Jestr

Journal of Engineering Science and Technology Review 15 (5) (2022) 70 - 76

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Stress Detection using Adaptive Neuro Fuzzy Inference System

Shohel Sayeed[1,*], Ali Afzalian Mand[1], Jakir Hossen[2], Amelia Binti Kassim[1] and Pa Pa Min[1]

[1]*Faculty of Information Science and Technology, Multimedia University, Malaysia*
[2]*Faculty of Engineering and Technology, Multimedia University, Malaysia*

Received 3 March 2022; Accepted 2 October 2022

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). The main purpose of this paper is to apply the adaptive neuro fuzzy inference system (ANFIS) on the stress-level detection from Video Interview dataset of 10 male subjects who were recording the video resume for the analysis purposes. The proposed method able to achieve an accuracy of 100 % with the re-clustering and ANFIS framework.

Keywords: ECG, ANFIS, Stress Detection, Re-clustering

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. 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 [1], 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. 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. To indicate the stress, it can be physiological or behavioral. [2] The physiological stress or calmness can be dectect by heart rate (HR), electro-dermal activity (EDA) and heart rate variability (HRV) and for the behavioral, it can be detected through smartphone activity statistics. [3] 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.

Despite stress has always been defined as a negative situation, this has overshadowed the bright side of the stress that can actually 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 [4][5]. This has also been studied earlier by [6] 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. Not only that, the journey to obtain the optimal arousal level for an optimum performance needs a series of particular 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 [7]. In this study, stress detection is done by analyzing the heart electrical signal as one of the biosensors known as the electrocardiogram signal (ECG). An ECG signal 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. 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. Nowadays, there are advanced progress in applying computing technologies [1-14] that have significant progress in artificial intelligence.

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

*E-mail address:shohel.sayeed@mmu.edu.my
ISSN: 1791-2377
doi:10.25103/jestr.155.10

the physiological signal through suitable biosensors such as Electrocardiogram (ECG) signal to understand the situation mentioned above. In summary, the main contribution of this paper can be concluded as follows:

- We collected Electrocardiogram (ECG) signals from 10 subjects who are recording the video resume and interview
- The Electrocardiogram (ECG) signals are preprocessed and selected the important features significant to the stress level
- The Adoptive Neuro Fuzzy System is used to design the framework of stress detection system
- Re-clustering approach is used for the misclassified stress level
- The proposed framework is analysed and evaluated

The rest of the paper is organized as follows. In Section 2, we discussed the related works for emotion recognition and stress detection using Electrocardiogram (ECG) signals. The proposed research method is explained in Section 3 with signals pre-processing and classification for stress detection. The analysis and evaluation results are discussed in Section 4. Finally, the paper is concluded with the final findings and futures work in Section 5.

2. Related Works

Haggs et.al [1] describes that "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". They described the process and methods for the automatic emotion detection for different types of bio-sensors. In this paper, the dataset was the three positive and negative states under variable arousal level using 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 90Hz, high-pass filtering at 0.5Hz and notch filtering at 50Hz. 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. This paper proposed neural network classifier as the classification method and 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.

In the early research of 2005, Wagner et.al [8] also dedicated to explore the physiological signal-based emotion recognition which has less paid attention compared to audio-visual emotion channels. The authors aimed 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. They 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, 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. This paper further discussed the three different classification methods which are the k-nearest neighbour (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%. Then, they claimed that it is much easier to detect emotion along the arousal axes rather than on the valence axes.

As one of the emotion recognitions, stress detection gained the attention from the researchers especially using the bio-sensors. 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 [5]. The 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. The authors used the same method as [8] to get the R peaks intervals, while the Welch algorithm was used to extract the power spectral density of HRV features. For classification, [9] used two types of features analysis methods which were the individual features analysis and the feature combination analysis. The first method is to analyse each of the features that has the best classification precision and accuracy, then to generate five clusters of features to be tested [5]. 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, RMSSD, AVNN, SDANN, LF, HF, NN50, pNN50, Mean GSR, Med GSR, STD GSR. 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. The authors concluded the GSR, heart rate, and HRV features contribute the most to the stress recognition.

Another notable paper for stress detection is presented by [9]. They used the Electrocardiography known as ECG which could be defined as the electrical activity of the heart recording process over time using electrodes placed on the human skin [9]. 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 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 [9]. 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 [10]. 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 0.32s respectively for the missed beat to detect the R peaks and exclude the noisy peaks. The pre-processed signals were used for the feature's extraction under time domain (TD) and frequency domain (FD). [9] extracted and selected time domain features skewness from the HRV signal and standard deviation for the ECG signal using the Pan-Tompkins algorithm [10]. 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" [9].

Shirvan et.al [11] presented the classification of stress level by analazying the functional near infrared spectroscopy (fNIRS) signals. They collected the fNIRS signal while the subject is taking the arithmetic task in a limited time. They extracted the non-linear and linear features from the signals and classified the level of stress. The authors [11] able to achieve the high accuracy for classification of high and low level of stress with 88.72% and even 96.92 % for the rest states. They argued that their proposed method able to classify the early state of stress in the real world. To present the novel method for stress detection, Goel et.al [12] introduced the stress detection of automobile drivers using the ECG signals. The signals filtering and denoising is done as the pre-processing for the MIT-BIH PhysioNet Multi-parameter Database. They extracted four different types of features from the ECG signals that are Isoelectric level, QRS complex, P wave and ST waves. They calculated the means, variance and standard deviation of each features and analysed with their baseline values. They were able to achieved 87% on the stress detection with their proposed method. Bichindaritz et.al [13] also used the same dataset as [12] to present the multi-level stress detection of the drivers. They extracted the 14 different types of features from ECGs called fiducial points (P, Q, R,S). They even added the variance of the interval features however, it didn't improve the results. So, the entropy features are added using the multiscale entropy analysis and proceed to the stress levels classification. The multi-level classification is done with the Weka (Waikato Environment for Knowledge Analysis). According to the analysis, they were able to achieve the highest accuracy on Multilayer Perceptron with 10 selected features. Behinaein et.al [14] introduced transformer architecture for stress detection using the Electrocardiogram (ECG) signals. They used the two public datasets called WESAD and SWELL-KW. The raw ECG signals are pass through the end-to-end deep network including the convolution layers, subnetworks, transformer encoder and fully connected layers. The spatiotemporal features that extracted from the convolutional layer was fed into the encoder. The fully connected layer takes charges the classification of the stress level. According to analysis, the highest accuracy for WESAD dataset is 91.1% after fine-tuning the 10% of test data. Similar to that, SWELL-KW also was able to achieve the 71.6 % after fine-tuning.

In current research works on stress detection, the authors used different kind of stress detection data such as ECG, Galvanic Skin Response (GSR) , (fNIRS) with different analysis and machine learning process. However, the researchers still fail to indicated the level of stress in different situation and condition. So, 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 analyse 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.

3. Research 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. The proposed stress detection framework included data collection, signal pre-processing, feature extraction, feature selection, re-clustering and finally the classification with Multi-Layer Perceptron classifier.

3.1. 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.

Since the emotion and stress indications are subjective and relevant to many personal factors such as gender, age, nutrition, and stereotypes, we have selected 10 subjects all male, all Malaysian and within a fixed age range to reduce other factors affecting the subjects' affective state. For example, referring to our fifth reference, Kim et *al*. [8] only used 4 males (all German ages 18-25) for 25 sessions of data collection each. We used 10 users for each 11 sessions of data collection. In other words, the dataset contains 110 samples of ECG signal that was collected from 10 subjects under 11 sessions. In order to get the proper ECG data, the students are required to obey the instructions such as: (i) No heavy meal for at least two hours before experiment, (ii) No coffee for at least two hours before experiment (iii) No medication at least one day before. The details list of the subjects is shown in Table 1.

Table 1. Subject Requirements

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

The experimental data are collected under 11 sessions with the video resume recording and listening the music one after another.

Firstly, Subject is explained on the requirements of video shooting for resume and allowed to rehearse using a mirror provided only once. The subjects will not be informed of the number of attempts they can have. They can re-attempts and improve the video quality after each attempt. After each attempt, the students are allowed to listen their favorite music. The process is repeated until the subject reached 3rd music section after the video interview attempt. And then, the subjects are interrupted by informing him that he has only 2 more attempts for video shooting. The subject is asked to complete a set of questionnaire and/or interview after the fifth attempt. In the last session, subject is asked to listen to his music for complete relaxation for about 3 minutes. The session and predicted labels are shown in Table 2.

Table 2. Data collection Session and expected labels

Attempt 1 (A1)	Music 1 (M1)	Attemt 2 (A2)	Music 2 (M2)	Attempt 3 (A3)	Music 3 (M3)	Interception	Attempt 4 (A4)	Music 4 (M4)	Attempt 5 (A5)	Last Session
High Stress	Relax	High Stress	Relax	High Stress	Relax	High Stress	High Stress	High Stress	High Stress	Relax

3.2. Signal pre-processing

The raw signal data normally included the noise or null values. Holding to the popular analytic concept in the computer science world, "Garbage in, garbage out" (GIGO) which dictates the importance of data preprocessing before any experimental process to be carried out, reason being the output can only be as accurate as of the data input. Thus, the raw ECG signals undergone pre-processing to improve the overall performance of the research. So, we performed the signal pre-processing in this framework with missing data handling and bandpass signal filtering. All the Not a Number (NaN) and empty column exist in the raw signal datasets were replaced by 0. This problem may cause the analysis to end with wrong inferences about the experimental data.

Real-worlds worlds ECG signal is believed to be distorted with a lot of background noise. In order to extract the R-peaks precisely and get the clean ECG singal, we performed the bandpass filtering. Here, we used Butterworth bandpass filtering which 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 0.5Hz to remove the frequency sways caused by the movement of the artefact. While, the high cut-off frequency was at 90.0Hz to remove sharp peaks in the signal that are considered as noise [1].

3.3. 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. Pyhrv is the open source python toolbox for Heart Rate Variability which allowed to perform R-peak detection and feature extraction in different domains including time, frequency and nonlinear modules.

3.3.1 R-Peak Detection

The R-peak in the QRS interval is the most vital feature for understanding the ECG data. This is based to the fact that the "R-peak detection in ECG is one such method that is widely used not only to diagnose heart diseases such as heart rhythm irregularities but also able to estimate heart-rate variability (HRV) which is closely related to stress detection"[15]. Therefore, more meaningful features could be extracted that can greatly manifest the ECG signal by validating the R-R intervals of the signals.

During the R-peaks detection using SciPy functionality[16], the distance for both datasets was set to a different value which is 50 for the MMU dataset.

This is because the distance indicates the minimum horizontal range on the data point that is needed between neighbouring peaks.Since the MMU dataset has R-peaks appearance more frequently in an imperfect manner, the distance is needed to be set shorter to enable the program to detect more accurate R-peaks. Thus, all the small unrelated peaks that could be noise peaks in the signal were discarded until the remaining peaks meet the desired requirement.

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 , frequency and nonlinear domains.

3.3.2. Time Domain Features

Naturally, all normal ECG signal will have a common amplitude and time interval for the R-peak occurrences along the signal. 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.

3.3.3. Frequency Domain Features.

"The frequency domain analysis uses high-frequency and low-frequency ranges to differentiate ventricular rhythm, atrial rhythm, parasympathetic and sympathetic activity signals" [17]. Therefore, in this paper, the frequency bands will be from the range of VLF in 0.00 – 0.04Hz, LF in 0.04 – 0.15Hz and HF in 0.15 – 0.40Hz [18]. The Power Spectral Density (PSD) estimation was calculated using Welch's Method [19] 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.

3.3.4. Non-linear Domain Features

Non-linear parameters are intended to improve the non-linear properties and unpredictability of the NNI series caused by various complex physiological dynamics of the human body leading to HRV (such as SNS and PNS). In recent years, HRV research has used a variety of mathematical techniques for nonlinear analysis. However, many of them have proven unsuitable for HRVA due to the lack of common methods in HRV research, complexity, or lack of proper scientific support [20]. The common methods are Poincare Plot, Sample Entropy and Detrended Fluctuation Analysis (DFA). In this paper, we used 4 features from Poincare Plot (SD1, SD2, ellipse_area, sd_ratio). Thus, there are 16 frequency domain features, 21-time domain feature and 4 nonlinear domain Features.

3.4. Features selection

Features selection is important to select the most relevant features as a large number of irrelevant features might increase the computational time and cause overfitting of the model. We performed this step to eliminate the unnecessary and irrelevance features from the dataset which will allows us to improve the accuracy and computation time. In this paper, 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.

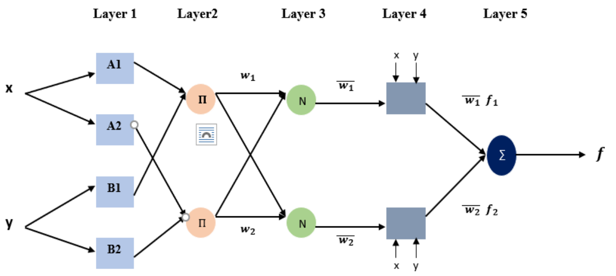

Fig. 1. The Adaptive Neuro Fuzzy Inference System

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 (ANOVA), 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. So, we divided the 8 clusters starting from three selected features until 10 features. Both selection techniques with different features clusters are tested for classification and we will only document the highest accuracy clusters in this paper.

3.5 Classification Methods

Adaptive Neuro Fuzzy Inference System

The adaptive neuro fuzzy inference system (ANFIS) is firstly introduced by J.S. Roger Jang in 1993. The main idea of this system is to combine the learning ability of the neural network and knowledge representation of the fuzzy logic ability. The ANFIS is the Type (3) fuzzy system is also known as Takagi-Sugeno Fuzzy System or universal estimator. As we assumed 2 inputs of (x, y), the IF-THEN rules of the Takagi-Sugeno Fuzzy System can be expressed as below:

(i) Rule 1: If x is A_1 and y is B_1, then $F_1 = p_1 x + q_1 y + r_1$,
(ii) Rule 2: If x is $A_2$and y is B_2, then $F_1 = p_2 x + q_2 y + r_2$,

The x and y are the inputs in the crisp set and the A_i and B_i are Fuzzy set, the output F_i is the fuzzy membership function and p_i , q_i and r_i are the parameters that will be determined during the training. The normal ANFIS system include 5 layers and is demonstrated in Figure 1.

Layer 1 is also known as fuzzification layers which determine the membership function belongs to the input values. The input values (for example x and y) pass through the fuzzification layer and the fuzzy sets (A_i and B_i)also known as linguistic labels that characterized by proper membership function to produce the fuzzy value. The output of this layer can be denoted as O_i^j for the ith node and jth layers and μ is the member
ship function. There are different kinds of membership function such as Gaussian, Trapezoidal, Triangular. Here, we used the bell-shaped Gaussian function for this paper. So, the output can be expressed as below:

$$O_i^1 \begin{cases} \mu_{A_i}(x), & i = 1,2 \\ \mu_{B_{i-2}}(x), & i = 3,4 \end{cases} \quad (1)$$

The second layer is known as rule layer which responsible for the firing the strength of fuzzy rules. In this layer, the IF-THEN rules of the Takagi-Sugeno Fuzzy System is used as mentioned above. The output node of this layer is also weight of the node and multiplication of the all coming inputs. The output node for the layer 2 is as shown as belows.

$$O_i^2 = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y) \quad , \quad i = 1,2 \quad (2)$$

The third layer (layer 3) is called the normalizing layer to evalued the computed firing strengths from the previous layer. The neurons are normalized and fixed by computing each rule divided by sum of all firing strength of total rule. The output of the third layer is formulated as

$$O_i^3 = \overline{w_i} = \frac{w_i}{w_1 + w_2} \quad , i = 1,2 \quad (3)$$

The fourth layer known as defuzzation layers takes the normalized values (w_i)and the consequence parameter set $\{ p_i , q_i , r_i \}$ as inputs. The output neuron can be calculated as followed:

$$O_i^4 = \overline{w_i} \times f_i = \overline{w_i}\,(p_i \times x + q_i \times x + r_i) , i = 1,2 \quad (4)$$

The last layer (Layer 5) is the output layer where a single neron are presented as the output. The output can be achieved by calculating the weighted average of the outputs from each rule.

$$O_i^5 = f(x,y) = \sum_i \overline{w_i}\, f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad , i = 1,2 \quad (5)$$

The ANFIS apply both forward pass and backward pass for the training. During the forward pass, only the consequence parameters are updated while the premise parameters are fixed. The final output is calculated after the consequence parameter are updated in layer 4. The backward pass, also called back propagation where the error propagated back to the layer 1 as the final output is calculated. The only premise parameters are updated during the back pass.

4. Results and Discussions

4.1 Classification Accuracy on MMU video interview dataset before re-clustering

The three different MLP architectures are evaluated for the stress level classification with different selected features. As we mentioned in 3.4, we will only present the accury with the highest cluster for both Chi squared and ANOVA features selection. The details of selected features for time (TD), frequency (FD) and Non-Linear Domain (NLD) are documented in Table 3 and Table 4**.**

Table 3. Feature Selection by Chi-squared

Features Name	Domain	Description
nni_max	TD	Maximum NNI [ms]
nni_diff_max	TD	Maximum NNI difference [ms]
nn20	TD	Number of NN interval differences greater 20 milliseconds
pnn20	TD	Ratio between NN50 and total number of NN intervals
fft_abs_vlf	FD	Absolute powers of very low frequency bands [$ms2$]
fft_total	FD	Total power over all frequency bands [$ms2$]
sd_ratio	NLD	Ratio between SD1 and SD2 (SD2/SD1)

Table 4. Feature Selection by ANOVA

Features Name	Domain	Description
nni_max	TD	Maximum NNI [ms]
rmssd	TD	Root mean of squared NNI differences [ms]
tinn_n	TD	N value of the TINN computation (left corner of the interpolated triangle at (N, 0))
tinn	TD	N value of the TINN computation (left corner of the interpolated triangle at (N, 0))
fft_peak_vlf	FD	Peak frequencies of very low frequency bands [Hz]
fft_abs_vlf	FD	Absolute powers of very low frequency bands [$ms2$]
sd_ratio	NLD	Ratio between SD1 and SD2 (SD2/SD1)

Based on Table 5, the classification accuracy on training data which is the MMU video interview dataset have shown that Cluster C with 7 number of features selected has the highest accuracy compares to the other 2 clusters. The cluster B and A have recorded 54.5% and 68.18 as the classification accuracy by using Chi 2 statistical calculation meanwhile 45.45 % and 68.18 % accuracy by using ANOVA statistical calculation. Meanwhile, the overall accuracy has recorded the highest accuracy at 72.72% by using Chi 2 and also the poorest accuracy at 45.45% while using ANOVA.

Additionally, Table 3 and 4 show that Cluster B with Chi2 consists of two features from the frequency domain, three features from time domain and one features from non-linear domain features same as the while the Cluster B with ANOVA. Even though the features are from the different domain, the result is not promising. So, we will try to re-cluster the stress levels and the features selection will perform again for better accuracy.

4.2. Clustering Methods

The MMU video interview dataset included two levels of stress which is high stress and Relax (0 or 1). According to the classification results in Table 5, we can see that some of the stress levels are misclassified. So, the result optimization was also done to improve the performance of the result by optimizing the dataset. One of the methods was re-clustering the labels of the MMU dataset by using unsupervised learning to overcome any misclassification on the expected outcome. Here, we used K-Means clustering to compare the performance of the stress detection with the new stress level from clustering approach.

The clustering is one of the unsupervised learning methods which enables to divide the entire dataset into the desired number of clusters based on the pattern of the data. Among the clustering approach, K-Means Clustering is a centroid-based algorithm and also known as distance-based algorithms [21][22]. We choose the number of K as 3 which are High Stress (HS), Low Stress (LS), and Relax (R) instead of only high stress and relax as proposed in the initial framework. Here, the data are divided into 3 clusters to predict the levels of the stress in classification phase.

4.3 Classification result on MMU video interview dataset after re-clustering

Based on the result of Table 5, we can see that same of the stress level are misclassified. So, we performed the clustering with K-Means clustering and Fuzzy C-means clustering. We need to re-select the features for datasets with clustered dataset for accurate evaluation. The new sets of features for both Chi-squared and ANOVA are documented in the following Table 6 and 7.

Table 5. Classification Accuracy for different clusters

	Cluster	No. of Features Selected	Chi^2	ANOVA
	A	5	54.5 %	45.45
	B	6	68.18	68.18
ANFIS	C	7	72.72	68.18

Table 6. Feature Selection by Chi-squared after clustering

Features Name	Domain	Description
hr_min	TD	Minimum heart rate [bpm]
hr_std	TD	Standard deviation of the heart rate series [bpm]
nni_diff_max	TD	Maximum NNI difference [ms]
fft_abs_lf	FD	Absolute powers of very low frequency bands [$ms2$]
fft_abs_hf	FD	Absolute powers of very high frequency bands [$ms2$]
fft-total	FD	Total power over all frequency bands [$ms2$]
sd_ratio	NLD	Ratio between SD1 and SD2 (SD2/SD1)

Table 7. Feature Selection by ANOVA after clustering

Features Name	Domain	Description
hr_std	TD	Standard deviation of the heart rate series [bpm]]
rmssd	TD	Root mean of squared NNI differences [ms]
sdnn	TD	Standard deviation of NN intervals [ms]
sdsd	TD	Standard deviation of NNI differences [ms]Peak
sd1	NLD	Standard deviation (SD1) of the major axis Ratio between SD1
sd_ratio	NLD	and SD2 (SD2/SD1)

Table 8. Classification Accuracy after K-Means Clustering

MLP	Cluster	No. of Features Selected	Chi^2	ANOVA
	A	5	95.95	81.81
ANFIS	B	6	100	90.9
	C	7	100	90.9

We performed the stress classification with the adaptive neuro-fuzzy inference system in 3 different features clusters. As we seen the Table 8, the three different stress level are classified after K-Means clustering approach. We can see the overall accuracy is significantly improved compare to before clustering. The highest accuracy of 100 % is able to achieve with the 6 and 7 selected features for both Chi squared and also ANOVA able to achieved the promising result.

The high difference the accuracy score is due to the less testing data in MMU video interview dataset. However, we can say the overall the accuracy is desirable after the K-Means Clustering approach. The main reason behind is that the ANOVA feature selection has only included the 1 frequency domain features compare to Chi squared. So, the feature selection is important for the stress level and it need to include all three different domains for the accurate stress level detection.

5. Conclusion

This paper has been mainly focused on the use of hybrid collaboration of neural network and fuzzy inference system. A thorough discussion on different clusters by using distinguish set of selected features based on the classification accuracy has been provided. And also, the re-clustering approach for the stress level classification which is mainly based on the given targets or labels. We used the implementation of fuzzy logic in the classification process to understand better on the stress level in human because the 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. Our proposed method is able to achieve the promising result with the 100 % accuracy rate after re-clustering. In the future, we are planning to explore more on different bio signals which can affect the human stress level and the bigger sample size for the testing human stress level detection.

References

1. 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, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 36–48.
2. F. Albertetti, A. Simalastar, and A. Rizzotti-Kaddouri, “Stress detection with deep learning approaches using physiological signals”. In International Conference on IoT Technologies for HealthCare, 2020, Springer, Cham, pp. 95-111.
3. P. C. Trimmer, E. S. Paul, M. T. Mendl, J. M. McNamara, and A. I. Houston, “On the evolution and optimality of mood States,” Behav. Sci. (Basel), vol. 3, no. 3, pp. 501–521, 2013.
4. L. Rapolienė, A. Razbadauskas, and A. Jurgelėnas, “The reduction of distress using therapeutic geothermal water procedures in a randomized controlled clinical trial,” Adv. Prev. Med., vol. 2015, p. 749417, 2015.
5. S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, “Stress detection in working people,” Procedia Comput. Sci., vol. 115, pp. 359–366, 2017.
6. R. M. Yerkes and J. D. Dodson, “The relation of strength of stimulus to rapidity of habit-formation,” J. Comp. Neurol. Psychol., vol. 18, no. 5, pp. 459–482, 1908.
7. K. Wang,Y. L Murphey, Y. Zhou, X. Hu, and X. Zhang . “Detection of driver stress in real-world driving environment using physiological signals,” In 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), vol. 1, pp. 1807-1814. 2019.
8. 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, 2005, pp. 940–943.
9. M. S. Athira, N. Sindhu, and S. Dr Jerritta, “Human Stress Detection using ECG and HRV Signals,” Int. J. Eng. Technol. Manag. Appl. Sci. www.ijetmas.comMay, vol. 5, no. 5, pp. 470–475, 2017.
10. J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” IEEE Trans. Biomed. Eng., vol. 32, no. 3, pp. 230–236, 1985.
11. Shirvan, R. A., Setaredan, S. K., & Nasrabadi, A. M. (2018). Classification of mental stress levels by analyzing fNIRS signal using linear and non-linear features. International Clinical Neuroscience Journal, 5(2), 55.
12. Goel, S., Kau, G., & Toma, P. (2017). A novel technique for stress recognition using ECG signal pattern. Current Pediatric Research, 21(4).
13. Bichindaritz, I., Breen, C., Cole, E., Keshan, N., & Parimi, P. (2017, June). Feature selection and machine learning based multilevel stress detection from ECG Signals. In International Conference on Innovation in Medicine and Healthcare (pp. 202-213). Springer, Cham.
14. Behinaein, B., Bhatti, A., Rodenburg, D., Hungler, P., & Etemad, A. (2021, September). A Transformer Architecture for Stress Detection from ECG. In 2021 International Symposium on Wearable Computers (pp. 132-134).
15. Park, J. S., Lee, S. W., & Park, U. (2017). R Peak Detection Method Using Wavelet Transform and Modified Shannon Energy Envelope. Journal of Healthcare Engineering, 2017.
16. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods, 17(3), 261-272.
17. C.-H. Lin, “Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier,” Comput. Math. Appl., vol. 55, no. 4, pp. 680–690, 2008.
18. O. Kwon et al., “Electrocardiogram sampling frequency range acceptable for heart rate variability analysis,” Healthc. Inform. Res., vol. 24, no. 3, pp. 198–206, 2018.
19. P. D. Welch, “Welchs Periodogram.pdf,” IEEE Trans. Audio and electroacoustic, vol. 15. pp. 70–73, 1967.
20. Caridade Gomes, P. M. (2019). Development of an open-source Python toolbox for heart rate variability (HRV) (Doctoral dissertation, Hochschule für angewandte Wissenschaften Hamburg).
21. Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.
22. Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. International Journal of Advanced Computer Science and Applications, 4(4).