Understanding Physiological Responses to Stressors during Physical Activity

Jin-Hyuk Hong Julian Ramos Anind K. Dey
Carnegie Mellon University, Human-Computer Interaction Institute
5000 Forbes, Pittsburgh, USA, PA 15213
hjinh7@gmail.com ingenia@andrew.cmu.edu anind@cs.cmu.edu

ABSTRACT
With advances in physiological sensors, we are able to understand people's physiological status and recognize stress to provide beneficial services. Despite the great potential in physiological stress recognition, there are some critical issues that need to be addressed such as the sensitivity and variability of physiology to many factors other than stress (*e.g.*, physical activity). To resolve these issues, in this paper, we focus on the understanding of physiological responses to both stressor and physical activity and perform stress recognition, particularly in situations having multiple stimuli: physical activity and stressors. We construct stress models that correspond to individual situations, and we validate our stress modeling in the presence of physical activity. Analysis of our experiments provides an understanding on how physiological responses change with different stressors and how physical activity confounds stress recognition with physiological responses. In both objective and subjective settings, the accuracy of stress recognition drops by more than 14% when physical activity is performed. However, by modularizing stress models with respect to physical activity, we can recognize stress with accuracies of 82% (objective stress) and 87% (subjective stress), achieving more than a 5-10% improvement from approaches that do not take physical activity into account.

Author Keywords
Stress recognition, physiological responses, physical activity

ACM Classification Keywords
H.m. Information Systems: Miscellaneous.

INTRODUCTION

Due to an overabundance of stress in our modern lives, being stressed is now regarded as a negative experience when failing to adequately respond to mental, emotional, or physical demands [1, 2]. Thus, stress is often produced when people are exposed to the demands and pressures from physical or mental activities in their daily life, or forced by their self-imposed demands, obligations and self-criticism [3,4]. Since stress can lead to significant health problems such as headaches, trouble sleeping and fatigue, it is necessary to properly understand and measure stress in the natural environment [5]. In behavioral science, people commonly use self-reports that periodically collect instantaneous measurements of perceived stress, but it is practically impossible to detect stress in a timely manner with the self-reporting approach due to its imposed burden and obtrusiveness. In clinical settings, cognitive and physical assessments are used such as the Mini-Mental state Examination, the Mental Status Questionnaire and physical tests of walking speed, grip strength and button pressing speed, but they are not designed for continuous stress recognition [6].

Stress recognition has been actively investigated in the area of affective computing based on various sensing systems and inference techniques. Outward expressions, such as speech, facial expression and behaviors, have been used to recognize emotion or stress, but they are still limited in their ability to correctly detect a person's affect in natural environments. Instead, psychophysiological measurements are now perceived as having the most potential for detecting changes in stress level since there are an increasing availability of sensors that can conveniently capture states of a human and his surrounding environment [6,7]. Since physiology responds to stress (*e.g.*, changes in stress hormone levels may lead to changes in heart rate, blood pressure, pupil dilation, and galvanic skin responses), the main task of stress recognition based on physiology is to understand the relationship between a person's psychophysiological responses and stress. It would be especially beneficial if some automated methods or algorithms existed that could recognize such states [8]. Moreover, the recent development of advanced physiological sensors makes it more practical to automate stress recognition, since they could provide an objective, continuous, unobtrusive and passive way to capture physiological responses in natural environments [5].

Despite the great potential in stress recognition, there are some practical issues in exploiting physiological responses. First, the physiological system reacts not only to changes in stress, but also to many other factors such as changes in physical or mental conditions; thus, a physiological change does not necessarily imply a stress change [5,9]. It often

UbiComp' 12, Sep 5 – Sep 8, 2012, Pittsburgh, USA.

responds to physical activity demands, physical discomfort, noise, changes in posture, lighting conditions and mental task demand and emotional stress. In fact, researchers have exploited physiological responses in recognizing physical activity [9]. Second, sensing platforms for detecting physiological signals have limitations [10]. These platforms are typically sensitive to noise, can miss sensor readings, and have issues with wearability and battery limits. Third, stress has wide variations in its physiological expression. For example, physical stress makes a sympathetic nervous response dominant but an adrenal response is dominant during cognitive stress. Finally, physiological responses to stress and physical activity are very individual [4,11]. People sometimes express different physiological responses to some degree for the same stressor, which indicates the necessity of personalizing affect models.

Motivated by these challenges, this paper presents an understanding of physiological changes, particularly in situations having multiple stimuli that may complexly affect people's physiological system. Three different stressors are administered to participants (mental arithmetic [12], loud sounds [13] and cold water pressor [14]) while three activities (sitting, walking, and bicycling) having different levels of intensity are performed. A variety of physiological channels, including respiration, electrocardiogram, skin conductivity and body temperature, are captured by two representative commercial sensors, and various data analysis methods of feature extraction and classification are used to understand the changes in physiological responses and recognize stress in the presence of physical activities. In order to detect stress even during physical activity, we apply an approach that consists of multiple stress models, each of which classifies stress during a specific activity.

In the following section, we provide an overview of related research on stress, physiological sensors, and approaches for performing stress recognition. We then provide details about our stress data collection with various stressors in the presence of physical activity. We present a framework to analyze physiological signals and our stress recognition method. The next section provides experimental results including an analysis of physiological features for each stimulus including stressors and physical activities and for stress recognition performance. We then discuss limitations of our work and remaining challenges, and conclude with perspectives related to future work.

RELATED WORK

Physiological stress recognition

When a situation is appraised as threatening and the individual is unable to cope with an appropriate reaction, emotional and physiological responses, often called *stress*, are stimulated. Stress has been investigated as a dimension of negative affect in psychological and medical research [4,5]. Since stress can lead to significant health problems, the understanding and assessment of stress is a crucial issue; stress must be detected in a timely manner and individuals must receive appropriate treatment. Over the past two decades, several markers of stress have been identified [5], where physiological measurements are now considered as the most common tool for detecting changes in stress level although facial expressions and keyboard-typing patterns have also been used [6,15].

In contrast to psychology where researchers often aim to achieve a pure understanding of physiological responses to stress, recent studies are more focused on the automation of stress detection and its application to real-life situations [5,8,15]. Healey and Picard used four types of physiological sensors including an electrocardiogram (ECG), eletromyogram (EMG), skin conductivity (also known as galvanic skin response (GSR)), and respiration. They collected physiological data during real-world driving tasks to measure drivers' stress levels [7]. Nasoz *et al*. designed a multi-modal intelligent car interface with negative affective states such as panic/fear, frustration/anger and boredom/fatigue [8]. To identify these states, they used physiological responses such as skin conductance, heart activity, respiration, muscle activity, and finger pressure.

In order to recognize stress and fatigue, Liao *et al*. proposed a unifying framework using a dynamic probabilistic decision-theoretic model, which included affective state recognition, active sensory action selection and user assistance [3]. They used four different types of inputs: physiological responses, physical appearance features, user performance and behavioral data, and focused on constructing an optimal feature set to improve the recognition performance instead of using all available features. Sierra *et al*. used a fuzzy logic approach that models galvanic skin response and heart rate as a result of stress induced by hyperventilation and a talk preparation task [15]. Plarre *et al*. designed a mobile platform that continuously captures various physiological channels to infer stress in the natural environment [5]. Although they applied a systematic approach to map a physiological response to perceived stress, their focus was on the development and evaluation of their sensing platform rather than on providing a deeper understanding of the relationship between physiological responses and stress or other factors like physical activity.

Affective computing with physiological responses

After Picard coined the term *affective computing* in the mid 1990's, Picard and Healey worked on the basics of a computational approach for analyzing affective physiological state [16]. Various features and machine learning techniques were applied to discover promising features and to model affect in a variety of situations [16,17]. Similar to other pattern recognition tasks based on sensory signals, this approach has several basic steps for exploiting physiological responses in modeling affect: 1) selecting specific physiological sensors and capturing signals, 2) extracting informative features from the raw sensory channels, 3) modeling the relationship between target affects and the extracted features.

As mentioned above, it has become much easier to collect a variety of physiological responses due to advances in sensor technology. Each physiological channel has its own distinct strengths and limitations in terms of accuracy, resolution or sampling rate, sensor's wearability, and power consumption. Since raw sensory signals are often not very useful for directly modeling affect, affect modeling is usually pursued with feature variables extracted from these raw signals. From the means and standard deviations of the raw signals to the dominant frequency and power of the raw signals in the frequency domain, a broad spectrum of features have already been introduced in modeling affect [16,17]. For certain physiological responses like ECG, a sophisticated algorithm (*e.g.*, QRS detection algorithm) has developed to extract features (*e.g.*, heart rate or RR interval) widely used in biomedical engineering [17,18]. The choice of features is a fundamental and highly problem-dependent task in the development of affect models, being conducted either using automated feature selection or expert knowledge [3].

Since even distinguishing features do not have sufficient capability or information required to model affect, researchers apply machine learning techniques to discover the complex relationship or recognize target affective states. The more a sensor or a user is exposed to the natural environment, the more uncertainty and complexity are introduced, making it impossible to address with manual approaches. A variety of machine learning methods can be applied to the affect modeling problem, such as k-nearest neighbors [8,11,16], linear discriminant analysis [7,11,17,18], multilayer perceptron [11], support vector machines [19], fuzzy logic [15], Bayesian networks [8], dynamic Bayesian networks [3].

Although much work has introduced advanced techniques and systems that accurately model affect and detect stress, we still need to consider issues about how to apply stress recognition in natural environments. Physiological responses may often be activated by other factors on the body such as speaking and physical activity. To use stress models with physiological responses in practice, we must be able to correctly detect stress even during other activities that affect one's physiology. We are motivated by the fact that few studies have studied the impact of such real-world factors, and the necessity to pursue work addressing the effect of physical activity, as it is a daily common event, on physiological responses and stress recognition.

DATA COLLECTION

We collected physiological data from 20 participants while they performed physical activities. Here, we describe our data collection method.

Materials and setup

The data collection was performed in a closed laboratory environment under controlled temperature. A number of physiological signals were recorded using two commercial sensors, the BioHarness BT and the SenseWear armband (more detail available in the next section), and an Android smartphone that connects to the BioHarness BT via Bluetooth for real-time monitoring of raw ECG signals (to verify the measurements were of high quality). The placement of the two sensors is shown in Fig. 1. Participants used a treadmill and a stationary bike to perform physical activity.

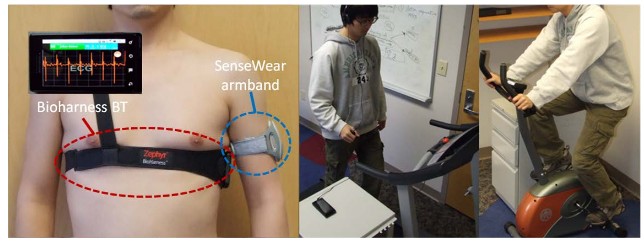

Figure 1. Sensor platform and exercise equipment

Participants

We recruited 20 participants (10 females and 10 males), mainly students from Carnegie Mellon University, ranging in age from 18 to 38 years old (average of 26.3, std. dev. of 5.3). The participants were asked to perform our data collection tasks on 4 different days, and reimbursed $30 for each day. 19 participants completed all four sessions.

Procedure

When a participant arrived at our laboratory, an experimenter took him/her to a meeting room to explain the data collection procedure. The experimenter then helped the participant wear the two sensors and verified the quality of the output signals, before initiating a data collection session.

As shown in Fig. 2, the data collection starts and ends with a baseline recording (10 and 5 minutes each), and consists of three sub-sessions of physical activity with different intensities: sitting, walking (around 2 mph) and bicycling (around 16 mph). Following the ACSM (American College of Sports Medicine)'s guideline, the treadmill speed was chosen to match the speed that causes a subject's heart rate to be over 50% of the maximal heart rate, in the warm-up period. Similarly, for bicycling speed, 60% of the maximal heart rate was kept as minimum. Each sub-session was divided into 4 tasks, 3 with stressors and 1 without any stressor. In order to provide a variety of stressful situations, we presented three different types of stressors [used in 12,13,14] to the participant, including 1) a mental arithmetic problem (denote *math*) as a mental stressor where the participant needs to repeatedly subtract a two-digit number from a 4-digit number (*e.g.*, keep subtracting 13 from 2,081), 2) a cold water pressor (denote *cold*) as a physical stressor where the participant keeps his/her hand in cold water (4°C), and 3) noisy sound (denote *sound*) as another mental stressor where a number of random loud sounds (*e.g.*, screaming and snoring) are presented.

Each task was administered for 3 minutes, followed by a short break (2 minutes) during which participants assessed their own stress levels (5-point Likert scale from 1: no stress and 5: most stressed) and filled out the NASA-TLX questionnaire [20]. The NASA-TLX method was designed to assess a subject's overall workload using a weighted multi-dimensional rating for six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. We exploit the NASA-TLX method to objectively evaluate the cognitive load induced by our three stressors. While this does not directly assess stress, it is intended to complement the self-assessment of stress. After both walking and bicycling, participants take a 10 minute long break to recover. The 3 physical activities are performed in a different order on each visit, and the stressors are presented in a counter-balanced order based on the Latin square method.

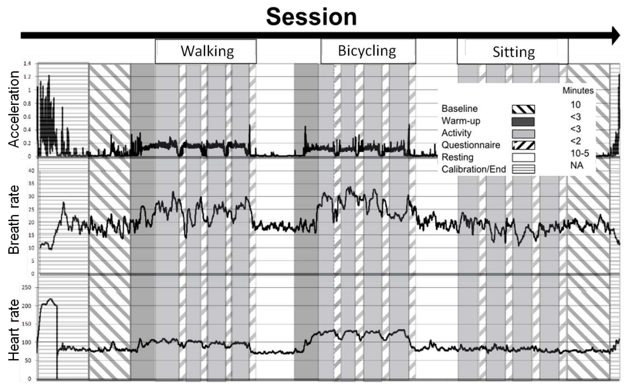

Figure 2. Overview of data collection

A FRAMEWORK FOR PHYSIOLOGICAL ANALYSIS

The basic structure for analyzing physiological responses and recognizing stress is illustrated in Fig. 3. Two popular sensors are used to continuously capture a variety of physiological responses, and the streaming signals are filtered and segmented with a constant frequency. After the preprocessing stage, we extract meaningful features from each segment composed of raw biosignals. In stress recognition, we apply two popular methods: naïve Bayes classifier (NB) and Bayesian networks (BN), where we discretize numeric variables into a nominal form before using them. Different feature sets are used in building stress recognition models, since not all physiological features are informative.

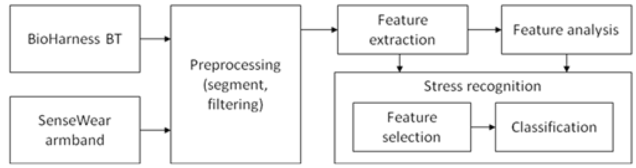

Figure 3. Overview of basic processes in physiological analysis

Feature extraction

We use the Bioharness BT and the SenseWear armband to capture the physiological responses as shown in Table 1. The BioHarness BT (Zephyr Technology) is a chest-worn band with embedded sensors measuring 3D acceleration, breathing, ECG and skin temperature. The SenseWear armband (BodyMedia) is an armband sensing platform that senses 2-D acceleration, heat flux, galvanic skin response, skin temperature and near-body ambient temperature.

All sensory signals from the two sensors are segmented into final samples for a given window (30 seconds in this paper) in the preprocessing stage, and a number of features are extracted from the signals of each segment, by using conventional statistics in time series and geometric analysis [16,17]. First, some basic statistics including average, standard deviation and median values are calculated for every physiological stream with low sampling frequencies in Table 1. Second, some features are extracted from raw ECG and breathing signals based on medical and physiological research literature [17,21]. For the raw ECG signals, we detect the QRS complex (representing depolarization of the ventricles) and its interval in the ECG waveform, such as the RR interval (RRI: interval between ventricular depolarization), and calculate four values from the Poincare geometry of the RR intervals. Poincare geometry quantifies self-similarity and fluctuations in periodic features like RR intervals. The mean and standard deviations of the distances of each RR interval and the next interval to two lines such as y=x and y=-x+2*RRM (the mean of all RR intervals against the next interval) are calculated as follows:

Channels	Physiological responses / movement features	Frequency	Sensor
Electrocardiogram	Heart rate, ECG amplitude, ECG noise	1 Hz.	Bioharness BT
Respiration	Breathing rate, breathing wave amplitude		
Temperature	Skin temperature		
Acceleration	XYZ acceleration minimums and peaks, posture, vector magnitude, peak acceleration		
Respiration	Raw signal of breathing	18 Hz	
Electrocardiogram	HR RR value		
Electrocardiogram	Raw signal of ECG	250 Hz	
Acceleration	XY acceleration peaks, XY acceleration average, XY acceleration MAD (Mean Absolute Difference)	0.033 Hz	SenseWear Armband
Temperature	Skin temperature, near-body temperature		
Electrodermal activity	GSR (Galvanic skin conductance)		
Heat	Heat flux		
Acceleration	Step counter, lying down, sleep, physical activity, energy expenditure, sedentary, moderate, vigorous, very vigorous, METs (Metabolic Equivalent Tasks)	0.017 Hz	

Table 1. Physiological responses measured in this study

$$D_1(\mathrm{i}) = \frac{\mathrm{rrs_i} - \mathrm{rrs_{i+1}}}{\sqrt{2}}$$

$$D_2(\mathrm{i}) = \frac{\mathrm{rrs_i} + \mathrm{rrs_{i+1}} + 2 \times RRM}{\sqrt{2}}$$

mean of D_1	$\frac{\sum_1^N D_1(i)}{N}$, n:number of rrs in a given window
mean of D_2	$\frac{\sum_1^N D_2(i)}{N}$
standard deviation of D_1	$\sqrt{\frac{1}{\mathrm{N}}\sum_1^{\mathrm{N}}(\mathrm{D_1(i)} - \mu(\mathrm{D_1}))^2}$
standard deviation of D_2	$\sqrt{\frac{1}{\mathrm{N}}\sum_1^{\mathrm{N}}(\mathrm{D_2(i)} - \mu(\mathrm{D_2}))^2}$

Additionally, the heart rate and breathing rate are calculated again from the raw sensory signals although the sensor also provides this information. The volume of inhalation and exhalation is analyzed from the raw breathing signal. (Note that the raw breathing signal basically reflects the changes in pressure on the sensor, where we use the value indirectly as a measurement for the volume of breathing.)

All movement features such as acceleration peaks and METs (metabolic equivalents), which are widely used for activity recognition, are excluded from our analysis of stress recognition since they are specifically sensitive to physical activity. Instead, 42 features obtained from the physiological responses are used in this paper. A group of features are calculated from the measurements at a low sampling frequency, including average, standard deviation, and median of heart rate (F1~3), of breathing rate (F4~6), of skin temperature (F7~9), of breathing wave amplitude (F10~12), of ECG amplitude (F13~15), and of ECG noise (F16~18). Five features are calculated by analyzing the QRS complex and the RR intervals (F19~F23). We calculate features from the raw signals of the measurements at a relatively high sampling frequency, including average and median of derived heart rate (F24, F25), average and median of RR intervals derived from raw ECG signals (F26, F27), breathing rate derived from raw breath signals (F28), 5 statistics (maximum, minimum, median, average, and standard deviation) each for the maximum volume of inhalation and exhalation (F29~F38). Also, the Armband provides heat flux (F39), skin temperature (F40), near-body temperature (F41), and GSR (F42).

Stress recognition during physical activity

For the given segments of extracted features, we formulate a basic classification problem where each segment of data corresponds to a binary state: *stress* or *no-stress*. In order to model the complex relationship between physiological features and stress, we use two probabilistic techniques: the naïve Bayes classifier (NB) and Bayesian networks (BNs). NB is a simple probabilistic model based on Bayes' theorem and independence assumptions of features. Despite its use of over-simplified assumptions, it often exhibits good performance on many real-world problems and the resulting model can be easily interpreted and understood. As a powerful technique that handles uncertainty in a complex domain, BNs probabilistically model a set of joint probability distributions over variables. While they are also based on Bayes' theorem, BN is not limited in the independence assumption and are capable of representing the complex relationship among variables as a directed acyclic graph, whose nodes correspond to the variables, and arcs to their causal dependency [22]. The probabilistic models in our implementation are based on the SMILE reasoning engine for graphical probabilistic models, from the Decision Systems Laboratory, at the University of Pittsburgh (http://dsl.sis.pitt.edu). For BN, we first learn the structure with the K2 algorithm and then apply the expectation-maximization (EM) algorithm to learn the parameters.

All the continuous variables are discretized into five states {verylow, low, medium, high, veryhigh} where each state has almost the same number of samples as the others, before being used in the probabilistic models. When building a model that recognizes stress, we optimize the model by selecting informative features for the target classification problem. Although we include a number of physiological features already used in other literature, no preliminary or "expert" selection was applied to clarify which physiological features are correlated with stressors or physical activity.

In order to recognize physiological stress during physical activity (or certain levels of intensity), we extend the basic framework (see Fig. 3) as a model (called M1) that takes physical activity into account in stress recognition as shown in Fig. 4. Instead of building a general model (called M2) that recognizes stress in all situations, we build multiple stress models (called M3), each of which is trained with samples collected when the corresponding physical activity is performed. M3s are responsible for managing the influences caused by their corresponding physical activity. This approach first classifies physical activity with movement features and then applies the corresponding stress recognition model by using physiological features.

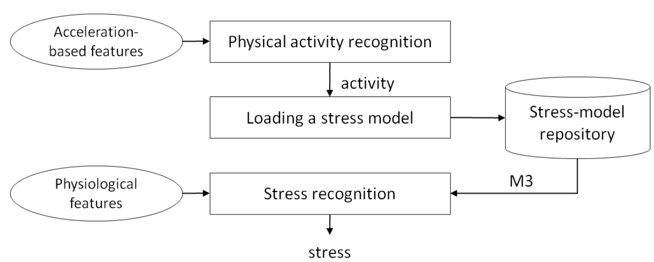

Figure 4. Overview of the physiological stress recognition during physical activity

ANALYSIS & EVALUATION

Evaluation schemes

In order to examine which physiological features are related to stressors or affected by physical activities, we first analyze the features using an independent two-sample *t*-test for each stimulus. Secondly, we build the stress model based on the following evaluation schemes. Three evaluation schemes are proposed to identify the complex relationship among stressors and physical activities with respect to physiological features.

1) *Physical activity* (sitting *vs*. walking *vs*. bicycling): use samples obtained from three physical activities having different intensity, and extract the physiological features impacted by physical activity.
2) *Objective stress* in physical activity: use samples collected when participants are involved in each of three physical activities (sitting, walking and bicycling), and extract promising features showing significant changes to stress. Here, the three stressors are regarded as one class, *stress*, against *no-stress* without any stressor.
3) *Subjective stress* in physical activity: same as objective stress, but instead of using the presence of a stressor to indicate *stress*, we use the participant's subjective rating (5-point Likert scale from 1: no stress and 5: most stressed) with each of the stressors, and no stressor. Any rating except 1 (=no stress) is regarded *as* stress.

For the evaluation, we analyzed the data that includes a period of 2 minutes before the end of each task, and the data was segmented with a window of 30 seconds. The recognition performance of stress models was evaluated using five-fold cross validation, where the data were randomly divided into five parts with four parts used for training and one part for testing. This process was repeated five times. In the following sections on stress modeling, we report the average performance over five folds.

Physiological change by stressors and physical activity

To examine the change in extracted features to each stressor, we analyzed their mean difference across our user population using an independent two-sample t-test. Table 2 shows features that can distinguish our two conditions (no-stressor vs. stressor) for different physical activities, at a level of statistical significance ($p<0.01$).

	Sitting	Walking	Bicycling
Math	F4,F6,F24,F25, F30,F33,F34,F37	F1,F3,F4,F6,F11, F24,F25,F28,F29, F30,F33,F34,F38	F4,F6,F25, F28,F30,F33, F34,F38
Sound	F2,F4,F6,F11, F17,F29,F33,F35, F37,F38	-	F4,F6,F11, F28,F33
Cold	F4,F6,F8,F10,F11 ,F28,F29,F30,F33 ,F35,F36,F37,F38	F2,F4,F5,F6,F8, F11,F33,F35,F38	F11,F33,F38

Table 2. Discriminating features for the stressors

In general, *math* as a mental stressor caused a certain degree of change in physiological features regardless of physical activity. *Sound* did not affect physiological responses as much as the other stressors, and decreased its effect when physical activity was performed. Among the three stressors, *cold* had the strongest influence, causing significant changes in many physiological features when no physical activity was performed. However, fewer features had significant changes for *cold* during walking or bicycling.

In this study, the values of several features such as F4 (average of breathing rate), F33 and F38 (standard deviation of maximum volume of inhalation and exhalation) were significantly changed by stressors across our user population. As shown in Fig. 5(a), although the average breathing rate (F4) increased in the presence of physical activity from 18 to 25 (times/min), it decreased when stressors were administered, especially for *math* and *cold*. However, the breathing rate increased for *sound* when people were sitting and bicycling. Although the three stressors all induced stress, they resulted in different physiological responses due to their distinct effects on the human physiological system.

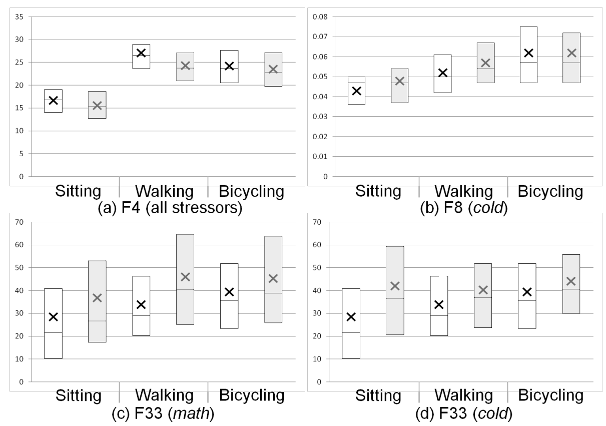

Figure 5. Change of physiological responses by stressors and physical activities (white and grey boxes represent tasks without/with stressors, respectively, and three horizontal lines of boxes: 1st/2nd/3rd quartiles, X: mean value)

Heart rate has been investigated as a stress indicator in previous research [7], but it is also very sensitive to physical activity. In our study, we saw no significant difference in the heart rate because of the influence of physical activity, between situations with and without stressors, over our entire population. However, when just examining the results from *math*, the heart rate for several subjects did change (F24, F25). Skin temperature was somewhat sensitive to *cold*, with its standard deviation (F8) increasing with *cold* (Fig. 5(b)). However, the high intensity physical activity of bicycling, also increased the standard deviation. Many breathing-related features changed over the population in the presence of stressors, especially *math* and *cold*, *e.g.*, the standard deviation of maximum volume of inhalation (F33) increased with these stressors despite the presence of physical activity (Fig. 5(c)(d)).

In addition to the general changes over population, physiological responses are subjective and vary by participant. Fig. 6 shows the distribution of average breathing rate (F4) for two subjects, illustrating the difference across individuals. With respect to physical activity, S7 did not change his breathing rate much, but S12 did. When S7 performs a physical activity, the breathing rate decreased significantly only with *math*. On the other hand, the breathing rate decreased for *cold*, whether or not S12 performs physical activity or not. The varied nature of the physiological responses is the main reason why personalized stress models are required.

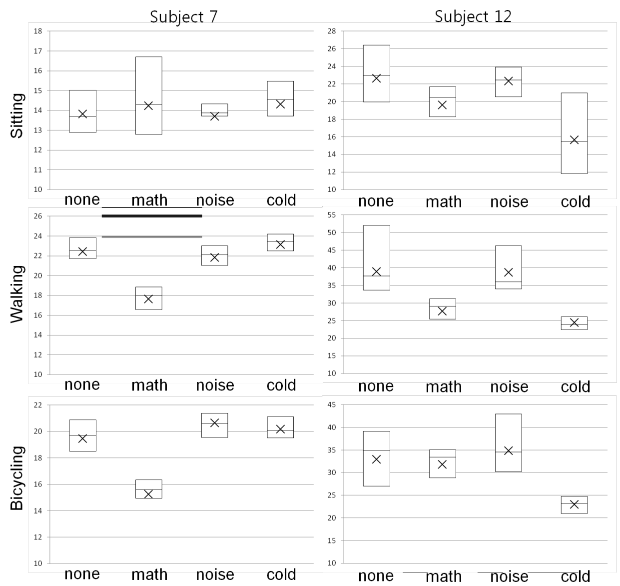

Figure 6. Example of individual difference

As opposed to the stressors which had a lesser influence on physiological responses, physical activity was dominant in causing changes for almost all features. This makes the recognition of stress in the presence of physical activity much more difficult. Table 3 presents the number of features that have significant differences ($p<0.01$) for each stressor and physical activity pair (averaged across all users). As stressors, *math* and *cold* had the largest impact on physiological responses, while physical activity affects the ability to discriminate between stressors across our user population.

p<0.01	Sitting	Walking	Bicycling
Math	4.3	5.1	4.1
Sound	1.5	0.1	1.2
Cold	4.7	3.5	1.6
Avg.	3.5	2.9	2.3

Table 3. Average number of physiological features that show discriminability to stressors with physical activity for each participant

In addition to the analysis on physiological changes by each stressor and physical activity across our entire population, we measured the information gain [23] of features from each participant to evaluate their robustness and usefulness for the personalized stress recognition in the presence of physical activity. We noted how often each was ranked in the top 10 among 42 features for three evaluation schemes as shown in Fig. 7. Influencing a broad channel of physiology such as ECG, skin temperature and breath, physical activity caused changes in many features, but it was relatively less sensitive for some features such as F39 (heat_flux) and F42 (GSR). With respect to objective stress, breath-related features showed higher information gain such as F4, F6, F33 and F36, in addition to F42 and F7 (average of skin temperature). Although few features overlapped between the two schemes of physical activity and objective stress, most features still had an effect on both.

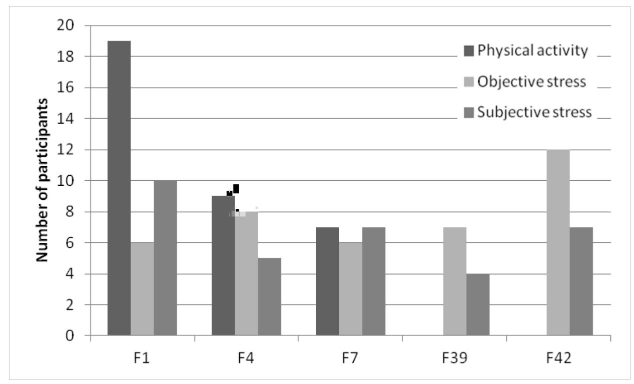

Figure 7. Information gain-based feature analysis for three evaluation schemes. The vertical axis represents the number of participants with the given feature in the top 10.

In contrast with objective stress, subjective stress was more related with physical activity since the participants got stressed not only from the three explicit stressors but sometimes also from some physical activities like bicycling. There are many ECG-related features (F1, F3, F19, F24, and F27) that have high information gain for both physical activity and subjective stress. However, some features (F39 and F42), that had high information gain with respect to objective stress, also obtained high information gain with subjective stress.

Understanding of subjective stress rating

We also examined participants' subjective ratings of stress, in response to stressors and physical activity. We validated our subjective rating with the NASA TLX, and found strong agreement between the two. Fig. 8 shows the distribution of subjective rates and the scores obtained from the NASA TLX, where the values vary with both stressors and physical activity.

Based on the subjective ratings, most subjects reacted to the three stressors, where stress was induced in decreasing order by *cold*, *math*, and *sound*, similar to the result from the changes in physiological responses. Physical activity also influenced the ratings for stress, *e.g.*, less stress was induced by *sound* when an intense physical activity was performed. In general, participants rated 68% of the 48 tasks (4 days × 3 physical activities × (3 stressors + no stressor)) as stressful (ratings>1). Three participants were exception with S01 and S09 only reporting stress for 18% of tasks and S18 reporting stress for 98% of tasks.

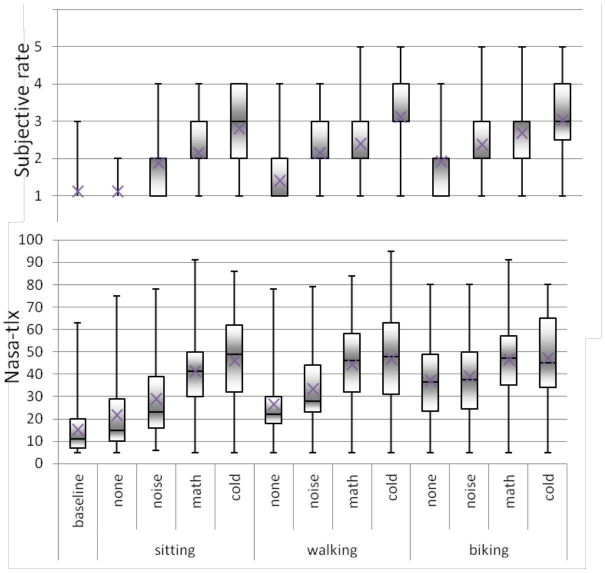

Figure 8. Analysis of subjective rates with the NASA TLX

Recognition of stressors in physical activity

In order to evaluate the applicability of stress models across situations with/without physical activity, we first built the model for situations with no physical activity (M3 for sitting) and applied it to other situations as shown in Fig. 9. It can model stress in the sedentary setting but, as expected, failed to achieve similar performance in the presence of physical activity, resulting in an accuracy drop of 14%.

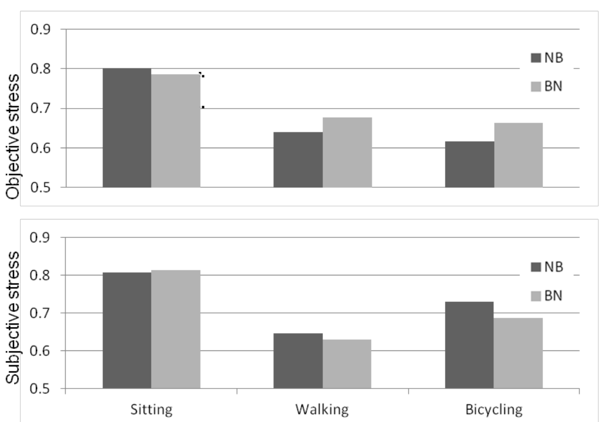

Figure 9. Effect of physical activity on stress recognition

Using our first evaluation approach, we evaluated our recognition of physical activity using movement features (average, standard deviation and median values were calculated from acceleration channels in Table 1) as shown in Fig. 10. Although there are individual differences, both NB and BN achieved high accuracy, around 97% on average.

Using our second evaluation approach, we measured the performance of objective stress recognition, between M1 (the stress model that uses physical activity to recognize stress) and M2 (the stress model that does not use physical activity to recognize stress) as shown in Fig. 11.

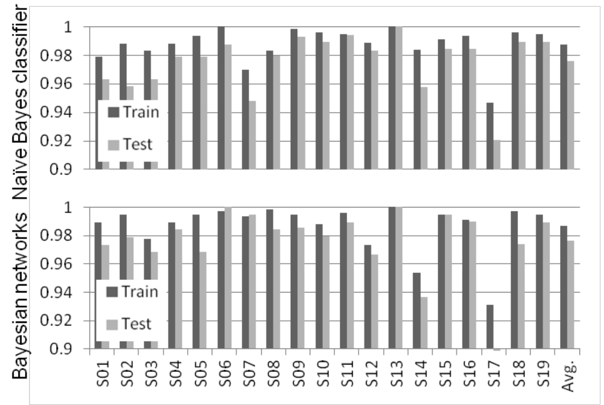

Figure 10. Physical activity recognition

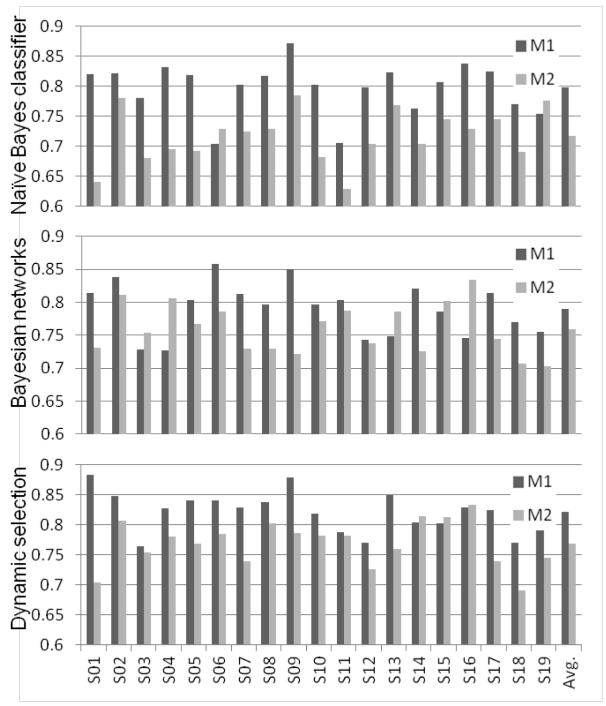

Figure 11. Objective stress recognition

On average, M1 results in accuracies of 79.8% (NB),79% (BN) and 82.1% (DY, in which NB or BN is dynamically selected based on their validation accuracy on training samples) which are significantly better (t=7.2, p<0.001; t=2.3, p<0.02; and t=4.9, p<0.001) than M2 which has accuracies of 71.8%, 76% and 76.9%, respectively. Maximum improvements of 13.6% (NB), 12.9% (BN) and 18% (DY) were obtained for S04, S09 and S01 when we considered physical activity in modeling stress. Inaccurate recognition of physical activity did not degrade the overall performance of stress recognition.

Using our third evaluation approach, we measured the performance of subjective stress recognition, as shown in Fig. 12. M1 has accuracies of 86.4% (NB), 87.4% (BN) and 87.8% (DY) which are 4.3%, 9.4% and 4.1% higher than for M2 (t=7.4, p<0.001; t=4.5, p<0.001; t=5.1, p<0.001). The maximum improvements are 13% (S11 with NB), 19% (S13 with BN) and 12% (S05 with DY).

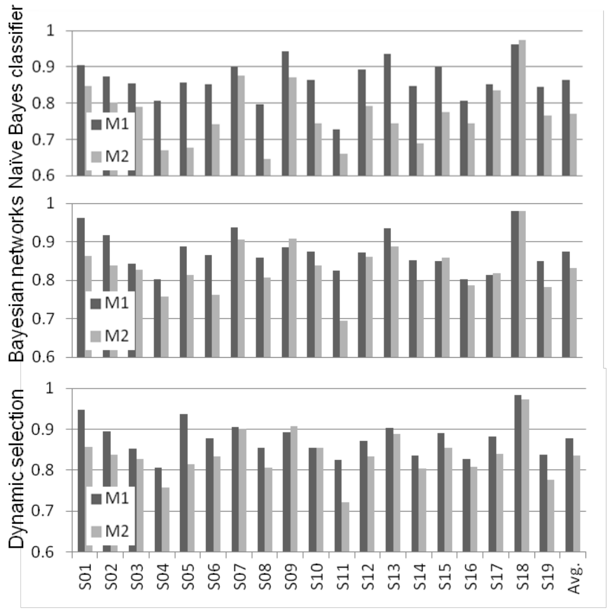

Figure 12. Subjective stress recognition

Subject	M3			M1	M2
	Sitting	Walking	Bicycling		
S01	95.0%	93.8%	95.2%	94.7%	85.7%
S02	85.8%	84.6%	98.5%	89.5%	83.7%
S03	72.1%	84.2%	100.0%	85.4%	82.8%
S04	70.1%	81.0%	88.8%	80.6%	75.8%
S05	95.6%	90.4%	96.9%	93.7%	81.5%
S06	87.7%	82.2%	96.2%	87.7%	83.4%
S07	76.5%	95.3%	100.0%	90.6%	90.1%
S08	88.8%	79.7%	87.7%	85.4%	80.7%
S09	91.8%	91.8%	86.1%	89.3%	90.7%
S10	81.3%	84.5%	90.6%	85.4%	85.4%
S11	78.2%	88.7%	80.3%	82.6%	72.1%
S12	79.2%	85.3%	100.0%	87.1%	83.3%
S13	77.7%	93.0%	100.0%	90.4%	88.7%
S14	79.0%	89.4%	84.2%	83.6%	80.5%
S15	84.5%	87.7%	96.9%	89.1%	85.4%
S16	64.0%	89.0%	96.9%	82.8%	80.8%
S17	88.3%	84.4%	93.8%	88.3%	84.1%
S18	95.2%	100.0%	100.0%	98.4%	97.4%
S19	82.8%	78.1%	93.6%	83.8%	77.6%

Table 4.Subjective stress recognition during physical activity (dark gray cell: M1 < M2, light gray cell: M3 < M2, note that M1 is better than M2 in all other cases)

Table 4 presents the result of subjective stress recognition (DY) for each physical activity. With a 97% accurate physical activity recognizer, M1 performs better than M2 for almost all users (S09 is an exception). This validates the hypothesis, that, in general, it is better to build stress models separately for each physical activity since the variation in physiological changes increases with physical activity.

The predictive ability of M1 is dependent on the performance of M3s (the stress models for recognizing stress in the presence of each physical activity) built for each physical activity. If any specific stress recognition model M3 performs poorly in a certain situation, a more accurate general stress recognizer may not be achievable. In the case of S09 (in dark gray cells), M3s had much lower accuracy for bicycling compared to M2. Because of this, M1 was inferior to M2 although has it has relatively high accuracy for the other activities. On the other hand, for some participants (*e.g.*, S03 and S04), M1 produces better overall results even though it has low accuracy in particular situations (light gray cells in Table 4).

Discussion

As already stated, physiological responses are highly subjective and sensitive to various factors such as mental or physical condition, activity and environment [7]. For instance, in addition to physical activity and stress, weather or the time of day may cause great variations in the physiological system. For accurate detection of stress, a model of the complex relationship that takes into account these various factors along with day-to-day and individual differences [16], is required.

Besides the fundamental issue of recognition, there are technical issues in physiological stress modeling that need exploration, such as 1) gathering high quality data, 2) developing or applying sophisticated techniques, and 3) evaluating systems in the real environment.

Due to physical activity that requires participants to move, the body-worn sensors may have noisy signals, or even have no signal as the sensors lose contact with the body. A mechanism to manage the data quality is required in such settings – we used time-consuming manual effort. It is also important to associate the sensory data with real stress, yet it is often difficult to obtain ground truth stress labels [24].

Although we applied probabilistic methods to stress recognition in this study, the performance could still be improved with other machine learning techniques like support vector machines. Also, a sophisticated technique like ensembling could be useful for modeling the complex relationship in sensory data with multiple features to improve reliability and accuracy [25].

Finally, we need to extend our protocol of data collection to natural environments [5,26]. As an initial stage of research attempting a systematic investigation into the structure of physiological stress relating to physical activity, we scoped our experimental design. The type and number of stressors and physical activities in the present data were inevitably restricted, as well as data collection being performed in a controlled setting. The results reported in this paper should be generalized by verifying our approach in a longitudinal field study that does not restrict any data collection process in natural environments. To what extent context independence in stress models can be found under varying stimulus conditions in natural environments will be an important topic for future research.

CONCLUSION

In this paper, we explored the influence of physical activity, as an important confound occurring commonly in real situations, on stress recognition with physiological responses, and recognized stress by modularizing stress models with respect to physical activity, instead of building a general stress model working in any physical situation. With three activities having different levels of physical intensity, we applied three stressors to induce stress, obtaining both objective and subjective measures of stress. A wide range of physiological features were examined, and we achieved stress recognition accuracies between 82% (objective) and 87% (subjective), improvements of more than 5-10% for many participants when compared to the approach without considering physical activity. At the same time, however, some challenging issues remain to generalize the results reported in this paper. More practical and realistic situations of stress in the presence of physical activity should be examined for stress recognition through a longitudinal field study. Also, we will incorporate more sophisticated algorithms like support vector machines to manage the increasing complexity of modeling a number of physiological features against various factors in the real environment.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Number: CNS-0910878, funded under the American Recovery and Reinstatement Act of 2009 (Public Law 111-5).

REFERENCES

1. The Stress of Life, Hans Selye, New York: McGraw-Hill, 1956.
2. http://www.medicinenet.com/stress/article.htm#what
3. W. Liao, W. Zhang, Z. Zhu, Q. Ji, and W. Gray, "Toward a decision-theoretic framework for affect recognition and user assistance," *Int. J. of Human-Computer Studies*, vol. 64, no. 9, pp. 847-873, 2006.
4. D. Mroczek and D. Almeida, "The effect of daily stress, personality, and age on daily negative affect," *J. of Personality*, vol. 72, no. 2, pp. 355-378, 2004.
5. K. Plarre, A. Raij, S. Hossain, A. Ali, M. Nakajima, M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," In *Proc. 10th Int. Conf. on Information Processing in Sensor Networks (IPSN)*, pp. 97-108, 2011.
6. L. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: an exploratory study," *Int. J. of Human-Computer Studies*, vol. 67, no. 10, pp. 870-886, 2009.
7. J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156-166, 2005.
8. F. Nasoz, C. Lisetti, and A. Vasilakosm "Affectively intelligent and adaptive car interfaces," *Information Science*, vol. 180, no. 20, pp. 3817-3836, 2010.
9. L. Ming, V. Rozgić, G. Thatte, L. Sangwon, A. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan, "Multimodal physical activity recognition by fusing temporal and cepstral information," *IEEE Trans. Neural and Rehabilitation Systems Engineering*, vol. 18, no. 4, pp. 369-380, 2010.
10. F. Wilhelm, M. Pfaltz, and P. Grossman, "Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion," *Interacting with Computers*, vol. 18, no. 2, pp. 171-186, 2006.
11. V. Kolodyazhniy, S. Kreibig, J. Gross, W. Roth, and F. Wilhelm, "An affective computing approach to physiological emotion specificity: toward subject-independent and stimulus-independent classification of film-induced emotions," *Psychophysiology*, vol. 48, no. 7, pp. 908-22, 2011.
12. B. Hughes, "Self-esteem and changes in heart rate during laboratory-based stress," *Psicológica*, vol. 24, vol. 1, pp. 79-91, 2003.
13. R. Rylander, "Physiological aspects of noise-induced stress and annoyance," *J. of Sound and Vibration*, vol. 277, no. 3, pp. 471-478, 2004.
14. J. Stocks, N. Taylor, M. Tipton, and J. Greenleaf, "Human physiological responses to cold exposure," *Aviation Space and Environmental Medicine*, vol. 75, no. 5, pp. 444-457, 2004.
15. A. Sierra, C. Avila, J. Casanova, G. Pozo, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Trans. Industrial Electronics*, vol. 58, no. 10, pp. 4857-4865, 2011.
16. R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, 2001.
17. J. Kim and E. Andre, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067-2083, 2008.
18. F. Agrafioti, D. Hatzinakos, and A. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affective Computing*, vol. 3, no. 1, pp. 102-115, 2012.
19. D. Wu, C. Courtney, B. Lance, S. Narayanan, M. Dawson, K. Oie, and T. Parsons, "Optimal arousal identification and classification for affective computing using physiological signals: virtual reality stroop task," *IEEE Trans. Affective Computing*, vol. 1, no. 2, pp. 109-118, 2010.
20. S. Hart, "NASA-task load index (NASA-TLX); 20 Years Later," In *Proc. the Human Factors and Ergonomics Society*, pp. 904-908, 2006.
21. P. Rainville, A. Bechara, N. Naqvi, and A. Damasio, "Basic emotions are associated with distinct patterns of cardiorespiratory activity," *Int. J. of Psychophysiology*, vol. 61, no. 1, pp. 5-18, 2006.
22. R. Neapolitan, *Learning Bayesian Networks*. Prentice Hall, 2004.
23. S. Kullback and R. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79-86, 1951.
24. F. Wilhelm and P. Grossman, "Emotions beyond the laboratory: theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment," *Biological Psychology*, vol. 84, no. 3, pp. 552-569, 2010.
25. J.-H. Hong and S.-B. Cho, "The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming," *AI in Medicine*, vol. 36, no. 1, pp. 43-58, 2006.
26. J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris, "Out of the lab and into the fray: towards modeling emotion in everyday life," *Pervasive 2010, LNCS*, vol. 6030, pp. 156-173, 2010.