

Stress Classification for Gender Bias in Reading

Nandita Sharma and Tom Gedeon

Research School of Computer Science,
Australian National University,
Canberra, Australia

{Nandita.Sharma, Tom.Gedeon}@anu.edu.au

Abstract. The paper investigates classification of stress in reading for males and females based on an artificial neural network model (ANN). An experiment was conducted, with stressful and non-stressful reading material as stimuli, to obtain galvanic skin response (GSR) signals, a good indicator of stress. GSR signals formed the input of the ANN with stressed and non-stressed states as the two output classes. Results show that stress in reading for males compared to females are significantly different ($p < 0.01$), with males showing different patterns in GSR signals to females.

Keywords: classification, artificial neural networks, galvanic skin response, gender stress, reading.

1 Introduction

The term, *stress*, was first coined by Hans Selye, who defined stress as “the non-specific response of the body to any demand for change” [1]. Stress is defined as the body’s psychological and physical reaction or response to the imbalance between demands and the resources available for a person. It is known as a natural alarm, resistance and exhaustion [2] system for the body to prepare for a fight or flight response to protect the body in case of threats and make the body adapt to changes. Stress has been identified as a serious and growing issue adversely impacting both individuals and society. It is widely accepted that stress has the potential to cause a variety of chronic illnesses ranging from cardiovascular diseases, diabetes and even some forms of cancer. Societies are affected by the stress problem due to large proportions of their citizens facing stress, which is resulting in high economic costs, especially in developed countries [3, 4]. Stress research has a wide range of potential applications including the capacity to improve personal operations, learning and increase work productivity [5], making it an interesting area of research in a variety of fields including posing a technical challenge to Computer Science. A number of computational techniques have been applied to define stress objectively and build simplistic stress models based on techniques such as Bayesian networks [6], support vector machines [7], and decision trees [8].

Due to complexity in its definition, stress cannot be measured directly and objectively but can be modeled in terms of other measures, including galvanic skin response (GSR) [9], or heart rate variability [10]. GSR, also known as *skin*

conductance or *electrodermal activity response*, is reliable as a measure of stress [10-12] and it is used in this paper to classify stress. When an individual is under stress, skin conductance is increased [6], which may result in moisture on the surface of the skin, or in a more extreme case leading to sweating. The increase in skin conductance is recorded as a change in GSR. GSR has been used as a measure in a biofeedback system to help individuals lower their stress levels. The system had Biopac sensors to take GSR readings and a competitive racing game component that induced stress [13].

Artificial Neural Networks (ANN) has been used for modeling stress but is at early stages of research. Physiological signals have been used as inputs to ANN for classifying mental workload with good accuracy rates [14]. Some emotions, such as fear and anger, are symptoms of stress, which have been differentiated using ANNs with GSR as inputs. This paper presents an ANN for determining whether there is a gender based difference in stress for reading. It uses GSR data obtained from a reading experiment comprising of *stressed* and *non-stressed* reading to obtain data sets for male, female and a combined data set, develops an ANN classifier for the data sets, and then provides an analysis of the results. An experiment was conducted, with stressful and non-stressful reading material as stimuli, to obtain galvanic skin response (GSR) signals, which is a good indicator of stress. GSR signals formed the input of the ANN with stressed and non-stressed states as the two classes. Three ANNs were built to investigate whether there is gender bias in stress. Results from the experiments show that stress in reading for males and females are significantly different ($p < 0.01$). The paper evaluates the classifiers, validates them, analyses classification rates, concludes the findings, and proposes further research.

2 Reading Experiment

Undergraduate Computer Science students over the age of 18 years old were recruited as experiment participants (after obtaining Ethics Approval from the Australian National University Ethics Committee). The participant group for the investigating gender bias experiment was made up of 10 males and 10 females with age between 18 and 24 years. Each participant had to understand the requirements of the experiment from written experiment instructions with the guidance of the experiment instructor before they filled in the experiment consent form. Afterwards, GSR electrodes were attached to the participant. The instructor notified the participant to start reading, which triggered a sequence of text paragraphs. After finishing the reading, participants had to do an assessment. An outline of the process of the experiment for an experiment participant is shown in Fig. 1.



Fig. 1. Overview of the experiment process

2.1 Experiment Set Up

Experiment participants had to silently read *stressed* and *non-stressed* types of paragraphs. Stressed paragraphs had stressful content, whereas the non-stressed paragraphs had content that created an illusion of meditation or soothing environments. There were three paragraphs for each type of paragraph and each paragraph was about the same length, approximately 360 words. Participants were told that they would be monitored on the way they interacted with text while they read and then had to do an assessment, based on what they read, after the reading.

The paragraphs were displayed on a 1050 x 1680 pixel Dell monitor. For consistency, paragraphs were displayed on the screen for 60 seconds and positioned at the same location of the computer screen for each participant. Each line of the paragraph had 70 characters.

Biopac GSR100C equipment was used to obtain GSR signals at 1 kHz. Disposable EL507 GSR electrodes were used, which were placed on the first (or index) and forth (ring) fingers of the participant's left hand.



Fig. 2. The experiment was conducted in a room with a consistent environment for participants. A computer monitor displayed the reading text. The participants wore GSR equipment on their left hand, which they rested on the arm rest while reading.

3 ANN Classifier for Determining Gender Bias

An ANN is inspired by biological neural networks with characteristics for learning and reacting. It is made up of interconnected processors, known as *artificial neurons*, connected by weighted links that pass signals between neurons. Feed-forward ANNs were developed, trained and tested for testing the null and alternative hypotheses:

- H1.** Null: There is no difference in GSR patterns for females and males during reading stressed and non-stressed material.
- H2.** Alternative: There is a difference in GSR patterns for females and males during reading stressed and non-stressed material.

A three-layer (plus the input layer) ANN was trained on a data set obtained from the experiment to test the hypotheses. Each of the ANNs developed was trained on one of the three data sets:

- DF.** Female data set – standardised GSR data for all female participants
- DM.** Male data set – similar to the female data set, but the data set contained data for male participants only
- DFM.** Male and Female data set – composed of the male and female standardized GSR data sets

An ANN was developed by training it on the male and female data set (DFM) to test H1. Classification rates for participants were obtained by swapping two females and males at a time from the training and testing sets at a time. In order to test H2, two ANNs were developed. Each ANN was trained on either the female data set (DF) or the male data set (DM). Classification rates of a participant was determined on the ANN trained using the opposite gender e.g. female participants were classified using the ANN that was trained on DM.

The Sigmoid function was chosen as the activation function to limit the output values of the ANN within the range of 0 (not-stressed) and 1 (stressed). The ANNs were developed using the Matlab Neural Networks Toolbox and trained using the Levenberg-Marquardt back-propagation algorithm.

In order to minimize individual bias, the GSR signals were standardized to a range within 0 and 1 using Equation (1) for each timestamp $T(i)$.

$$\text{Standardised_GSR}_{T(i)} = (\text{Raw_GSR}_{T(i)} - \text{Minimum_GSR}_{T(i)}) / \text{RangeGSR}_{T(i)} \quad (1)$$

The size of the raw GSR data set for a participant reading a particular type of paragraph was 180,000, which was a result of a sampling rate of 1 kHz. This data set is not feasible to extract suitable features. Therefore, a maximum value was obtained for every 1000th interval to reduce the data set size to 180. This gave a data set a granularity level of one second. Subsequently, a running maximum algorithm was used with a window of five seconds. GSR after each second, SGSR_t , was defined as the maximum $\text{Standardised_GSR}_{T(i)}$ from a window of size five as described in Equation (2).

$$\text{SGSR}_t = \max\{\text{Standardised_GSR}_{T(i)}\} \text{ where } i \leq j < i+5 \quad (2)$$

Each SGSR_t formed an input to the ANNs. Research has showed that physiological responses affected by stress appears a short time after an individual has been exposed to stressful stimulus [10, 15]. The experiment data, depicted in Fig. 4, shows an example of the claim. Also, a participant was given 60 seconds to read a paragraph, so as a consequence, SGSR_t values over a 30 second time interval was used to develop a sample in the data set for the ANNs. An input list, I_k , for a sample in the ANN data set is defined in Equation (3).

$$I_k = [\text{SGSR}_k, \text{SGSR}_{k+1}, \text{SGSR}_{k+2}, \dots, \text{SGSR}_{k+29}] \quad (3)$$

Classification of stress for reading is non-linear and the input fed to the ANN was discretised and segmented. The topology of the ANN was obtained empirically based

on capturing prominent characteristics appearing in the data (described in Section 4) and experimental trials. The characteristics were observed over approximately 30 second time intervals so GSR data over 30 second time intervals formed inputs to the ANN, thus the ANN had 30 input neurons. Rules of thumb were incorporated in determining the number of neurons in the hidden layers of the ANN. The first hidden layer was designed so that the number of neurons was no more than twice the number of neurons in the input layer and the number of neurons in the second layer was less than the square root of the number of neurons in the first hidden layer. Accordingly, 50 neurons were allocated to the first hidden layer, 7 neurons were in the second hidden layer and one neuron in the final layer. Moreover, the topology was defined so that the ratio of free parameters to the number of training patterns was at most 3.

4 Results and Discussion

The GSR data for reading stressed and non-stressed material obtained from the reading experiment for each participant is shown in Fig. 3. By observation, prominent characteristics of stress, spikes or short periods with high GSR values, in the data

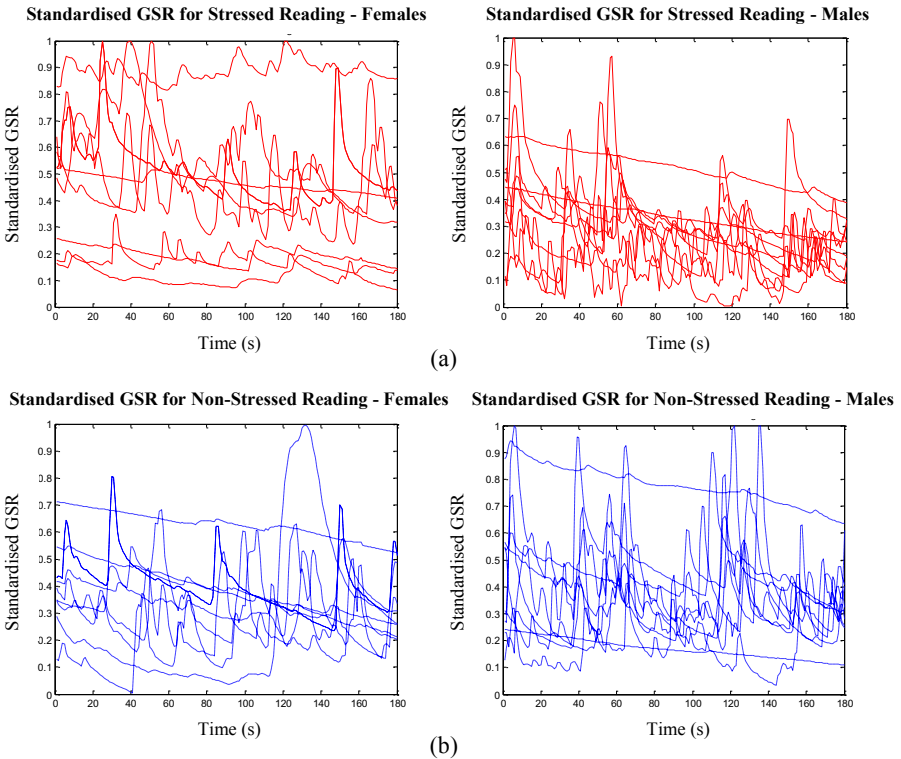


Fig. 3. Standardised GSR values were obtained for participants in the reading experiment. Each participant read three stressed and three non-stressed paragraphs in consecutive order for stressed and non-stressed readings respectively. (a) Standardised GSR values for male and female data sets while stressed reading (b) Standardised GSR values for male and female data sets while non-stressed reading.

generally span more than 20 seconds. Accordingly, the inputs for the ANN were defined to include this span, but not a span as large as the time period given to read a paragraph, which was 30 seconds. This generated 6,040 data samples, which were used to train the ANN.

A 10-fold cross-validation method was used to evaluate the ANNs trained on DF, DM and DFM. The ANN that was trained on the female data set performed the best. The ANN that was trained on the male data set did not perform much worse, however, the ANN which was trained on the training set with both the male and female data performed the worst. This variation suggests that characteristics in male and female data sets may not be similar. The performance values of the ANN on the three different training sets are compared in Table 1.

Table 1. Statistics for 10-fold cross-validation on the ANN for the reading data sets

Data set	Correct rate	Sensitivity	Specificity
Female	91%	91%	91%
Male	81%	77%	84%
Male and Female	70%	71%	70%

Average classification rates for correct classification for each of the participant samples were calculated over a set of five trials. For each trial, the ANN was trained using the training set, which varied for the hypotheses that were tested. The classification rates for stress for male and female participants on the ANN trained on DFM is shown in Table 2.

Table 2. Average correct classification rates for samples tested on the ANN trained using samples of both the genders

Participant ID	Correct classification rates for female samples	Correct classification rates for male samples
1	68%	50%
2	33%	34%
3	42%	62%
4	71%	60%
5	33%	64%
6	55%	58%
7	76%	56%
8	51%	15%
9	78%	14%
10	31%	44%

The Student's t-Test gives a p-value of 0.004 ($p < 0.01$), which shows that we have strong evidence that the null hypothesis, H_1 , does not hold. As a result, there is no evidence that suggests that there are no differences in GSR patterns for females and males during reading stressed and non-stressed material, which suggests investigating existence of difference in GSR patterns.

The classification rates for stress for each female participant on the ANN trained on DM and for each male participant on the ANN trained on DF are given in Table 3 to determine whether there is a difference in patterns in GSR for males and females.

Table 3. Average correct classification rates for samples tested on the ANN trained using samples of the opposite gender

Participant ID	Correct classification rates for female samples	Correct classification rates for male samples
1	56%	54%
2	62%	24%
3	34%	45%
4	45%	54%
5	27%	43%
6	43%	54%
7	38%	60%
8	8%	10%
9	47%	34%
10	38%	35%

The classification rates for male and female samples have a p-value of 0.16 ($p > 0.1$), which means that there is strong evidence that males show different characteristics and patterns in GSR signals for stress to females. Thus, H2 hypothesis holds. Classification results in Table 2 and Table 3 and statistical analyses show that males and females respond to stress in reading with different GSR signals.

5 Conclusion and Future Work

ANN was used to classify stress in reading for females and male. We showed that stress in males and females are significantly different ($p < 0.01$) i.e. gender has influence in the separability of GSR characteristics. The results are consistent with previous studies [16]. Future possible extension to the reported work includes modeling GSR features for males and females on other types of classifiers (e.g. SVM) to generalize reported patterns, investigating discrepancies within the male data set and developing a method for obtaining an improved classifier for a male stress data set. In order to obtain a classifier with a higher classification rate, inputs of an ANN could be defined in terms of multiple primary measures of stress e.g. a combination of GSR and HRV measures.

References

1. Selye, H.: Confusion and controversy in the stress field. *Journal of Human Stress* 1(2), 37–44 (1975)
2. Hoffman-Goetz, L., Pedersen, B.K.: Exercise and the immune system: a model of the stress response? *Immunology Today* 15(8), 382–387 (1994)
3. The-American-Institute-of-Stress. America's No. 1 Health Problem - Why is there more stress today? (August 5, 2010), <http://www.stress.org/americas.htm>
4. Lifeline-Australia, Stress Costs Taxpayer \$300K Every Day (2009)
5. Kompier, M., Cooper, C.L.: Preventing stress, improving productivity: European case studies in the workplace. Routledge (1999)

6. Liao, W., et al.: A real-time human stress monitoring system using dynamic bayesian network (2005)
7. Dou, Q.: An SVM ranking approach to stress assignment. University of Alberta (2009)
8. Zhai, J., Barreto, A.: Stress recognition using non-invasive technology (2006)
9. Shi, Y., et al.: Galvanic skin response (GSR) as an index of cognitive load. ACM (2007)
10. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems* 6(2), 156–166 (2005)
11. Labbé, E., et al.: Coping with stress: the effectiveness of different types of music. *Applied Psychophysiology and Biofeedback* 32(3), 163–168 (2007)
12. Ferreira, P., et al.: License to chill!: how to empower users to cope with stress. ACM (2008)
13. Bersak, D., et al.: Intelligent biofeedback using an immersive competitive environment. In: 2001: Paper at the Designing Ubiquitous Computing Games Workshop at UbiComp. (2001)
14. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psycho physiological measures and artificial neural networks. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45(4), 635 (2003)
15. Rimmele, U., et al.: Trained men show lower cortisol, heart rate and psychological responses to psychosocial stress compared with untrained men. *Psychoneuro endocrinology* 32(6), 627–635 (2007)
16. Wang, J., et al.: Gender difference in neural response to psychological stress. *Social Cognitive and Affective Neuroscience* 2(3), 227 (2007)