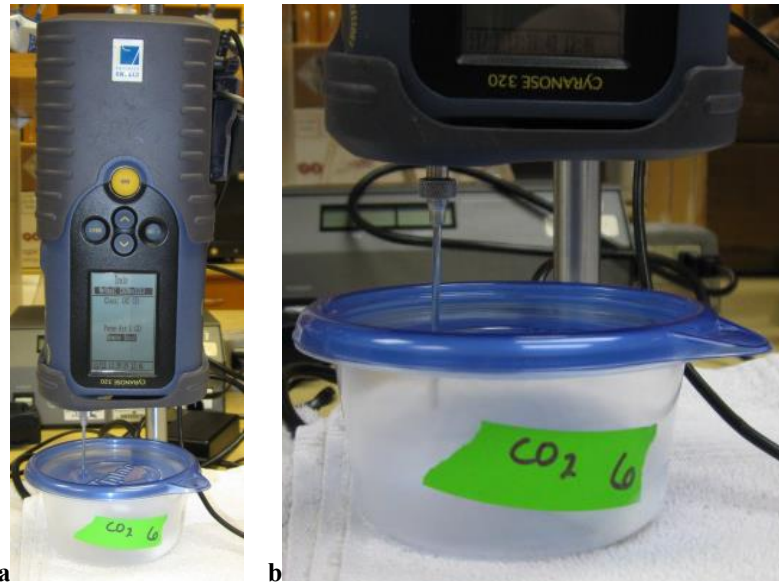


Figure 2. Carbon dioxide smellprints were collected by inserting the C320 snout into a plastic container containing a small piece of dry ice.

## STATISTICAL ANALYSIS

A pilot study found PCnose® was insufficient in analyzing similar breath samples (Whysong et al., 2014) so all statistical analyses were completed externally using JMP® (SAS Institute, Inc., Cary, North Carolina). JMP®'s Discriminant platform was used to complete many linear discriminant analyses (LDA) to compare data standardization techniques as described below (DA1 through DA21). Later three methods of data reduction were observed to see if they could improve LDA results (DA22 through DA27). Finally the “Tube” and “No Tube” apparatuses were compared through LDA results (DA28 through DA31).

Some standardization techniques utilized averaged data. When analyzed by the C320, each sample had a smellprint consisting of 32 recorded resistance values, one per sensor. Averages were taken over each sensor. For instance, data was collected from six carbon dioxide samples with the C320, each consisting of 32 recorded data points, for a total of 192 recorded data points for carbon dioxide. To standardize some data subsets, the average carbon dioxide resistance for the six samples was calculated for each sensor. These 32 averages, one per sensor, were then used in the relevant data standardizations prior to further statistical analysis, such as LDA. Some analyses were completed using average Pre-

Study, BSR, or BLR in the data standardization. These averages were calculated in the same way except an average per sensor per subject was calculated and then used in the standardizations, unless otherwise noted.

$$R_{S(c,n)} = \frac{R_{(c,n)}}{R_{CO2(n)}} \quad (1)$$

$$R_{S(c,n)} = R_{(c,n)} - R_{CO2(n)} \quad (2)$$

$$R_{S(c,n)} = \frac{R_{AR(c,n)} - R_{CO2(n)}}{R_{BR(c,n)}} \quad (3)$$

$$R_{S(c,n)} = \frac{R_{(c,n)} - R_{CO2(n)}}{R_{Pre(n)}} \quad (4)$$

$$R_{S(c,n)} = \frac{R_{AR(c,n)}}{R_{BR(c,n)}} \quad (5)$$

$$R_{S(c,n)} = \frac{R_{AR(c,n)}}{R_{BR(c,n)} \times R_{CO2(n)}} \quad (6)$$

$$R_{S(c,n)} = \frac{R_{Pre(n)}}{R_{CO2(n)}} \quad (7)$$

$$R_{S(c,n)} = \frac{R_{Post(n)}}{R_{CO2(n)}} \quad (8)$$

$$R_{S(c,n)} = \frac{R_{(c,n)}}{R_{Pre(n)}} \quad (9)$$

$$R_{S(c,n)} = \frac{R_{(c,n)}}{R_{Pre(n)} \times R_{CO2(n)}} \quad (10)$$

$$R_{S(c,n)} = R_{AR(c,n)} - R_{BR(c,n)} \quad (11)$$

$$R_{S(c,n)} = \frac{R_{AR(c,n)} - R_{BR(c,n)}}{R_{CO2(n)}} \quad (12)$$

$$R_{S(c,n)} = R_{AR(c,n)} - R_{BR(c,n)} - R_{CO2(n)} \quad (13)$$

$$R_{S(c,n)} = R_{Pre(n)} - R_{CO2(n)} \quad (14)$$

$$R_{S(c,n)} = R_{Post(n)} - R_{CO2(n)} \quad (15)$$

$$R_{S(c,n)} = R_{(c,n)} - R_{Pre(n)} \quad (16)$$

$$R_{S(c,n)} = \frac{R_{(c,n)} - R_{Pre(n)}}{R_{CO2(n)}} \quad (17)$$

$$R_{S(c,n)} = R_{(c,n)} - R_{Pre(n)} - R_{CO2(n)} \quad (18)$$

$$R_{S(c,n)} = \frac{R_{(c,n)} - \mu_{(c,n)}}{\sigma_{(c,n)}} \quad (19)$$

where:  $c$  = sample class (Pre-Study, BSR, ASRField, ASRClinic, BLR, ALRField, ALRClinic, or Post-Study)

$n$  = C320 sensor number (1 through 32)

$R_s$  = standardized resistance ( $k\Omega$ )

$R$  = raw resistance recorded by C320 ( $k\Omega$ )

$R_{CO_2}$  = average resistance for carbon dioxide samples ( $k\Omega$ )

$R_{AR}$  = raw resistance recorded by C320 for the after run class (ASRField, ASRClinic, ALRField, or ALRClinic) ( $k\Omega$ )

$R_{BR}$  = average resistance for the before run (BSR or BLR) samples ( $k\Omega$ )

$R_{Pre}$  = average resistance for the Pre-Study class ( $k\Omega$ )

$R_{Post}$  = average resistance for the Post-Study samples ( $k\Omega$ )

$\mu$  = average resistance for all subjects ( $k\Omega$ )

$\sigma$  = standard deviation for all subjects ( $k\Omega$ ).

#### ***Data Analysis 1 (DA1)***

DA1 was a LDA on all data collected from all participants during all participation periods. Raw data was utilized.

#### ***Data Analysis 2 (DA2)***

DA2 was a LDA on all data collected from all subjects during all participation periods. Each recorded data point was standardized using carbon dioxide according to equation 1.

#### ***Data Analysis 3 (DA3)***

DA3 was a LDA on all data collected from all subjects during all participation periods. Each recorded data point was standardized using carbon dioxide according to equation 2.

#### ***Data Analysis 4 (DA4)***

DA4 was a LDA on data collected from all subjects during all participation periods. Each recorded data point for the after run classes (ASRField, ASRClinic, ALRField, and ALRClinic) was standardized using the respective before run (BSR and BLR) average according to equation 3. The raw Pre-Study and Post-Study data were included in the LDA but the raw BSR and BLR classes were not.

#### ***Data Analysis 5 (DA5)***

DA5 was a LDA on only data collected from subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. The Pre-Study data was utilized to standardize each data point for the other classes according to equation 4. The Pre-Study class was not included in the LDA.

#### ***Data Analysis 6 (DA6)***

DA6 was a LDA on data collected from all subjects during all participation periods. Each recorded data point for the after run classes (ASRField, ASRClinic, ALRField, and ALRClinic) was standardized using the respective before run (BSR and BLR) average according to equation 5. The raw Pre-Study and Post-Study data were included in the LDA but the raw BSR and BLR classes were not.

#### ***Data Analysis 7 (DA7)***

DA7 was a LDA on data collected from all subjects during all participation periods. Each recorded data point for the after run classes (ASRField, ASRClinic, ALRField, and ALRClinic) was standardized using the respective after before

(BSR and BLR) average according to equation 6. Each Pre-Study and Post-Study data point was standardized using average carbon dioxide data according to equations 7 and 8, respectively.

***Data Analysis 8 (DA8)***

DA8 was a LDA on only data collected from subjects during the 2010 participation periods. The Pre-Study data was utilized to standardize each data point for the other classes according to equation 9. The raw Pre-Study data was not included in the LDA.

***Data Analysis 9 (DA9)***

DA9 was a LDA on only data collected from subjects during the 2010 participation periods. The Pre-Study and carbon dioxide data were used to standardize each data point for the other classes according to equation 10. The raw Pre-Study data was not included in the LDA.

***Data Analysis 10 (DA10)***

DA10 was a LDA on data collected from all subjects during all participation periods. Each recorded data point for the after run classes (ASRField, ASRClinic, ALRField, and ALRClinic) was standardized using the respective before run (BSR and BLR) average according to equation 11. The raw Pre-Study and Post-Study data were included in the LDA but the raw BSR and BLR classes were not.

***Data Analysis 11 (DA11)***

DA11 was a LDA on data collected from all subjects during all participation periods. Each recorded data point for the after run classes (ASRField, ASRClinic, ALRField, and ALRClinic) was standardized using the respective before run (BSR and BLR) and carbon dioxide data according to equation 12. The raw Pre-Study and Post-Study data were included in the LDA but the raw BSR and BLR classes were not.

***Data Analysis 12 (DA12)***

DA12 was a LDA on data collected from all subjects during all participation periods. Each recorded data point for the after run classes (ASRField, ASRClinic, ALRField, and ALRClinic) was standardized using the respective before run (BSR and BLR) data according to equation 13. Each Pre-Study and Post-Study data point was standardized using average carbon dioxide according to equations 14 and 15, respectively. The raw BSR and BLR classes were not included in the LDA.

***Data Analysis 13 (DA13)***

DA13 was a LDA on data only collected from subjects during the 2010 participation periods. The Pre-Study data was utilized to standardize each data point for the other classes according to equation 16. The raw Pre-Study data was not included in the LDA.

***Data Analysis 14 (DA14)***

DA14 was a LDA on only data collected from subjects during the 2010 participation periods. The Pre-Study and

carbon dioxide data were used to standardize each data point for the other classes according to equation 17. The raw Pre-Study data was not included in the LDA.

***Data Analysis 15 (DA15)***

DA15 was a LDA on only data collected from subjects during the 2010 participation periods. The Pre-Study and carbon dioxide data were utilized to standardize each data point for the other classes according to equation 18. The raw Pre-Study data was not included in the LDA.

***Data Analysis 16 (DA16)***

DA16 was a LDA on data collected from all subjects during all participation periods. The average and standard deviation were calculated for each class (Pre-Study, BSR, BSRField, BSRClinic, BLR, ALRField, ALRClinic, and Post-Study) for all subjects for each sensor. Each data point was then standardized using the relevant average and standard deviation according to equation 19.

***Data Analysis 17 (DA17)***

DA17 was a LDA on only data collected from subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. Raw data was utilized.

***Data Analysis 18 (DA18)***

DA18 was a LDA on data collected from all subjects in the 2010 participation periods. The Pre-Study data was used to standardize each data point for the other classes according to equation 16. The raw Pre-Study data was included in the LDA.

***Data Analysis 19 (DA19)***

DA19 was a LDA on only data collected from subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. The Pre-Study data was used to standardize each data point for the other classes according to equation 16. The raw Pre-Study data was included in the LDA.

***Data Analysis 20 (DA20)***

DA20 was a LDA on data collected from all subjects during all participation periods. The Pre-Study data was used to standardize each data point for the other classes according to equation 16. The raw Pre-Study data was included in the LDA.

***Data Analysis 21 (DA21)***

DA21 was a LDA on data collected from all subjects during the 2008 participation period and only the subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. The Pre-Study data was used to standardize each data point for the other classes according to equation 16. The raw Pre-Study data was included in the LDA.

***Data Analysis 22 (DA22) and Data Analysis 23 (DA23)***

JMP®'s Principal Components platform was used to complete two principal component analyses (PCA): DA22 used the same data from DA19; DA23 used the same data from DA21. The optimum number of principal components to represent the data was selected by choosing those with an eigenvalue close to one and accounting for 90% of the cumulative variation along with reviewing the scree plot. The optimum number of principal components were saved and then used in a LDA for each analysis (DA22 and DA23).

***Data Analysis 24 (DA24) and Data Analysis 25 (DA25)***

DA24 used the same data from DA19 while DA25 used the same data from DA21. The Stepwise Variable Selection function of JMP®'s Discriminant platform was utilized to evaluate which C320 sensors discriminate well. Through this function the user can review the p-values and F ratios (Inc., 2013) for each sensor and then review the LDA results after removing sensors from or adding sensors to the model.

***Data Analysis 26 (DA26) and Data Analysis 27 (DA27)***

JMP®'s Partial Least Squares platform was used to complete two partial least squares regression (PLS) analyses: DA26 used the same data from DA19; DA27 used the same data from DA21. The Nonlinear Iterative Partial Least Squares method was specified and the Leave-One-Out Cross Validation validation was utilized to determine the optimum number of latent variables to extract using the Root Mean PRESS. The Voet T<sup>2</sup> statistic was then utilized to compare models with different latent variables (Inc., 2013). The selected model had the smallest number of latent variables with a significance level above 0.10 (Inc., 2013). These latent variables were then saved and used in a LDA for DA26 and DA27.

***Data Analysis 28 (DA28) and Data Analysis 29 (DA29)***

DA28 used only the data collected from subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. DA29 utilized data collected from all subjects during the 2008 participation period and only the subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. A LDA was completed for DA28 and DA29 on the BLR, ALRField, ALRClinic, and Post-Study data which had all been standardized using the Pre-Study data according to equation 16. This data had been collected using the “No Tube” method, or by inserting the C320 snout into a stopcock inserted in the breath storage bag's luer port (fig. 1a).

***Data Analysis 30 (DA30) and Data Analysis 31 (DA31)***

DA30 used only the data collected from subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. DA31 utilized data collected from all subjects during the 2008 participation period and only the subjects who completed the Pre-Study and subsequent collections during the 2010 participation periods. A LDA was completed for DA30 and DA31 on the BLR, ALRField, ALRClinic, and Post-Study data which had been standardized using the Pre-Study data according to equation 16. This data had been collected using the “Tube” method, or by inserting

the C320 snout into the series of tubing connected to the breath storage bag's luer port (fig. 1b).

## **PRELIMINARY RESULTS AND DISCUSSION**

The cross validation results were summarized for 21 LDA completed on different combinations of data and standardization techniques in table 5A.1. Initially the cross validation results were compared for DA1 through DA16. DA5, DA8, DA9, DA13, DA14, and DA15 each correctly classified 100% of the data, so their canonical plots were compared to each other and to that of the raw data (fig. 5A.1a). All canonical plots presented the points and 95% confidence ellipse for each class in the two dimensions that best separated the classes.

The canonical plot for the raw data, or DA1, (fig. 5A.1a) presented clear separation between the Post-Study and other sample classes but significant overlapping occurred between the other sample classes. The DA5 canonical plot did not accurately display due to extreme separation between some classes. Both models were rejected.

The DA8 plot (fig. 5A.1b) presented clearer separation between the sample classes than DA1. The Post-Study class was clearly separated from the other classes. The LI training load classes were clearly separated from the HI training load classes. Some separation occurred between the classes within each training load. These results indicated standardizing the data with the Pre-Study class improved the model. The DA9 plot (fig. 5A.1c) was similar to the DA8 plot, which suggested adding carbon dioxide to the standardization method did not significantly improve the model over only standardizing with the Pre-Study data.

The DA13 plot (fig. 5A.1d) presented greater separation between all sample classes than the previous analyses. The Post-Study class was clearly separated from the other classes. The LI training load classes were clustered together away from the HI training loads which were also clustered together. The LI training load classes were clustered together closer than the HI training load classes which was expected since it was thought a HI training load would induce a greater physical response in the athletes. These results indicated utilizing the Pre-Study class to standardize data provided a better model for VOCs in the breath due to training loads. The DA14 plot (fig. 5A.1e) presented similar results as DA13. However, there was less separation between the ALRField and Post-Study classes which indicated there could be confusion in identifying samples from these two collection times. This suggested adding an additional standardization variable of carbon dioxide provided worse results than only standardizing with the Pre-Study data. The DA15 plot (fig. 5A.1f) presented similar results as those for DA13 and DA14, which verified the suggestion that adding an extra standardization variable of carbon dioxide did not improve the prediction model.

A comparison of DA1 through DA16 indicated standardizing with the before run (BLR or BSR) or carbon dioxide data does not improve the prediction model. Standardizing by subtracting the Pre-Study (eq. 16) data provided a better model

than dividing by the Pre-Study data and was the selected method to pursue. However, the LDA analyses using the equation 16 technique were completed without the raw Pre-Study data which did not allow for a direct comparison of the effects of the cumulative training loads on breathprints. DA17 through DA21 were completed to determine how adding the raw Pre-Study data back into the analyses would affect the model.

Cross validation results (table 5A.1) were compared for DA17 though DA21. Results were best (96% correct classification) when data was narrowed to only include athletes who completed the Pre-Study and subsequent collection times, as in DA19. However, this excluded many subjects from the study so the results from DA21 were also reviewed. DA21 included subjects from Summer 2008 and those subjects from the 2010 participation periods who provided Pre-Study and subsequent samples. The DA21 cross validation correctly classified 81.4% of the data. The DA19 (fig. 3a) and DA21 (fig. 3b) plots presented significant separation from the Post-Study and the other sample classes. Both models clustered the LI training load classes together. The HI training load classes were also clustered together. Significant overlapping occurred between the LI training loads in both models while less overlapping occurred between the HI training load classes. However, the DA21 plot overlapped the ALRField and ALRClinic sample classes.

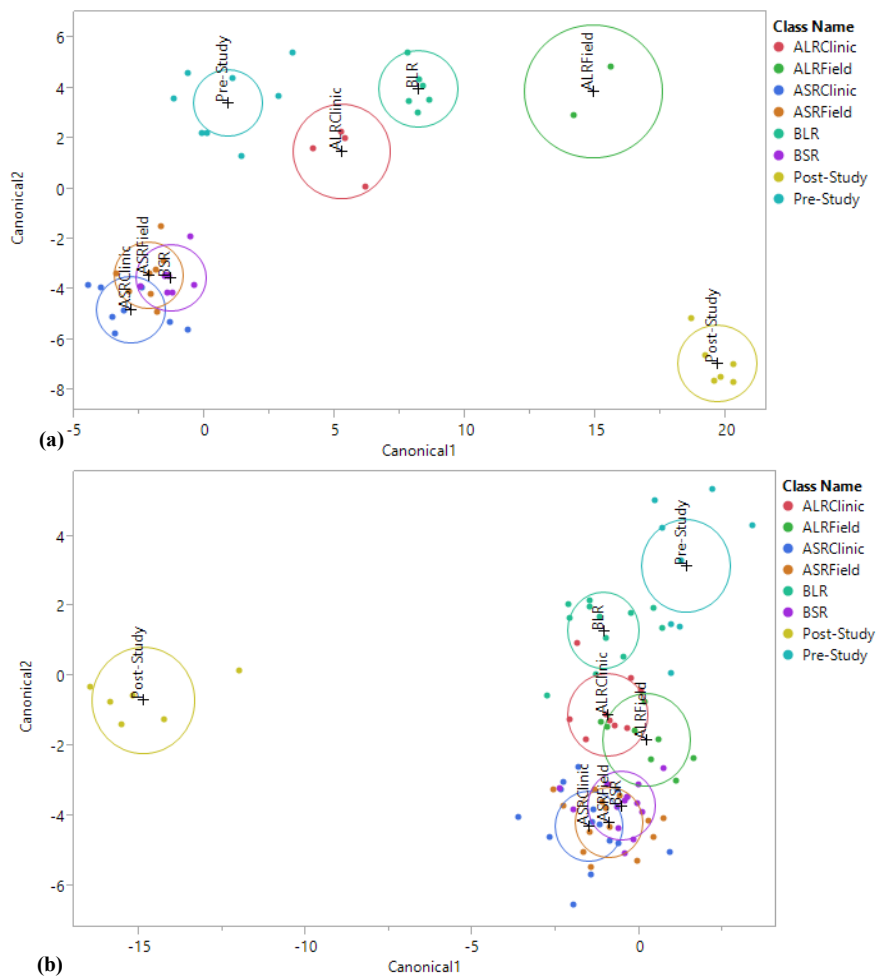


Figure 3. DA19 (a) and DA21 (b) canonical plots with 95% confidence ellipses for each mean, comparing all sample classes.

Previous researchers found enoses can correctly classify between 80% and 100% of breathprints collected from ill patients (table 2). In the present study, the selected models correctly classified 81.4% (DA21) and 96% (DA19) of the data. These results exceeded expectations based on the previous studies. The DA19 and DA21 subsets of data, and standardization technique were used for the rest of the study.

**Table 2. Enoses have been observed to detect illnesses in breath. CDA is canonical discriminant analysis and DA is discriminant analysis.**

Enose System	Patient Sample Class 1	Patient Sample Class 2	Patient Sample Class 3	Statistical Methods	Correctly Classified (%)	Source
C320	Non-Small Cell Lung Cancer	Healthy		PCA and CDA	80	(Dragonieri et al., 2009)
		Chronic Obstructive Pulmonary Disease			85	
C320	Malignant Pleural Mesothelioma	Healthy Asbestos Exposure		PCA and CDA	84.6 80.8	(Dragonieri et al., 2012)
C320	Mild Asthma Severe Asthma	Young Controls		PCA and LDA	100	(Dragonieri et al., 2007)
		Old Controls			90	
C320	Malignant Mesothelioma	Controls		PCA and CDA	95	(Chapman et al., 2012)
		Other Asbestos-Related Diseases	Controls		88	
Piezoelectric Quartz Crystals	Uremia	Chronic Renal Insufficiency	Chronic Renal Failure	DA	86.78	(Lin et al., 2001)

PCA, PLS, and stepwise LDA were completed for the DA19 and DA21 data subsets. LDA were then completed on the reduced data sets (table 5A.2). DA22 and DA23, LDA after PCA, did not improve cross validation results nor did they improve separation of classes in the resulting plots (e.g. fig. 5A.2a). A stepwise LDA removed three sensors from the analysis (DA24 and DA25) but did significantly improve the cross validation results or separation in the plots (e.g. fig. 5A.2b). LDA after PLS, DA26 and DA27, improved cross validation results but not separation between classes in the plots.

The “No Tube” (fig. 1a) and “Tube” (fig. 1b) apparatuses were compared using the BLR, ALRField, ALRClinic, and Post-Study classes of the DA19 and DA21 data subsets. DA28 and DA29 used the No Tube apparatus while DA30 and DA31 used the Tube apparatus. DA28 through DA31 correctly classified 100% of the data (table 5A.3) and the plots (fig. 5A.3) presented similar separation between the sample classes.

## CONCLUSION

Overtraining occurs in athletes who exceed their optimum performance point by completing more training loads than rest periods and can have serious consequences. This preliminary study found a Cyranose® 320 enose can discriminate between breathprints of collegiate long distance runners after completing cumulative and acute training loads. An external statistical model standardizing data with the Pre-Study breath samples (eq. 16) allowed the C320 to correctly classify 96% of the athlete’s breathprints.

Future work should focus on developing a large library of breathprints from athletes after acute and cumulative training loads which will assist in creating a more robust C320 statistical prediction model. Pre-Study breath samples should be

collected to ensure the subsequent data can be indirectly standardized using equation 16. Direct standardization of data should be explored by using carbon dioxide as a baseline purge for the C320 unit. The tube apparatus should be used to analyze samples with the C320 to help prevent contamination from outside air sources.

It is believed the C320 enose will provide a clear, rapid, simple, and noninvasive method for clinically diagnosing the onset of overtraining. Additionally, as the unit is small and handheld, it can be used in the field to provide immediate feedback on training stresses, allowing athletic staffs to make immediate adjustments to training regimens for training optimization.

#### ACKNOWLEDGEMENTS

The authors would like to thank Dr. Mark Rogers, Dr. Gregory Beato, and the members of the Virginia Tech long distance running program for their interest and participation in this study which would not have been possible without them. The authors would also like to thank VCOM's Harvey Peters Foundation, Virginia Tech's Institute for Critical Technology and Applied Science, and Virginia Tech's Graduate Student Assembly's Graduate Research Development Program who provided monetary support for this project.

#### REFERENCES

- Anglem, N., Lucas, S. J., Rose, E. A., & Cotter, J. D. (2008). Mood, illness and injury responses and recovery with adventure racing. *Wilderness Environ Med, 19*(1), 30-38. doi: 10.1580/07-WEME-OR-091.1
- Armstrong, L. E., & VanHeest, J. L. (2002). The unknown mechanism of the overtraining syndrome: clues from depression and psychoneuroimmunology. *Sports Med, 32*(3), 185-209.
- Bell, L. M., & Ingle, L. (2013). Psycho-Physiological Markers of Overreaching and Overtraining in Endurance Sports: A Review of the Evidence. *Medicina Sportiva, 17*(2), 81-97. doi: 10.5604/17342260.1055272
- Booth, C. K., Probert, B., Forbes-Ewan, C., & Coad, R. A. (2006). Australian army recruits in training display symptoms of overtraining. *Mil Med, 171*(11), 1059-1064.
- Chapman, E. A., Thomas, P. S., Stone, E., Lewis, C., & Yates, D. H. (2012). A breath test for malignant mesothelioma using an electronic nose. *Eur Respir J, 40*(2), 448-454. doi: 10.1183/09031936.00040911
- de Meij, T. G., Larbi, I. B., van der Schee, M. P., Lentferink, Y. E., Paff, T., Terhaar Sive Droste, J. S., . . . de Boer, N. K. (2014). Electronic nose can discriminate colorectal carcinoma and advanced adenomas by fecal volatile biomarker analysis: proof of principle study. *Int J Cancer, 134*(5), 1132-1138. doi: 10.1002/ijc.28446
- Di Natale, C., Macagnano, A., Martinelli, E., Paolesse, R., D'Arcangelo, G., Roscioni, C., . . . D'Amico, A. (2003). Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosens Bioelectron, 18*(10), 1209-1218.
- Dragonieri, S., Annema, J. T., Schot, R., van der Schee, M. P., Spanevello, A., Carratu, P., . . . Sterk, P. J. (2009). An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. *Lung Cancer, 64*(2), 166-170. doi: 10.1016/j.lungcan.2008.08.008
- Dragonieri, S., Schot, R., Mertens, B. J., Le Cessie, S., Gauw, S. A., Spanevello, A., . . . Sterk, P. J. (2007). An electronic nose in the discrimination of patients with asthma and controls. *J Allergy Clin Immunol, 120*(4), 856-862. doi: 10.1016/j.jaci.2007.05.043
- Dragonieri, S., van der Schee, M. P., Massaro, T., Schiavulli, N., Brinkman, P., Pinca, A., . . . Sterk, P. J. (2012). An electronic nose distinguishes exhaled breath of patients with Malignant Pleural Mesothelioma from controls. *Lung Cancer, 75*(3), 326-331. doi: 10.1016/j.lungcan.2011.08.009
- Dutta, R., Morgan, D., Baker, N., Gardner, J. W., & Hines, E. L. (2005). Identification of Staphylococcus aureus infections in hospital environment: electronic nose based approach. *Sensors and Actuators B, 109*, 335-362. doi: 10.1016/j.snb.2005.01.013
- Fens, N., van der Schee, M. P., Brinkman, P., & Sterk, P. J. (2013). Exhaled breath analysis by electronic nose in airways disease. Established issues and key questions. *Clin Exp Allergy, 43*(7), 705-715. doi: 10.1111/cea.12052

- Finaud, J., Lac, G., & Filaire, E. (2006). Oxidative stress : relationship with exercise and training. *Sports Med*, 36(4), 327-358.
- Halson, S. L., & Jeukendrup, A. E. (2004). Does overtraining exist? An analysis of overreaching and overtraining research. *Sports Med*, 34(14), 967-981.
- Hockstein, N. G., Thaler, E. R., Torigian, D., Miller, W. T., Jr., Deffenderfer, O., & Hanson, C. W. (2004). Diagnosis of pneumonia with an electronic nose: correlation of vapor signature with chest computed tomography scan findings. *Laryngoscope*, 114(10), 1701-1705. doi: 10.1097/00005537-200410000-00005
- Inc., S. I. (2013). *JMP® 11 Multivariate Methods*. Cary, NC: SAS Institute Inc.
- Kodogiannis, V. S., Lygouras, J. N., Tarczynski, A., & Chowdrey, H. S. (2008). Artificial odor discrimination system using electronic nose and neural networks for the identification of urinary tract infection. *IEEE Trans Inf Technol Biomed*, 12(6), 707-713.
- Lakier Smith, L. (2003). Overtraining, excessive exercise, and altered immunity: is this a T helper-1 versus T helper-2 lymphocyte response? *Sports Med*, 33(5), 347-364.
- Lin, Y., Guo, H., Chang, Y., Kao, M., Wang, H., & Hong, R. (2001). Application of the electronic nose for uremia diagnosis. *Sensors and Actuators B*, 76(1-3), 177-180.
- Margonis, K., Fatouros, I. G., Jamurtas, A. Z., Nikolaidis, M. G., Douroudos, I., Chatzinikolaou, A., . . . Kouretas, D. (2007). Oxidative stress biomarkers responses to physical overtraining: implications for diagnosis. *Free Radic Biol Med*, 43(6), 901-910. doi: 10.1016/j.freeradbiomed.2007.05.022
- Morgan, W. P., Brown, D. R., Raglin, J. S., O'Connor, P. J., & Ellickson, K. A. (1987). Psychological monitoring of overtraining and staleness. *Br J Sports Med*, 21(3), 107-114.
- Nagle, H. T., Gutierrez-Osuna, R., & Schiffman, S. S. (1998). The How and Why of Electronic Noses. *IEEE Spectrum*, 35(9), 22-31.
- Pavlou, A. K., & Turner, A. P. (2000). Sniffing out the truth: clinical diagnosis using the electronic nose. *Clin Chem Lab Med*, 38(2), 99-112. doi: 10.1515/CCLM.2000.016
- Peluso, M. A., & Guerra de Andrade, L. H. (2005). Physical activity and mental health: the association between exercise and mood. *Clinics (Sao Paulo)*, 60(1), 61-70. doi: /S1807-59322005000100012
- Rock, F., Barsan, N., & Weimar, U. (2008). Electronic nose: current status and future trends. *Chem Rev*, 108(2), 705-725. doi: 10.1021/cr068121q
- Thaler, E. R. (2002). Candidate's thesis: the diagnostic utility of an electronic nose: rhinologic applications. *Laryngoscope*, 112(9), 1533-1542.
- Voss, A., Baier, V., Reisch, R., von Roda, K., Elsner, P., Ahlers, H., & Stein, G. (2005). Smelling renal dysfunction via electronic nose. *Ann Biomed Eng*, 33(5), 656-660.
- Wang, P., Tan, Y., Xie, H., & Shen, F. (1997). A novel method for diabetes diagnosis based on electronic nose. *Biosens Bioelectron*, 12(9-10), 1031-1036.
- Whysong, C., Brolinson, P. G., Mallikarjunan, P., Misra, H. P., & Jia, Z. (2014). Chemosensory Evaluation of Training and Oxidative Stress in Long Distance Runners, A Brief Report (In Review). *Journal of the American Osteopathic Association*, 10.
- Yamada, Y., Takahashi, Y., Konishi, K., & Katsumi, I. (2007). Association of odor from infected root canal analyzed by an electronic nose with isolated bacteria. *J Endod*, 33(9), 1106-1109.

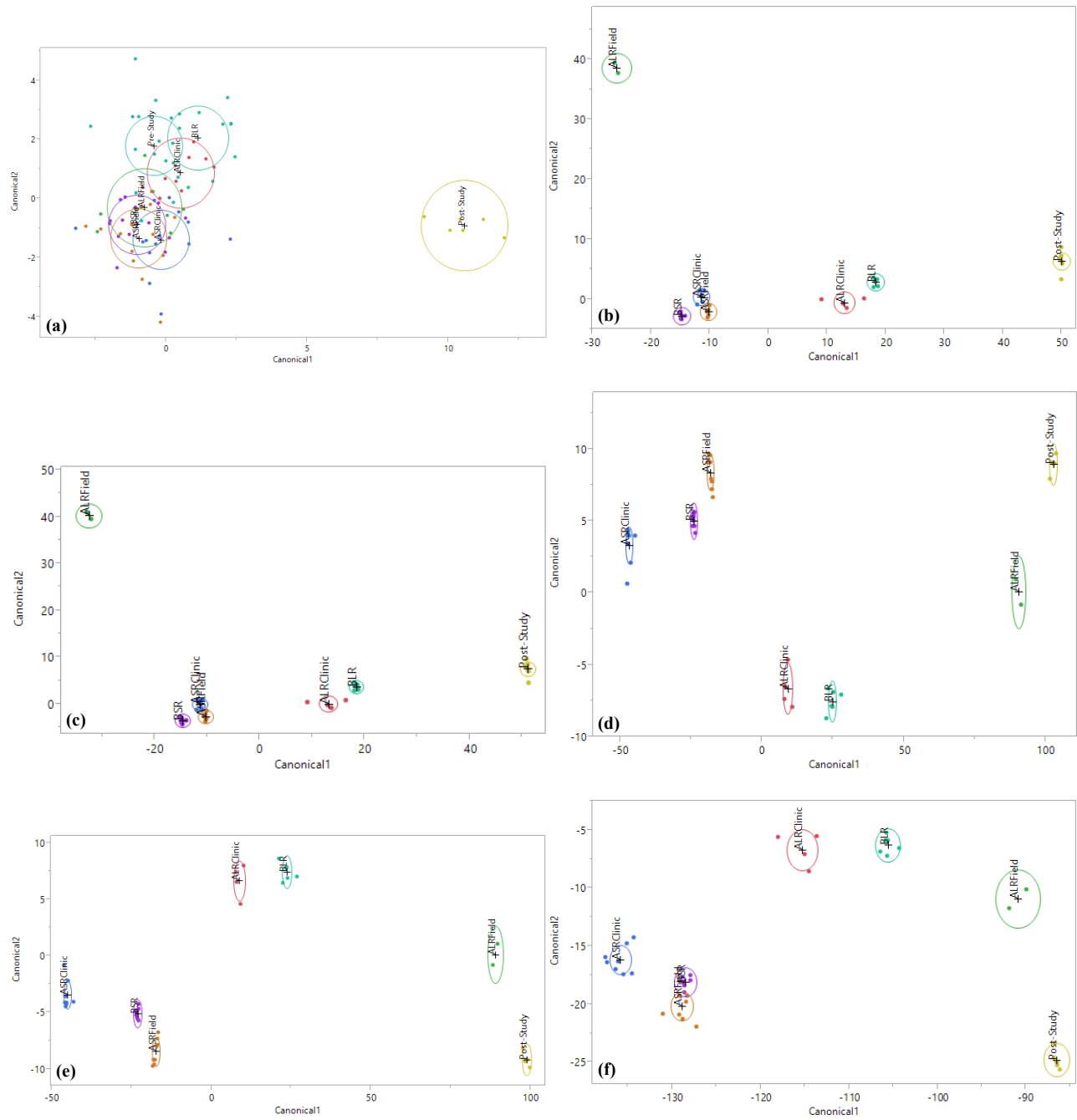
## APPENDIX CHAPTER 5A (5A)

Table 5A.1. Cross validation results were compared for 21 LDA using different standardization techniques.

SUMMARY OF ANALYSIS						RESULTS	
DA	Included Data		Standardization Methods		Excluded Data	Misclassified	
	Collection Periods	Subjects	Data	Equation		Number	Percent
DA1	All	All				22	23.91
DA2	All	All	CO <sub>2</sub>	1		22	23.91
DA3	All	All	CO <sub>2</sub>	2		22	23.91
DA4	All	All	CO <sub>2</sub> , AR	3	BSR, BLR	6	9.091
DA5	Spring & Fall 2010	Pre-Study & Subsequent Collections	CO <sub>2</sub> , Pre	4	Pre-Study	0	0
DA6	All	All	BR	5	BSR, BLR	4	6.061
DA7	All	All	CO <sub>2</sub> , BR	6, 7, 8		5	7.576
DA8	Spring & Fall 2010	All	Pre	9	Pre-Study	0	0
DA9	Spring & Fall 2010	All	CO <sub>2</sub> , Pre	10	Pre-Study	0	0
DA10	All	All	BR	11	BSR, BLR	5	7.576
DA11	All	All	CO <sub>2</sub> , BR	12	BSR, BLR	5	7.576
DA12	All	All	CO <sub>2</sub> , BR	13, 14, 15	BSR, BLR	4	6.061
DA13	Spring & Fall 2010	All	Pre	16	Pre-Study	0	0
DA14	Spring & Fall 2010	All	CO <sub>2</sub> , Pre	17	Pre-Study	0	0
DA15	Spring & Fall 2010	All	CO <sub>2</sub> , Pre	18	Pre-Study	0	0
DA16	All	All	$\mu$ , $\sigma$	19		78	84.78
DA17	Spring & Fall 2010	Pre-Study & Subsequent Collections				4	8
DA18	Spring & Fall 2010	All	Pre	16		4	7.143
DA19	Spring & Fall 2010	Pre-Study & Subsequent Collections	Pre	16		2	4
DA20	All	All	Pre	16		20	21.74
DA21	All	All (2008) Pre-Study & Subsequent Collections (2010)	Pre	16		16	18.6

Table 5A.2. Cross validation results were compared for six LDA using three data reduction methods. Data was standardized using the Pre-Study class and equation 16. No data classes were excluded from DA 22 through DA27.

Description of Analysis						Results		
DA	Included Data		LDA Preprocessing Methods			Misclassified		
	Collection Periods	Subjects	Data Reduction Method	Number of PCA	C320 Sensors Removed	Number of Latent Variables	Number	Percent
DA22	Spring & Fall 2010	Pre-Study & Subsequent Collections	PCA	9			14	28
DA23	All	All (2008) Pre-Study & Subsequent Collections (2010)		6			58	67.44
DA24	Spring & Fall 2010	Pre-Study & Subsequent Collections	Stepwise LDA		7, 20, 22		1	2
DA25	All	All (2008) Pre-Study & Subsequent Collections (2010)			7, 12, 22		14	16.28
DA26	Spring & Fall 2010	Pre-Study & Subsequent Collections	PLS			3	0	0
DA27	All	All (2008) Pre-Study & Subsequent Collections (2010)				4	0	0



**Figure 5A.1. DA1 (a), DA8 (b), DA9 (c), DA13 (d), DA14 (e), and DA15 (f) canonical plots with 95% confidence ellipses for each mean, comparing the ALRClinic (red), ALRField (green), ASRClinic (blue), ASRField (orange), BLR (teal), BSR (purple), Post-Study (yellow), and Pre-Study (light blue) classes.**

Table 5A.3. Cross validation results were compared for LDA utilizing two different sample analysis apparatuses: No Tube (fig. 1a) and Tube (fig. 1b). Data was standardized using the Pre-Study class and equation 16.

Analysis Summary					Results	
DA	Included Data			Apparatus	Misclassified	
	Collection Periods	Subjects	Data Classes		Number	Percent
DA28	Spring & Fall 2010	Pre-Study & Subsequent Collections	ALR, ALRField, ALRClinic, Post-Study	No Tube	0	0
DA29	All	All (2008) Pre-Study & Subsequent Collections (2010)			0	0
DA30	Spring & Fall 2010	Pre-Study & Subsequent Collections	ALR, ALRField, ALRClinic, Post-Study	Tube	0	0
DA31	All	All (2008) Pre-Study & Subsequent Collections (2010)			0	0

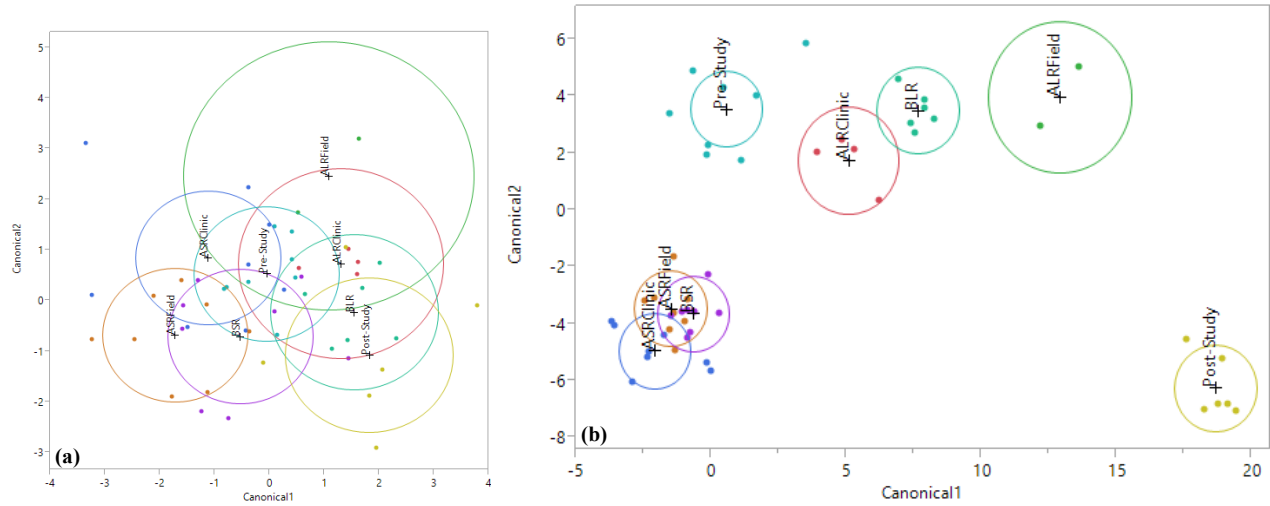


Figure 5A.2. DA22 (a) and DA24 (b) canonical plots with 95% confidence ellipses for each mean, comparing the ALRClinic (red), ALRField (green), ASRClinic (blue), ASRField (orange), BLR (teal), BSR (purple), Post-Study (yellow), and Pre-Study (light blue) classes.

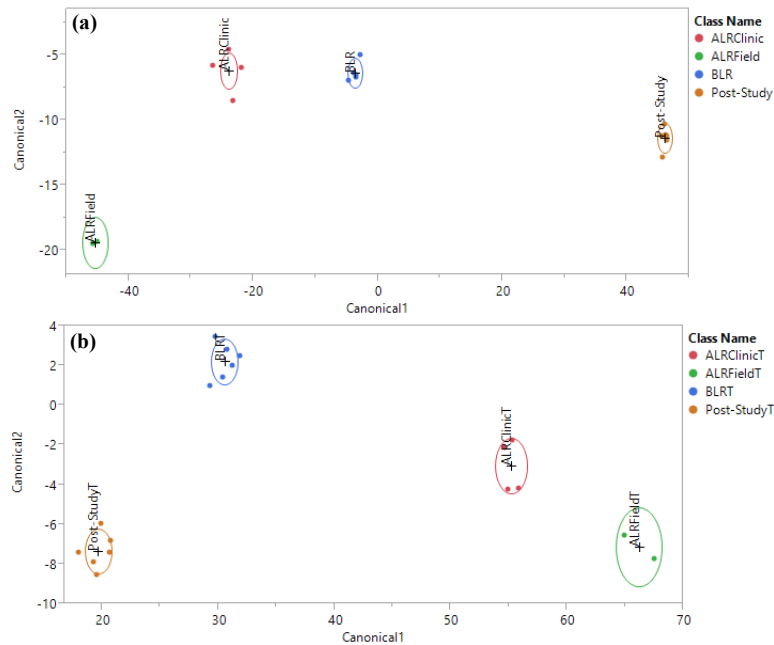


Figure 5A.3. DA28 (a) and DA30 (b) canonical plots with 95% confidence ellipses for each mean.