International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 4659 - 4667

Machine learning based stress detection system

S. K. Patil[1], S. N. Magdum-Gaji[2], Pranali Dilip Kadu[3], Pragati Rajendra Borase[4], Avanti Ravindra Mokashi[5]

Dept. of E&TC Enggineering, Smt. Kashibai Navale College of Engineering,

Savitribai Phule Pune University, Pune

[1]skpatil_skncoe@sinhgad.edu

[2]shraddha.magdum_skncoe@sinhgad.edu

[3]pranalikadu08@gmail.com

[4]pragatiborase212@gmail.com

[5]avantimokashi2000@gmail.com

Abstract

The main aim of this paper is to design and identify stress using machine learning approach. Proper detection of stress can conveniently prevent many psychological as well as physiological problems such as cardiovascular sicknesses, arrhythmia, diabetes. In our project we will be using three important features from ECG signal to detect stress - QT interval, RR interval and EDR. Using machine learning the model is trained with a set of combinations of ECG features and different SVM types. The model which gives better accuracy is selected. This project will be helpful in biomedical applications if the required performance level is achieved. Until now, only HRV(heart rate variability) is used as the main feature to detect stress. We will be using three important features from ECG signal to detect stress-QT interval, RR interval, EDR. The outcome of this project will be to select the best model with greater accuracy using all the features.

Keywords- *Stress detection, ECG Derived Respiration (EDR), Machine Learning, MATLAB*

I. INTRODUCTION

These days' people live at a frantic pace and in a nearly constant state of competition. In current scenario , mental stress is becoming an unavoidable part of our daily life because of various social and professional responsibilities. In the last two decades, researchers have realized that there is an important relationship between the physical health of an individual and his/her emotional state. Stress is one of the major factors that contributes to physical, emotional, mental and also behavioral changes. When a person is unable to balance between the demands that are placed on him/her and his/her ability to cope with them, then it causes pressure on mental health which creates stress. Stress can be constructive or many times destructive. Constructive stress makes you work, it keeps you active, busy and motivated. On the contrary, destructive stress makes you dull, inactive, scared and you feel lonely. By so far, one can divide stress into two types, Short-term and the other long term or chronic stress. Short-term stress is situational as soon as the situation changes, the stress is gone or over. Long-term stress is a long term problem and causes long term harmful consequences. Stress has a degenerative effect over time.

There are traditional as well as scientific methods to detect people under stress:

A. Questionnaire: Psychiatrist provides a big questionnaire and based on the answers, they decide whether one is under stress or not. This method has its own limitations and drawbacks because many

ISSN: 2233-7853 IJFGCN

times the answers are not factual. Sometimes some of the questions in the questionnaire are not appropriate.

B. The other method is the sensor measuring method. The limitation of this method is, it is time-consuming and a bit expensive. So we are going to use machine learning technique to detect stress using ECG signals. There are several bio-signals available (i.e. ECG, EMG, Respiration, GSR etc.) which are helpful in detecting stress levels as these signals show characteristic changes with stress induction.

Another advantage of using ECG signals is that the respiratory signal information can also be derived using the ECG signals and we don't have to use separate sensor to calculate the respiratory rate.

Different parameters of an ECG such as RR, ST,QT intervals as well as heart rate variability (HRV) are affected by stress. Reduced heart rate variability, increased QT dispersion, and changed respiratory pattern are found to be led by stress. People with these changes have the utmost risk for emerging fatal ventricular arrhythmias." PREVENTION IS BETTER THAN CURE". Therefore, detection or identification of stress is crucial in preventing the initiation of cardiac complexities.

II. DESIGN AND DRAWING

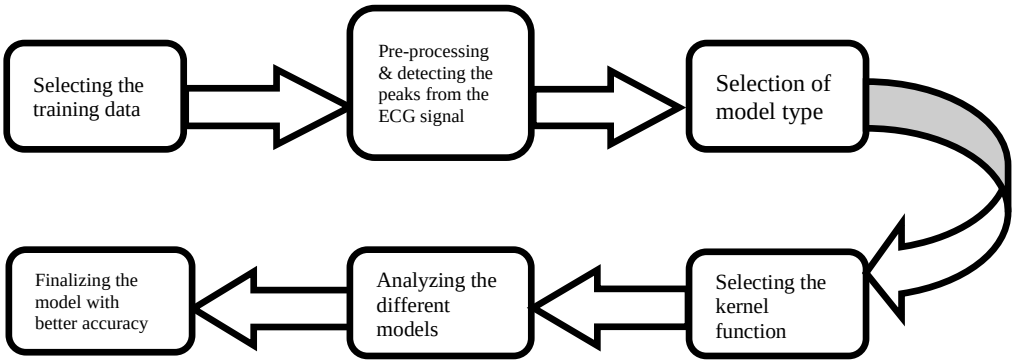

Fig. 1 Design of stress detection System

Block Diagram Description

1. Selecting and training the data:

ECG-ID Database available at Physionet was used where recordings of ECG signal of some individuals were available. The recordings were selected for a duration of 10 seconds. A total of 20 persons recordings were used for analysis.

2. Preprocessing of the ECG signal:

ISSN: 2233-7853 IJFGCN

The ECG signal is corrupted by different high frequency noise such as powerline interference and low frequency noise such as baseline wandering and electrode impedance. Such ECG signals which are contaminated with noise may lead to wrong interpretations. So various filters have been used to denoise the signal.

a. Low pass filter: Electromagnetic fields are caused by a powerline. These are the source of powerline interference in ECG signals as well as other bioelectrical signal recorded from the body surface. It is usually represented by a 50 or 60 Hz sinusoidal interference along with harmonics. Powerline interference should be removed from the ECG signal because it superimposes the low frequency ECG waves such as P wave and T wave.
b. High Pass filter:When the base axis (x axis) of a signal appears to wander or is seen to be tilted above or below the straight line it results in baseline wander. Due to this the entire signal appears to be shifted from its normal base. This is caused due to improper electrodes (electrode-skin impedance), movement and breathing(respiration) of the person.
   For removing this baseline wander we have used Butterworth filter. The output of this filter design will result the ECG signal to be centered along the base axis. The filtered ECG signal is then used for finding the peaks in the ECG signal. After noise removal from the ECG signal, the values for the performance parameter – QT interval, RR interval, EDR (ECG derived respiration) are calculated, and the dataset is prepared which is further used for training the machine learning models.
3. Selection of model type:

The data obtained from the processing of ECG signal is then further used for training and validation. To train and validate the model we have used Classification Learner app from MATLAB's Machine statistics and machine learning toolbox. We have used supervised machine learning to train the data. The performance parameters used are QT interval, RR interval and ECG derived respiration.
There are exactly two class labels in our data- Stressed and Not stressed so we have used support vector machine (SVM) for classification. SVM is a supervised machine learning algorithm which is mostly used for classification problems. In this algorithm each data item acts as a point and it is plotted in n-dimensional space. Here n it represents the number of features or performance parameters. Then classification is done by finding the hyper-plane which distinctively differentiates the two classes. We have trained the data using linear SVM, quadratic SVM, and cubic SVM.

4. Selecting the kernel function:

SVM classifies the data by choosing the proper hyper-plane. The optimal hyper-plane is chosen which segregates the two classes more efficiently. Sometimes the support vectors which are the data points cannot be classified linearly. When linear hyper-plane cannot be implemented, we use additional feature called as the kernel function. By using kernel function, the low dimensional input space is transformed into a higher dimensional space. Thus, in case of non-separable problem we have used kernel function so that the data can be classified into respective classes. We have tested our data by using Gaussian and cubic kernel functions.

5. Analyzing the different models:

ISSN: 2233-7853 IJFGCN

To analyze and study the different features, the model was first trained using only one performance parameter using the default kernel function and the hold-out cross-validation scheme was chosen with a degree of 50% in the classification learner App. By analyzing the scatter plot and confusion matrix the accurate and efficient model is selected.

a. Scatter plots are based on the idea of horizontal and vertical axes, like our basic co-ordinate system, to plot the data points or support vectors. It is beneficial to represent the co-relation between two variables.
b. Confusion matrix can be thought of as a matrix of n rows and n columns which can be used as a tool to evaluate the performance of a classification model. It compares the actual target values with the predicted values. The actual target variable has two values, positive and negative. The columns depict the actual values of the target variable and the rows of the matrix depict the predicted values of the target variable. Thus, proper study of the scatter plot and confusion matrix is carried out and then we have chosen the efficient model.

6. Finalizing the model with better accuracy:

After training the model with different features, with different SVM types and after proper analysis of the scatter plot, confusion matrix for each of the SVM type, the model with best performance is chosen.

	A	B	C	D	E
1	Sr.no	QT Interval	RR Interval	EDR(ECG derived respiration)	Label
2	1	0.378	0.788	0.785	Not Stressed
3	2	0.196	0.82	0.845	Stressed
4	3	0.412	0.86	0.788	Not Stressed
5	4	0.192	0.756	0.9318	Stressed
6	5	0.222	1.04	1.0396	Stressed
7	6	0.364	0.734	0.738	Not Stressed
8	7	0.378	0.7	0.695	Not Stressed
9	8	0.378	0.856	0.9329	Not Stressed
10	9	0.324	0.958	1.0292	Stressed
11	10	0.198	0.976	0.987	Stressed
12	11	0.36	0.764	0.8341	Not Stressed
13	12	0.374	0.694	0.715	Not Stressed
14	13	0.3649	0.726	1.0278	Stressed
15	14	0.372	0.668	0.6971	Not Stressed
16	15	0.368	0.768	0.7044	Not Stressed
17	16	0.374	0.676	0.6717	Not Stressed
18	17	0.342	0.994	1.077	Stressed
19	18	0.254	0.97	1.044	Stressed
20	19	0.224	0.962	0.965	Stressed
21	20	0.421	0.646	0.6418	Not Stressed

Fig. 2 Training Database

III. SOFTWARE DESCRIPTION

- Version: R2020b:

MATLAB is a high performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in mathematical notation.

The name MATLAB stands for matrix laboratory, in industry, MATLAB is a tool of choice for high productivity research, development and analysis.

- The MATLAB language:

ISSN: 2233-7853 IJFGCN

This is high level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features.

- The MATLAB Environment:

This is the set of tools and facilities for managing the variables in the workspace and importing and exporting the data. It also includes tools for developing, managing and debugging. The toolbox which we will be using are:

1.SIGNAL PROCESSING TOOLBOX:

This is helpful to visualize peak points in the signal and also used to visualize abrupt changes in the signal.

2.DSP SYSTEM TOOLBOX:

This is used to remove high-frequency noise in measured data. Removing high-frequency noise allows the signal of interest to be more compactly represented and enables more accurate analysis.

- MATLAB CLASSIFICATION LEARNER APP:

The Classification Learner App trains models to classify data. It is used to explore data, select features, train models and asses the results. Supervised Machine Learning can be performed by supplying a known set of input data(observations and examples) and known responses to the data(eg. Labels or classes). After training multiple models, their validation errors are compared and then the best model is selected.

Fig. 3 Classification learner app tools

IV. TESTING AND DEPLOYMENT

ISSN: 2233-7853 IJFGCN

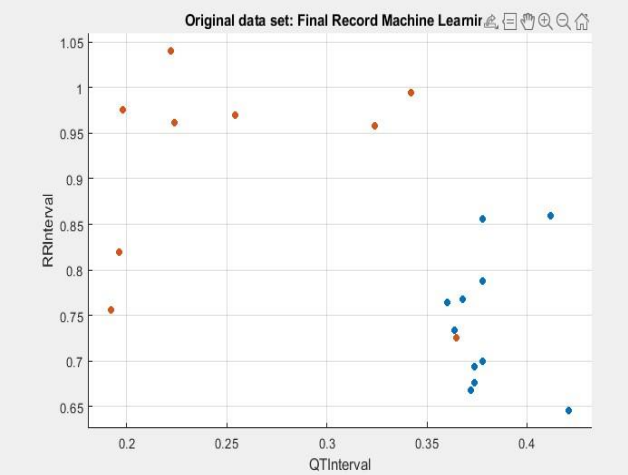

Fig. 4 scatter plot of original data

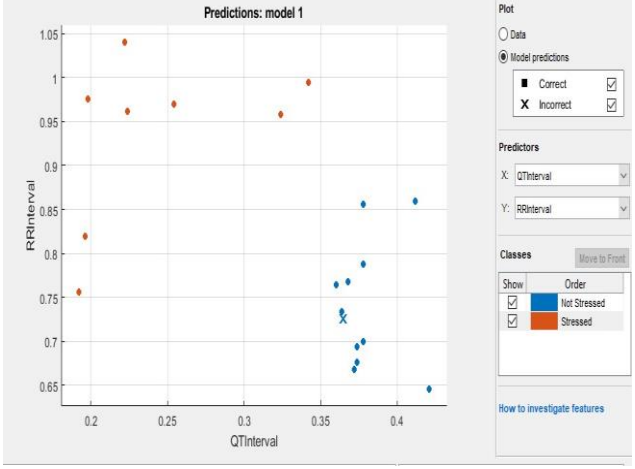

Fig. 5 scatter plot of linear SVM using all the three features

ISSN: 2233-7853 IJFGCN

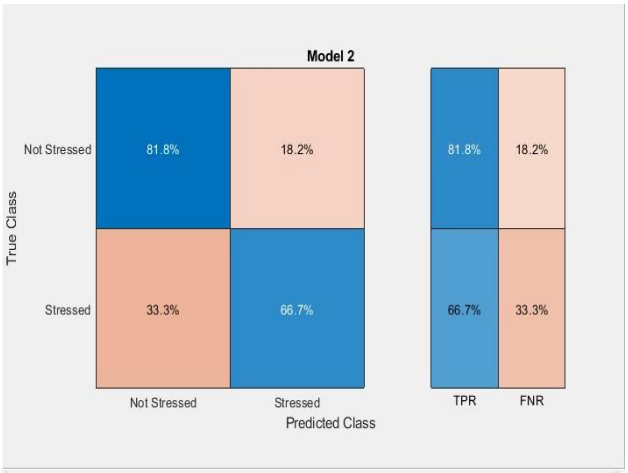

Fig. 6 Confusion matrix of linear SVM using only one feature

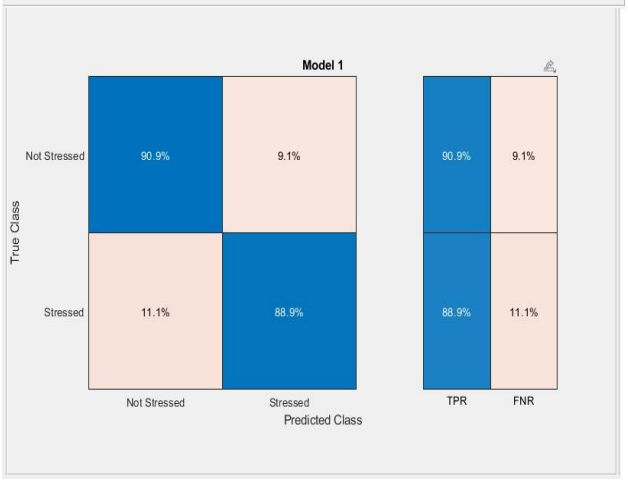

Fig. 7 Confusion matrix of linear SVM using all three features

V. RESULTS AND DISCUSSION

Model accuracy for different SVM types using only one feature:

Model Name	QT interval	RR interval	EDR (ECG derived respiration)
Linear SVM	85%	75%	90%
Quadratic SVM	85%	85%	90%
Cubic SVM	85%	85%	85%

Recall value: 66.7%

Model accuracy for different SVM types using any two features:

Model Name	QT, RR	QT, EDR	RR, EDR
Linear SVM	90%	95%	80%

ISSN: 2233-7853 IJFGCN

Quadratic SVM	90%	95%	80%
Cubic SVM	95%	95%	80%

Recall value: 77.8%

Model accuracy for different SVM types using all the three features:

Model Name	Features used	Accuracy
Linear SVM	QT interval, RR interval,	90%
Quadratic SVM	EDR (ECG derived	95%
Cubic SVM	Respiration	95%

Recall Value: 88.9%

Model accuracy for different SVM types using all the three features and different kernel function.

Model type	Kernel function	accuracy
Linear SVM	Gaussian	90%
	Cubic	95%
Quadratic SVM	Gaussian	90%
	Linear	90%
	Cubic	95%
Cubic SVM	Gaussian	90%
	Quadratic	95%

Recall Value: 88.9%

VI. FUTURE SCOPE & CONCLUSION

1. Thus, by proper comparison and analysis between different SVM models using different kernel functions stress can be identified at early stages preventing further complexities.
2. Traditional methods sometimes do not provide accurate result and there is a delay in remedial action which needs to be taken to prevent further consequences. Thus, the design of this system will be beneficial and will avoid the time delay.

ISSN: 2233-7853 IJFGCN

3. The above work does not provide 100% accuracy. Different techniques and algorithms can be implemented to improve the accuracy.
4. As wearable health devices that monitor everything from blood pressure to respiratory rate become more popular, this idea can be used in those devices for the timely detection of stress.

REFERENCES

[1] K.J.Paavonen, H. Swan, K.Piippo, L.Hokkanen, P.Laitinen, M.Viitasalo Department of Medicine ,University of Helsinki, Haartmaninkatu 4, 00290 Helsinki, Finland ”Response of the QT interval to mental and physical stress in types LQT1 and LQT2 of the long QT Syndrome” 2017

[2] ArpanaBhide, RajasekharDurgaprasad, LatheefKasala, VanajakshammaVelam, NarendraHulikal Department of Physiology, Sri Padmavathi Medical College for Women, Sri Venkateswara Institute of Medical Sciences, Tirupati, Andhra Pradesh, India. Sep 19-2015.

[3] Monika Chauhan, ShivaniVora, DipakDabhi Department of computer Engineering CGPIT, UkaTarsadia University Bardoli, India “Effective Stress Detection using Physiological Parameters” 2017

[4] KarthikSoman, Varghese Alex, Chaithanya Srinivas “Analysis of Physiological Signals in Response to Stress using ECG and Respiratory Signals of Automobile Drivers.” M. Tech Bio Medical Dept. Asst. Prof. Bio Medical Dept. VIT University, Vellore, Tamil Nadu, India.

[5] L.Vanitha Assistant Professor /ECE G.R. Suresh Professor/ECE “HYBRID SVM CLASSIFICATION TECHNIQUE TO DETECT MENTAL STRESS IN HUMAN BEINGS USING ECG SIGNALS” 2013 International Conference on Advanced Computing and Communication Systems (ICACCS -2013), Dec. 19 – 21, 2013, Coimbatore, INDIA .

[6] Physiobank (http://www.physionet.org/physiobank/database/drivedb/)

[7] (https://www.mathworks.com/)

[8] Adnan Ghaderi, JavadFrounchi “Machine Learning-based Signal Processing Using Physiological Signals for Stress Detection” 22nd Iranian Conference on Biomedical Engineering(ICBME 2015), Iranian Research Organization for Science and Technology (IROST), Tehran, Iran, 25-27 November 2015.

[9] Sanjeev Kumar Saini ,Rashmi Gupta “A Review on ECG Signal Analysis for Mental Stress Assessment” Ambedkar Institute of Advanced Communication Technologies & Research (AIACTR), Delhi, India 978-93-80544-34-2019 IEEE.

ISSN: 2233-7853 IJFGCN