Stress detection during job interview using physiological signal

Ali Afzalian Mand[1], Md. Shohel Sayeed[1], Md. Jakir Hossen[2], Muhammad Amer Ridzuan bin Zuber[1]
[1]Department of Information Security, Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia
[2]Department of Mechanical Engineering, Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia

Article Info

ABSTRACT

A job interview can be challenging and stressful even when one has gone through it many times. Failure to handle the stress may lead to unsuccessful delivery of their best throughout the interview session. Therefore, an alternative method which is preparing a video resume and interview before the actual interview could reduce the level of stress. An intelligent stress detection is proposed to classify individuals with different stress levels by understanding the physiological signal through electrocardiogram (ECG) signals. The Augsburg biosignal toolbox (AUBT) dataset was used to obtain the state-of-art results. Only five selected features are significant to the stress level were fed into neural network multi-layer perceptron (MLP) as the optimum classifier. This stress detection achieved an accuracy of 92.93% when tested over the video interview dataset of 10 male subjects who were recording the video resume for the analysis purposes.

Corresponding Author:

Ali Afzalian Mand
Faculty of Information Science and Technology, Multimedia University
Ayer Keroh Lama Street, Melaka 75450, Malaysia
Email: ali.mand@qnet.net

1. INTRODUCTION

A job interview can be tough even if one has gone through it many times. Becoming excessively nervous during the interview preparation could lead to a high level of stress and anxiety which makes one's life more miserable plus, the competition for placement is very high. Some might fail to control their stress and causes them to unsuccessfully deliver their best version throughout the interview session. Other than reducing performance in interview session, physical symptoms such as emotional overwhelm, absenteeism, and misplaced anxiety are also indicating a person is in a stress condition [1]. Therefore, an alternative method which is preparing a video resume and interview prior to actual company interview could reduce the stress. In the literature by [2], the level of calmness can be indicated based on the low heart rate variability (HRV) while any potential mental stress and frustration can be seen during the high rise of HRV. Not only that, HRV is also evidence to be taken into consideration as a measurement that reflects heart activities and overall autonomic health [3]. Plus, HRV alone is more sensitive than the heart rate in measuring the stress level [4]. As it can be observed, the decision making in life not just rely on the requirement and conditions but also on emotional states which basically are based on the experience. Trimmer *et al.* [5] claimed that "human emotion can be considered as the fluctuating dispositions to make a positive and negative evaluation". Many researchers had put their effort to study the human affective state and asserted that valence (positivity/negativity) and arousal (degree of mental alertness or activation) are the two key dimensions in human emotion. While facial expression can be represented in three subdivision which are beginning, apex and end, by the facial magnitude [6], the study by Kaul *et al.* [2] has also divided the emotion into four quadrants labelled by arousal as the y-axis and valence as the x-axis as shown Figure 1. Despite stress has always been defined as a negative situation, this has overshadowed the bright side of the stress that can uplift

the level of productivity and increase the quality of life, while the uncontrolled high level of stress (distress) is the one that causes health problems and reduces the working performance [7].
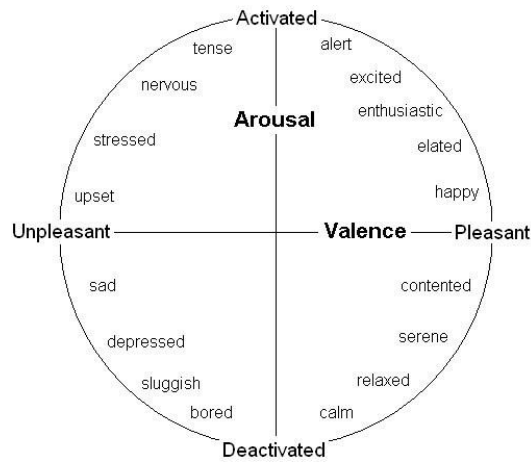

Figure 1. The 2-D representation of emotional state [5]

This has also been studied earlier by [8] who proposed an empirical relationship between arousal and performance which explains the increment of performance is related to the level of physiological or mental arousal but only up to an optimal arousal point as shown in Figure 2. Not only that, but the journey also to obtain the optimal arousal level for an optimum performance needs a series of actions which allow the body to adapt to fit, called as habituation which refers to the reduction in physiological responses elicited by repeated exposures to a repeated homotypic stressor. Biosignals are one of the most accurate input parameters for emotion recognition as it possesses the ability to be both robust and unobtrusive against various environmental situations which other emotion recognition inputs lack it. In this study, stress detection is done by analyzing the heart electrical signal as one of the biosensors known as the electrocardiogram signal (ECG). Also, in another study proposed the idea to use ECG for stress recognition by requiring the parameters extracted from the raw intractable long-term HRV and able to optimize by only using the ultra-short-term raw ECG waveform [9].
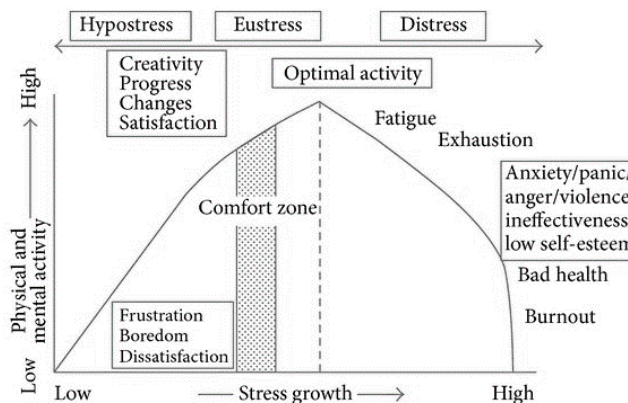

Figure 2. Human response to stress curve [10]

An ECG signal as shown in Figure 3, is a graph to illustrates the contractile activity which refers to the beating and strength of the electrical response of the heart. Briefly, it consists of six main components which are the P-wave, the PR-interval, the QRS-complex, the ST-segment, the QT-interval, and the T wave. One of the components, P-wave is the effect of the depolarization of right and left atria which was used as a feature that gave 87% of recognition rate to differentiate stress level [1]. Understanding each component

function will increase the effectiveness of the ECG signal application. It has been a popular discussion about any possible contribution to the diagnosis of heart diseases, emotion recognition, and health condition. Thus, with these ideas, many researchers have proposed to demonstrate the best method to understand the ECG signal and able to accurately recognize early signs of heart disease in ECG signal to improve the treatment as well as save people's lives. Therefore, this paper presents the issues and challenges related to stress detection where the best ECG signal features for stress detection with high accuracy are not yet clearly justified and there is lack of understanding of the relationship between the stress level at different states and condition. Based on these issues, this research aims to analyze and highlight the selected important features that are strongly related to the stress level and to study the relationship between the different condition and the stress levels. Nowadays, there are advanced progress in applying computing technologies that have significant progress in artificial intelligence.
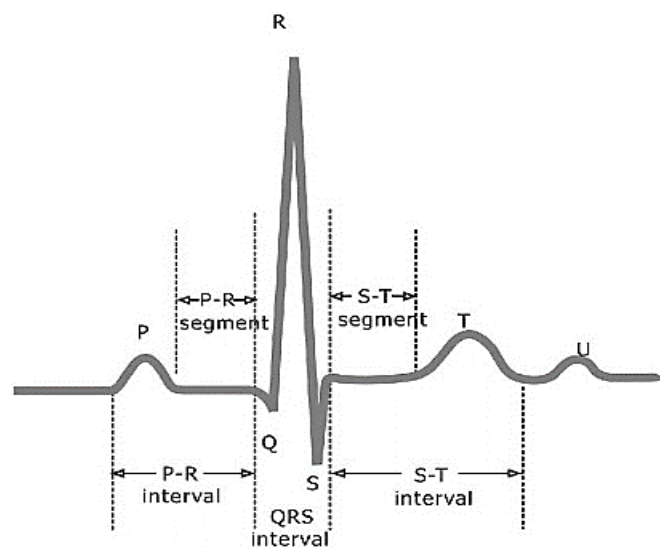

Figure 3. Illustration of ECG signal morphology [10]

       To fill the current gap in literature, we proposed an intelligent stress detection framework aims to classify a stressed individual from a normal person by understanding the physiological signal through suitable biosensors such as ECG signal to understand the situation mentioned above. In summary, the main contribution of this paper can be concluded: i) we collected ECG signals from 10 subjects who are recording the video resume and interview, ii) the ECG signals are preprocessed and selected the five important features significant to the stress level, iii) the neural network multi-layer perceptron (MLP) is used to design the framework of stress detection system, and iv) the proposed framework is analyzed and evaluated with current benchmark datasets and literatures.

       The rest of the paper is organized as the following. In section 1, we discussed the related works for emotion recognition and stress detection using ECG signals. The proposed research method is explained in section 2 with signals pre-processing and classification for stress detection. The analysis and evaluation results are discussed also in section 3. Finally, the paper is concluded with the final findings and futures work in sections 4. "The biosensors could be the first step towards the automatics emotion recognition system by proposing an approach to classify the emotion using multiple biosensors such as electromyography (EMG), skin conductance (SC), blood volume pulse (BVP), ECG, respiration and skin temperature" [5]. Another author [11] has also demonstrated that the ECG and EMG has a strong capability and effectiveness in detecting stress level at high accuracy by inducing multiple stressful environments. In this paper, the dataset was the three positive and negative states under variable arousal level using The International Affective Picture System (IAPS). The authors used a regular set of features values which are running mean, running standard deviation and the slope to replace the raw signals for the classification processes for each biosensor. Focusing on the ECG signal, the signal pre-processing steps were withdrawing the global mean value from the raw signal. Then, the signal undergone low-pass filtering at the frequency of 90 Hz, high-pass filtering at 0.5 Hz and notch filtering at 50 Hz. The important features such as heart rate (HR), HRV and interbeat interval (IBI) between successive heartbeat were calculated. The authors also claimed that HRV is affected by the sympathetic and parasympathetic vagus nerve which appears as a good benchmark for the interim

dominance of one of those signals. In another study, the HRV is mostly studied to be associated with low parasympathetic activity that is detected by an increase in the low frequency and decrease in the high frequency [3]. This paper proposed neural network classifier as the classification method and the classification results were shown in Table 1. The output dictates that the estimation of the valence value is more difficult compares to the estimation of arousal value, however more improvement works could be done in the future.

Table 1. Classifier results [5]

	Arousal		Valence	
Bandwidth	0.1	0.2	0.1	0.2
Correct	89.73%	96.58%	63.76%	89.93%
Wrong	10.27%	3.42%	36.24%	10.07%

Wagner *et al.* [12] dedicated this research to compare and implement a different type of features extractions and classification methods to give a robust output for the emotion detection of four different types of emotions which are anger, sadness, pleasure, and joy. The data for this paper was the signals collected from the four-channel of biosensor attached to the subjects which were the ECG, EMG, SC, and respiration change (RSP) while they were listening to the emotion inducer song that was picked personally. The list of biosensors channel appears to be similar to the [5] but without BVP and skin temperature. Wagner *et al.* [12] proposed two techniques for reducing the dimension of the features. First technique was by excluding a few features from the high dimension features array using analysis of variance (ANOVA), sequential forwards selection (SFS), and sequential backward selection (SBS). The second technique was by extracting a new set of features from the initial set. Figure 4 visualizes the classification error of these features' selection methods. principal component analysis (PCA) and fisher projection were used in the second technique. These dimension reduction methods consider all data in the features array including the noise exist in the features to extract new features set [12]. Figure 5 visualizes the sample output of using the fisher projection. This paper further discussed the three different classification methods which are the k-nearest neighbor (KNN), multilayer perceptron (MLP) and linear discriminant function (LDF). The authors also tested the combinations of different classifiers and features selection methods that finally concluded that the join force between LDF and SFS gives the best result for the recognition of those four emotions at 92.05%, valence at 86.37% and arousal at 96.59%. Not only that, MLP specifically is also being discussed as a good classifier to estimate the length of hidden message in steganography with accuracy around 90.00% which has made this study firmer [13]. Then, they claimed that it is much easier to detect emotion along the arousal axes rather than on the valence axes [12].
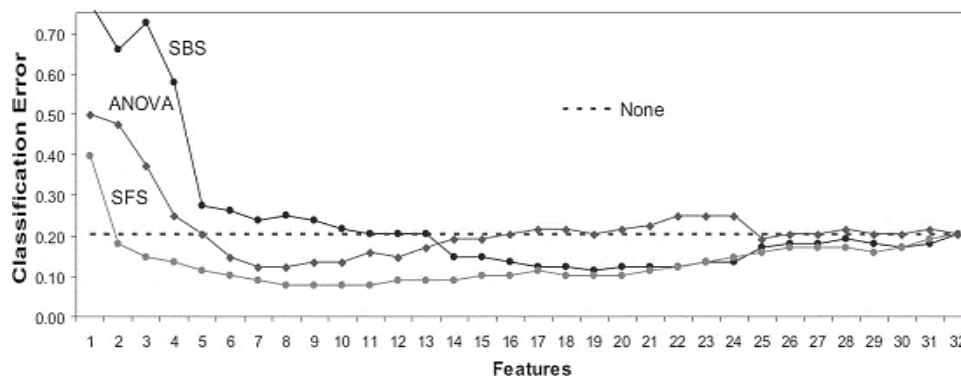

Figure 4. Classification error against features [12]

A stress recognition in working people using the support vector machine (SVM) as a parametric classifier and the KNN as non-parametric classifier was proposed by [14]. The SWELL knowledge work (SWELL-KW) dataset collected from the knowledge workers which is a large-scale multimodal action data was used in this paper. The proposed stress detection system was focusing on the ECG and galvanic skin response (GSR) sensors to extract the desired features which based on the GSR, HRV frequency domain, HRV, and heart rate statistical features. Wagner *et al.* [12] used the same method to get the R peaks intervals,

while the Welch algorithm was used to extract the power spectral density of HRV features. For classification, Sriramprakash *et al.* [14] used two types of features analysis methods which were the individual features analysis and the feature combination analysis. The first method is to analyze each of the features that has the best classification precision and accuracy, then to generate five clusters of features to be tested [14]. The result shows that cluster 5 was the best cluster of dominant features for stress recognition with the classification accuracy of 66.52% using KNN and 72.82% using SVM. The cluster 5 features are mean HR, MAD HR, root mean square of successive differences (RMSSD), average N–N interval (AVNN), standard deviation of the average of N–N intervals (SDANN), low frequency (LF), high frequency (HF), mean number of times an hour in which the change in successive normal sinus, NN intervals exceeds 50 ms (NN50), pNN50, mean GSR, median GSR, STD GSR. Addition to this, not all extracted features are good indicators or candidates to be used for stress detection, some of good surrogates would be mean NN, STD HR, HF when handling with ultra-short HRV waveform [4]. Upon this result, the authors furthered their analysis to improve the classification accuracy by adding new features to the cluster 5 which were the total average power (TAP) including the 2-norm or maximum singular value of ECG heart rate features, the energy of heart rate signal and the energy of heart variability features. Fortunately, the classification result of the improved set of features tremendously increased to 92.75% using the SVM classifier with RBF kernel. Sriramprakash *et al.* [14] concluded the GSR, heart rate, and HRV features contribute the most to the stress recognition. Electrocardiography could be defined as the electrical activity of the heart recording process over time using electrodes placed on the human skin [15]. In this paper, the authors were discussing over the reliable stress information ability to be detected during a stress-induced ECG signal and short HRV signal. The authors also quoted a suggestion stated by the European Society of Cardiology that advised the researchers to consider at least 5 minutes for the heart rate measurement for ECG signal in HRV signal extraction, otherwise, other values under that minimum time will cause uncertainty on the result. The methods and materials to attain the objectives of this paper were explained clearly that the authors will use the HRV signal and ECG signal for stress detection [15]. The proposed methodology started with the ECG signal pre-processing using the discrete wavelet transform (DWT)-based wavelet denoising algorithm that is believed can be applied on any type of physiological signal without the specification of cut-off and sampling frequency [15]. The pre-processed signals were then decomposed into 8 levels to extract the QRS wave. The back-search methodology of maximum and minimum beat detection was implemented at 0.32 s respectively for the missed beat to detect the R peaks and exclude the noisy peaks.
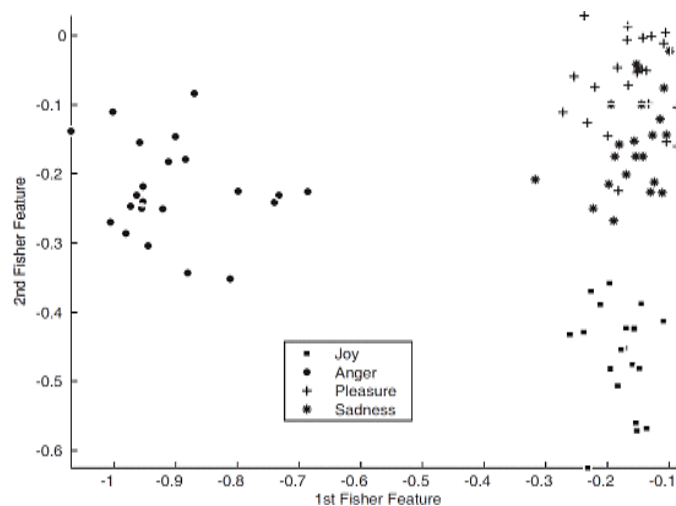

Figure 5. A sample set of features [12]

The pre-processed signals were used for the feature's extraction under time domain (TD) and frequency domain (FD). Athira *et al.* [15] extracted and selected time domain features skewness from the HRV signal and standard deviation for the ECG signal using the Pan-Tompkins algorithm [16]. In conclusion, a simple classification logic was defined by using the threshold value of those two features. The result claimed that "the signal with a value of standard deviation over 0.152 is considered as stress ECG while below the value is normal. Meanwhile, the signal with skewness value below 2.5 is labelled as normal ECG and any signals with skewness value greater than 2.5 are labelled as stress ECG" [15]. The frequency band analysis is to classify and analyze five types of human emotions which are the happiness, disgust, fear,

sadness, and neutral using HRV signals taken from ECG. The data used in this study were the ECG signals collected from 20 different subjects over 10 trials. The emotions were induced by altogether 50 different audio-visual stimuli (video clips) that represent the five emotions [17]. All ECG signals were filtered from the mismatch of electrode impedance, power line frequency, wandering of the baseline signal, and motion artefacts using 3rd order Butterworth filter with 0.002-100 Hz cut-off frequency. The Pan-Tompkins algorithm [16] was used to derive HRV signals from the ECG signals. This method has also been used in the study by [15]. For the statistical feature's extraction, DWT was used with different wavelet functions for LF and high frequency (HF) bands extraction such as db7, sym8, and coif5 [17]. In this study, the LF bands used lies in the range of 0.03-0.12 Hz, while the HF bands lie in the range of 0.12-0.488 Hz which fall below the universal frequency range as proposed by [18], [19]. The range defined for LF is mainly connected to sympathetic activity while the HF range is associated with parasympathetic activity [20].

Murugappan *et al.* [17] also mentioned about having 14 decomposition levels of input HRV signals to extract the wavelet coefficients and ECG features of those sub-bands. However, the very low frequency (VLF) at the frequency of 0.004-0.04 Hz was not included in this study as it is considered as not important in the emotional changes based no HRV signals. This is also supported by [21] in their paper stated that a shorter window was concluded to be reliable to extract HRV features since the 100 second windows and 300 second windows also performed at a similar rate even though some low frequency power might be lost due to shorter window size [21]. Murugappan *et al.* [17] only focus on the most studied statistical features which were the average frequency band power and the standard deviation, and these features were considered as the main features in this study. Next, the classifier of all five emotions used in this study was the KNN and linear discriminant analysis (LDA). The generated results showed that 5NN reached the maximum classification accuracy which better than LDA. The average maximum average classification rates for KNN were 69.754% while LDA was 67.808%. This is because there is a lot of overlapping characteristics in the emotional features that are not able to be differentiated by using linear boundary in the LDA classifier. Plus, LDA only permits the linear or quadratic relationship between the input and output. In features wise, the main features give better classification rate compared to the statistically features. Other than that, the HF itself was reported not being responsive in the classification performance but the frequency range of LF and HF was best for disgust and neutral emotion classification while the standard deviation was the good features for happiness and fear emotion recognition [17]. Murugappan *et al.* [17] claimed that the recognition over the arousal and valance axes is complex due to the individual perception upon the emotional experience over time when faces the same emotional stimuli.

2.　METHOD

This section portrays the whole methodology flow starting from the programing language, data used, signal preprocessing, features extraction and selection, and the classification algorithms. Python3 was used as the main programming language in the whole framework process as it is user-friendly and has high community support. We also used two sets of data, benchmark dataset to evaluate the accuracy of classification engine and a dataset we collected in Multimedia university, Malaysia while the subjects went through a mock video interview session.

2.1.　Benchmark dataset (AUBT dataset)

The benchmark dataset was taken from the Institute of Computer Science, University of Augsburg, Germany under their study on different features extraction and classification methods that was conducted by [12]. It contains a few sets of bio-signal data that were collected from four-channel of a biosensor which is EMG, ECG, SC and RSP. Briefly, EMG recording is produced from the electrical activity in skeletal muscles which can be used to diagnose disorders affecting the connection between nerves and muscles. SC is where skin becomes a good electric conductor when physiologically aroused by internal or external stimuli for a very short time. While the RSP refers to the process of breathing based on the breathing rate. However, this research only focuses on the ECG emotion dataset that has 100 samples of ECG signals recorded from four different types of emotion such as joy, sadness, anger and please. Wagner *et al.* [12] used songs that were personally picked by the subjects to induce a different type of emotions in them. This dataset will be addressed as Augsburg biosignal toolbox (AUBT) dataset after this.

2.2.　Data collection (video interview dataset)

The experimental dataset was self-collected in the Faculty of Business (FOB), Multimedia University, Malaysia. This dataset was used to measure the subjects' stress when compiling a video resume to assess their readiness to accept this new trend of the job application. It contains 110 samples of ECG signal that was collected from 10 subjects under 11 sessions. The subjects' status has been represented in Table 2.

Table 2. Subject requirements

Recruitment of Subjects	Descriptions
Course Program	Faculty of Business final year student
Course	None (random)
Age	21-23 years old
Gender	Male
Health Condition	No heart problem and healthy

2.3. Signal pre-processing

There are two main steps of signal pre-processing in this framework which is missing data handling and bandpass signal filtering. All the not a number (NaN) and empty column exist in the raw signal were replaced by 0 (Null). This problem may cause the analysis to end with wrong inferences about the experimental data [22]. Butterworth bandpass filtering was used to only allow the signal that lies between two specific frequencies which are low and high cut-off frequency. The low cut-off frequency was at 0.5 Hz to remove the frequency sways caused by the movement of the artefact, while the high cut-off frequency was at 90.0 Hz to remove sharp peaks in the signal that are considered as noise [5]. The order for this bandpass was the 3rd order as mentioned in [17].

2.4. Features extraction

In order to extract the ECG features, two predefined Python3 libraries were used for the feature's extraction such as Pyhrv and Scipy. Two domains which are time and frequency domain were used to analyze the inter-beat variation, the HRV. The features extraction started with the identification of the R peaks of each signal and the index of the peaks were collected. Then, the normal-to-normal interval (NNI) of the R peaks was prepared to be fed into the algorithm of time domain and frequency domain.

Naturally, all normal ECG signal will have a common amplitude and time interval for the R-peak occurrences along the signal as shown in Figure 6. However, the high-frequency components become weaken and the QRS complex becomes wider when the activation pulse is unable to pass through the pulse threshold which is the normal conduction path track. Thus, extracting statistical features at any point of the time interval based on time domain will be useful for the ECG signal analysis such as maximum, minimum, mean, and the NNI difference of the heart rate. Other than that, it would be the standard deviation of the heart rate series and NNI, the root mean of squared NNI difference, and the ratio between number of NNI difference greater than 50ms and total number of NNI. "The frequency domain analysis uses high-frequency and low-frequency ranges to differentiate ventricular rhythm, atrial rhythm, parasympathetic and sympathetic activity signals" [18]. Therefore, in this research, the frequency bands will be from the range of VLF in 0.00-0.04 Hz, LF in 0.04-0.15 Hz and HF in 0.15-0.40 Hz [19]. The power spectral density (PSD) estimation was calculated using Welch's method [23] from the NNI series. The PSD was used to compute all frequency domain parameters based on the frequency bands given such as the peak frequencies of all frequency bands, absolute powers of all frequency bands, the relative powers of all frequency bands, logarithmic powers of all frequency bands, normalized powers of the LF and HF frequency bands, LF/HF ratio, and total power over all frequency bands.
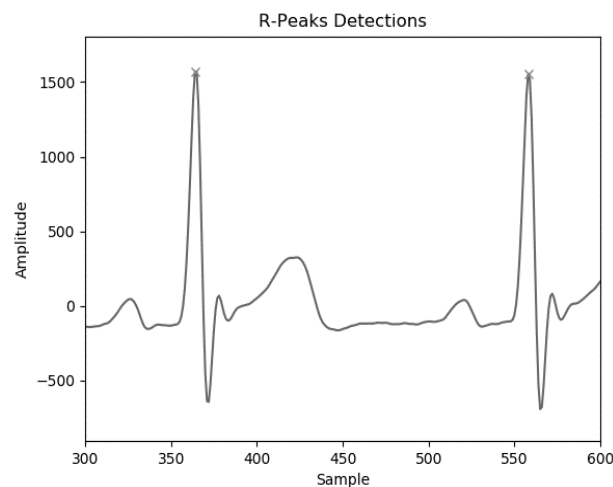

Figure 6. Illustration of RR-interval

2.5. Features selection

Features selection is important to select the most relevant features as many irrelevant features might increase the computational time and cause overfitting of the model. Thus, in this research, there were two methods for features selection which is based on univariate statistical tests, the chi-squared test and ANOVA. The chi-squared test calculates a statistic that has a chi-squared distribution. In this paper, it calculates chi-squared between the targets and features individually then select several desired features based on the best chi-squared scores the chi-squared test decides whether the relationship between two downright features of the sample would mirror their actual relationship in the populace based on the best chi-squared scores. Analysis of variance, a statistical technique succeeded by Ronald Fisher in 1918, is the extent from the t-test and the z-test which were believed to have issues with only permitting the nominal-level features to have two classes, while the ANOVA significantly resolves whether a feature could manifest the potential contrast between two or more classes. Overall, there were 8 clusters starting from three selected features until 10 selected features. Both features selection techniques were tested on all clusters to find which cluster has the highest classification accuracy.

2.6. Classification algorithms

Although Wagner *et al.* [12] concluded the combination of LDF and SFS gives the best results for the recognition of those four emotions, there is a lot of overlapping characteristics in the emotional features that are not able to be differentiated by using linear boundary in the LDA classifier [17]. Therefore, only these three different classifiers were used to compare the performance which are the KNN, MLP and SVM as there is no conclusive claim that justifies the best algorithm so far due to each algorithm advantages and disadvantages as shown in Table 3. KNN is one of the simplest non-parametric classifiers that is often called as the lazy learning algorithm. Non-parametric means that the model is an instance-based learning classifier which its structure is determined from the dataset instead of creating rules compared to other classification methods. This classifier works based on the feature's similarity determined by the distance between the object and the neighbors. The classification of an object is justified by the majority vote from its k nearest neighbors which obtained the shortest distance between the objects. Thus, the new object will be appointed to the most common class among the shortest distance neighbors. SVM is a classifier which is one of the supervised learning that able to work with limited and high dimensional training data. It aims to search for the best hyperplane to separate two different classes of data and maximize the margin between a hyperplane. These two goals are achieved based on statistical learning theory, which contains neural networks, polynomial classifiers, and radial basis functions (RBF). MLP is a neural network that simulates the human brain neural network learning processes. MLP exists in a composition of an input layer with single or multiple input channels based on the number of features input, one or more hidden layers, and an output layer that give the classification results based on the labels or classes. Each neuron is interconnecting to each other and the link between the neurons is identified as the weight value that represents the features. An activation function such as the rectified linear unit (ReLU) function will compute the weight of each neuron. Generally, the backpropagation method will be used to optimize the error in an artificial neural network (ANN) by passing the weight as the input from the next layer to the previous layer to update the previous weight values. The final weights are then used by the output layer for classification. The classifier variants of each classifier are shown in Table 3.

Table 3. Classifier variants

Classification Algorithms	Classifier Variants	Number of Features Selected
KNN	3 Nearest Neighbors (3NN)	
	5 Nearest Neighbor (5NN)	
	10 Nearest Neighbor (10NN)	
	4 Hidden Layers (MLP4)	Between 3-10
MLP	6 Hidden Layers (MLP6)	
	8 Hidden Layers (MLP8)	
SVM	Kernel=rbf	

3.   RESULTS AND DISCUSSION

3.1. Features selection

The proposed method for features extraction and features selection for stress detection was applied on both AUBT datasets and video interview datasets. Based on the AUBT datasets as the benchmark result, chi-squared able to select the significant features that result in higher accuracy better than ANOVA, as shown in Table 4. Out of the 8 clusters, only one cluster with 5 significant features was selected by both techniques

for the stress detection. The classifiers performance has also outrun the performance stated by [12] using the same dataset as shown in Tables 4 and 5.

Table 4. Highest classification accuracy from each classifier

Classifier	No. of Features	Chi^2	ANOVA
5NN	5	90.0%	70.0%
MLP6	5	90.0%	70.0%
SVM	5	100.0%	60.0%

Table 5. Classification accuracy comparison of benchmark dataset

Results		Benchmark	
Classifier	Accuracy	Classifier	Accuracy
5NN (Chi^2)	90.0%	5NN (SFS) [2]	86.36%
MLP6 (Chi^2)	90.0%	MLP6 (SFS) [5]	87.50%
SVM (Chi^2)	100.0%	SVM [7]	92.72%

The 5 significant features selected by chi-squared, and ANOVA are shown in Tables 6 and 7. It shows that the combination of time and frequency domain features selected by chi-squared has higher classification accuracy rather than time domain features alone selected by ANOVA. It also indicates that the domination of frequency domain features in the cluster is able to discriminate the different level of stress better. Therefore, *tinn_n, fft_abs_vlf, fft_abs_lf, fft_abs_hf, and fft_total* are the main features to classify the level of stress in the video interview dataset. Correlation analysis is also performed to understand the relationship between the five selected ECG features. Based on Table 8, the correlation coefficient value of *fft_total* and *fft_abs_vlf* has recorded the largest positive value of nearly 1.0 which is 0.9177. This value indicates a significant positive correlation between both features, while the correlation between *tinn_n* and *fft_abs_vlf* was found as the lowest negative coefficient value compares to the other combination which is -0.3283. This means those two features are negatively correlated to each other. Two scatter plots were produced to visualize the correlation of those features as shown in Figures 7 and 8.

Table 6. Feature selection by chi-squared

Features Name	Domain	Description
tinn_n	TD	N value of the TINN computation at (N, 0)
fft_abs_vlf	FD	Absolute powers of VLF bands [ms^2]
fft_abs_lf	FD	Absolute powers of LF bands [ms^2]
fft_abs_hf	FD	Absolute powers of HF bands [ms^2]
fft_total	FD	Total power over all frequency bands [ms^2]

Table 7. Feature selection by ANOVA

Features Name	Domain	Description
nni_mean	TD	Mean NNI [*ms*]
nni_min	TD	Minimum NNI [*ms*]
hr_mean	TD	Mean heart rate [*bpm*]
hr_max	TD	Maximum heart rate [*bpm*]
tinn_n	TD	N value of the TINN computation at (N, 0)

Table 8. Correlation between Selected Features by Chi-squared

Correlation	*tinn_n*	*fft_abs_vlf*	*fft_abs_lf*	*fft_abs_hf*	*fft_total*
tinn_n	1.0				
fft_abs_vlf	**-0.3283**	1.0			
fft_abs_lf	-0.2275	0.8090	1.0		
fft_abs_hf	-0.2441	0.5730	0.6308	1.0	
fft_total	-0.2992	0.8863	**0.9177**	0.8455	1.0

3.2. Classification result on video interview dataset

Based on Table 9, MLP6 shows the highest classification accuracy at 68.18%, followed by 5NN and SVM at 67.27% and 61.82% respectively. It has a better classification performance on the new dataset then SVM that was initially recorded the highest accuracy during the training phase using AUBT dataset in Table 5. This is thought to be caused by the overfitting that occurred in the SVM model training where it fits the training dataset extremely well including the outliers. Overfitting also contributes to a degradation of the model's generalization properties and results in its unreliable result on new dataset [24]. It is for this reason that the MLP6 was chosen as the main classifier model to be used rather than 5NN and SVM to study the Video Interview dataset in deep. To optimize this result, a data optimization process was done by re-clustering the video interview dataset using k-means clustering algorithm into three different classes which

are high stress, low stress, and relax instead of just having high stress and relax as proposed in the initial labels. Not only that, in [25] also proposed to have three classes of stress (low, moderate, high) to get a better data representation. It is believed that the changing between the high stress to relax is intercalated with some other indistinct stress levels such as low stress [8]. The re-clustering result was inspected in order to look into each subject condition in each session.

Surprisingly, a subject was suspected as an outlier to the group as it has recorded a consistent high stress condition in all sessions. This might be caused by various technical reasons such as improper placement of the patches on the subject's body or introduction of medication and caffeine-contained drink prior to the data collection sessions. Further observations were done on the subject stress level by excluding the outlier subject from the dataset. Classification accuracy was again calculated using MLP6 and surprisingly the classification accuracy hit 92.93%. It can be said that that subject was an outlier to the group which leads to a low accuracy result as mentioned Table 10 with 88.18% accuracy. Another classification result was generated by the same classification model using MLP6 and the accuracy hiked up from 88.18% to 92.93%. Table 10 shows a clearer overview of the changes in the number of subjects in each label with the classification accuracy.
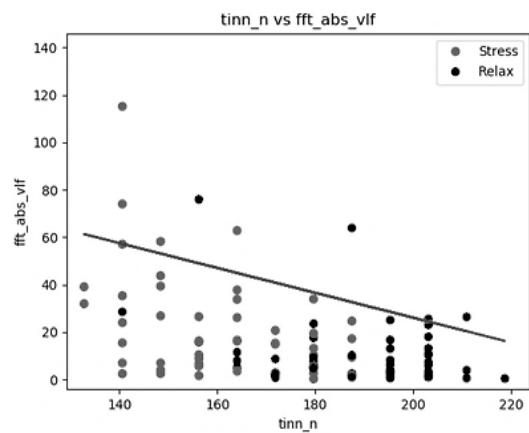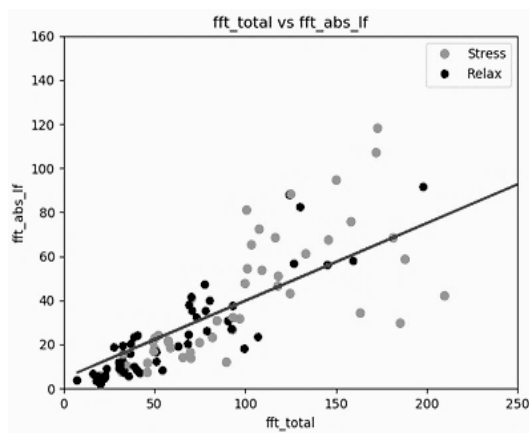

Figure 7. Correlation between *fft_total* and *fft_abs_lf* on benchmark dataset

Figure 8. Correlation between *fft_total* and *fft_abs_lf* on benchmark dataset

Table 9. Classifier performance comparison based on classification accuracy

Classifier	Video Interview Dataset	AUBT Dataset
5NN	67.27%	90.00%
MLP6	68.18%	90.00%
SVM	61.82%	100.00%

Table 10. Subjects distribution among different set of labels

Level of Stress	Initial Label	Re-Clustering	Outlier Withdrawal
High Stress	60	18	7
Low Stress	0	59	59
Relax	50	33	33
Accuracy	68.18%	88.18%	92.93%

4. CONCLUSION

A thorough discussion on different classifiers by using distinguish set of selected features based on the classification accuracy has been provided. We proposed MLP6 to be the best classifier compared to 5NN and SVM. Five features were selected as the main features for stress detection in ECG signal which are *tinn_n, fft_abs_vlf, fft_abs_lf, fft_abs_hf,* and *fft_total*. These five features are from both time and frequency domain instead of only focusing on only one domain. Further observation shows that the initial high stress and relax label was not adequate to describe the whole subject condition throughout the sessions, indeed low stress level exists in between those two. It is believed that the changing between the high stress to relax is intercalated with some other indistinct stress levels such as low stress. By applying all these findings, the classification result could be improved up to 92.39% using MLP6. This paper concludes that the stress level

is a subjective measure where it can be varying between person, but a person is able to adapt to the same particular stress that could help them to lower their stress level and perform a better action. Plus, the changing between high stress and relax is intercalated with some other indistinct stress levels such as low stress. In this case, the subjects were recording the job video interview for a few times before the actual video, thus allowing them to learn their mistakes, adapt to the situation and surrounding, change their approaches and improve their performance in a low stress condition throughout the video recording sessions.

This research has been mainly focused on the use of neural network and other supervised learning algorithm for the stress level classification which is mainly based on the given targets or labels. However, there are many different adaptations, tests, and experiments that can be done in the future. Implementation of fuzzy logic in the classification process could be very interesting to understand better on the stress level in human. Fuzzy rules and membership functions are subjective same as the level of stress that has the overlapping characteristics that could not be differentiated using linear boundary. Thus, more different stress levels could be recognized during the video recording session rather than just high stress, low stress and relax. This could bring more realistic results to be discussed. Other than that, it is also suggested that the study on the combination of time and frequency domain features should be given extra attention. Another approach that could be done is by mix-matching each extracted feature rather than only relying on the automatic statistical methods to the find significant features for specific level of stress. This can help to understand more on the contribution of each feature towards the stress detection process in ECG signal.

REFERENCES

[1]   S. Goel, G. Kau, and P. Toma, "A novel technique for stress recognition using ECG signal pattern," *Current Pediatric Research*, vol. 21, no. 4, pp. 674–679, 2017.
[2]   A. Kaul, A. S. Arora, and S. Chauhan, "ECG based human authentication using synthetic ECG template," in *2012 IEEE International Conference on Signal Processing, Computing and Control*, Mar. 2012, pp. 1–4, doi: 10.1109/ISPCC.2012.6224341.
[3]   H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: a meta-analysis and review of the literature," *Psychiatry Investigation*, vol. 15, no. 3, pp. 235–245, Mar. 2018, doi: 10.30773/pi.2017.08.17.
[4]   R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, Dec. 2019, doi: 10.1186/s12911-019-0742-y.
[5]   P. Trimmer, E. Paul, M. Mendl, J. McNamara, and A. Houston, "On the evolution and optimality of mood states," *Behavioral Sciences*, vol. 3, no. 3, pp. 501–521, Aug. 2013, doi: 10.3390/bs3030501.
[6]   A. Pise, H. Vadapalli, and I. Sanders, "Facial emotion recognition using temporal relational network: an application to E-learning," *Multimedia Tools and Applications*, Nov. 2020, doi: 10.1007/s11042-020-10133-y.
[7]   L. Rapolienė, A. Razbadauskas, and A. Jurgelėnas, "The reduction of distress using therapeutic geothermal water procedures in a randomized controlled clinical trial," *Advances in Preventive Medicine*, vol. 2015, pp. 1–10, 2015, doi: 10.1155/2015/749417.
[8]   R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, no. 5, pp. 459–482, Nov. 1908, doi: 10.1002/cne.920180503.
[9]   B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, and B.-T. Zhang, "Deep ECGNet: an optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals," *Telemedicine and e-Health*, vol. 24, no. 10, pp. 753–772, Oct. 2018, doi: 10.1089/tmj.2017.0250.
[10]   A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: first steps towards an automatic system," in *Lecture Notes in Computer Science*, 2004, pp. 36–48, doi: 10.1007/978-3-540-24842-2_4.
[11]   S. Pourmohammadi and A. Maleki, "Stress detection using ECG and EMG signals: a comprehensive study," *Computer Methods and Programs in Biomedicine*, vol. 193, Sep. 2020, doi: 10.1016/j.cmpb.2020.105482.
[12]   J. Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification," in *2005 IEEE International Conference on Multimedia and Expo*, pp. 940–943, doi: 10.1109/ICME.2005.1521579.
[13]   F. K. Gomis, M. S. Camara, and I. Diop, "Feature reduction algorithm for universal steganalysis," in *Learning and Analytics in Intelligent Systems*, Springer International Publishing, 2020, pp. 451–457, doi: 10.1007/978-3-030-36778-7_50.
[14]   S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, "Stress detection in working people," *Procedia Computer Science*, vol. 115, pp. 359–366, 2017, doi: 10.1016/j.procs.2017.09.090.
[15]   M. S. Athira, N. Sindhu, and S. D. Jerritta, "Human stress detection using ECG and HRV signals," *International Journal of Engineering Technology, Management and Applied Sciences*, vol. 5, no. 5, pp. 470–475, 2017.
[16]   J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.
[17]   M. Murugappan, S. Murugappan, and B. S. Zheng, "Frequency band analysis of electrocardiogram (ECG) signals for human emotional state classification using discrete wavelet transform (DWT)," *Journal of Physical Therapy Science*, vol. 25, no. 7, pp. 753–759, 2013, doi: 10.1589/jpts.25.753.
[18]   C.-H. Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier," *Computers and Mathematics with Applications*, vol. 55, no. 4, pp. 680–690, Feb. 2008, doi: 10.1016/j.camwa.2007.04.035.
[19]   O. Kwon *et al.*, "Electrocardiogram sampling frequency range acceptable for heart rate variability analysis," *Healthcare Informatics Research*, vol. 24, no. 3, 2018, doi: 10.4258/hir.2018.24.3.198.
[20]   J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," in *IFMBE Proceedings*, Springer Berlin Heidelberg, 2009, pp. 1366–1369, doi: 10.1007/978-3-540-89208-3_324.
[21]   J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, Jun. 2005, doi: 10.1109/TITS.2005.848368.
[22]   D. Priya and R. Sivaraj, "A review of missing data handling methods," *International Journal On Engineering Technology and Sciences-IJETS*, vol. 2, no. 2, pp. 58–68, 2015.

[23] R. Bartels and T. Peçanha, "HRV: a pythonic package for heart rate variability analysis," *Journal of Open Source Software*, vol. 5, no. 51, Jul. 2020, doi: 10.21105/joss.01867.

[24] H. K. Jabbar and R. Z. Khan, "Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study)," in *Computer Science, Communication and Instrumentation Devices*, 2014, pp. 163–172, doi: 10.3850/978-981-09-5247-1_017.

[25] N. Keshan, P. V Parimi, and I. Bichindaritz, "Machine learning for stress detection from ECG signals in automobile drivers," in *2015 IEEE International Conference on Big Data (Big Data)*, Oct. 2015, pp. 2661–2669, doi: 10.1109/BigData.2015.7364066.

BIOGRAPHIES OF AUTHORS

Ali Afzalian Mand ⓘ �Ⓢ🄿 served as a Lecturer at the Faculty of Information Science and Technology, Multimedia University, Malaysia. Graduated from Multimedia University in MSc IT. Research field in information technology, health informatics and machine learning. He can be contacted at email: ali.mand@qnet.net.

Md Shohel Sayeed ⓘ �contactⓈ🄿 has been a member of Multimedia University since 2001 and now he serves as an Associate Professor and the Chairperson of Centre for Intelligent Cloud Computing (CICC) of the Faculty of Information Science and Technology. He received his Doctor of Philosophy from Multimedia University in Engineering, specializing in hand signature verification, and holds Master of Information Technology degree from the University Kebangsaan Malaysia, specializing in industrial computing. He received his Bachelor of Science degree in Agricultural science from Bangladesh Agricultural University. Dr. Shohel has more than 24 years meritorious working experience and he holds a challenging career which combines research, versatile administration, and excellent teaching. His core research interest is in the area of Biometrics, big data, cloud computing, artificial intelligence, information security, image and signal processing, pattern recognition and classification. He has published over 80 research papers in international peer-reviewed journals and international conference proceedings as a result of his research work. His research works have been published by high ranked peer-reviewed journals such as IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), International Journal of Pattern Recognition and Artificial Intelligence (IJPRI), Expert Systems with Applications, Discrete Dynamics in Nature and Society (DDNS) as well as several peer-reviewed international journals. He can be contacted at email: shohel.sayeed@mmu.edu.my.

Md. Jakir Hossen ⓘ �Ⓢ🄿 is graduated in Mechanical Engineering from the Dhaka University of Engineering and Technology (1997), Master's in Communication and Network Engineering from Universiti Putra Malaysia (2003) and Ph.D. in Smart Technology and Robotic Engineering from Universiti Putra Malaysia (2012). He is currently a Senior Lecturer at the Faculty of Engineering and Technology, Multimedia University, Malaysia. His research interests are in the area of artificial intelligence (fuzzy logic, neural network), inference systems, pattern classification, mobile robot navigation, and intelligent control. He can be contacted at email: jakir.hossen@mmu.edu.my.

Muhammad Amer Ridzuan bin Zuber ⓘ �Ⓢ🄿 has received his Bachelor of Science (Bioinformatics) from Multimedia University (MMU 2019). He is currently pursuing his Master of Engineering (Computer and Microelectronic Systems) in Universiti Teknologi Malaysia (UTM). His research interest is in field of artificial intelligence, data science, and software and hardware validation system. He can be contacted at email: ridzuber@gmail.com.