International Conference on Communication and Signal Processing, July 28 - 30, 2020, India

A Decision Tree Optimised SVM Model for Stress Detection using Biosignals

Alana Paul Cruz, Aravind Pradeep, Kavali Riya Sivasankar and Krishnaveni K.S

***Abstract*—In our work we propose a machine learning model based on human bio signals to detect human stress. Detecting stress properly can help in preventing a large number of mental and physical scenarios which lead to abnormalities in cardiac rhythm or depression and more. In our work we selected ECG as the bio signal and extracted its features. The advantage of taking ECG as the bio signal is, information about respiratory signals - EDR (ECG Derived Respiration) feature can be easily derived without any extra sensors. Among those unique features we chose ECG derived Respiration, Respiration Rate, QT interval. For training and validation of our new model we used Physionet's "drivedb" database. Our proposed model uses Optimised Support Vector Machines (SVM) using decision trees. Our experimentation results show better accuracy in detecting stress**

***Index Terms*—ECG,Machine Learning, Stress Detection, SVM**

I. Introduction

STRESS can be considered as the physical, mental or emotional phenomenon causing psychological tension. Stress can be externally influenced or internally influenced. It initiates the "fight or flight" response. This response is mainly a reaction between neural network systems. Eustress and Distress are two kinds of stresses. Eustress is the kind of stress that has positive impact to an individual or it is something which stresses a person but ultimately it make them motivated or energized. But on the other hand, distress affects an individual negatively i.e. an individual gets shut down due to stress, anxiety or trauma. Stress leads to a particular type of heart attack without the involvement of sudden breaching of plaques or clogged blood vessels. It is called takotsubo cardiomyopathy, or stress cardiomyopathy [1][2][3]. Fig. 1 shows the other effects due to stress. From this it is evident that proper identification of stress is very crucial. In order to identify stress, we have used Machine Learning techniques.

Alana Paul Cruz is with Department of Computer Science and Engineering, Amrita School of Engineering ,Amrita Vishwa Vidyapeetham, Amritapuri. (e-mail: alanapaulcruz@am.students.amrita.edu).

Aravind Pradeep is with Department of Computer Science and Engineering, Amrita School of Engineering Amrita Vishwa Vidyapeetham, Amritapuri. (e-mail: aravindpradeep@am.students.amrita.edu).

Kavali Riya Sivasankar is with Department of Computer Science and Engineering, Amrita School of Engineering Amrita Vishwa Vidyapeetham, Amritapuri. (e-mail: riyasivasankar@am.students.amrita.edu).

Krishnaveni K.S is with Department of Computer Science and Engineering, Amrita School of Engineering Amrita Vishwa Vidyapeetham, Amritapuri.(e-mail: krishnaveniks@am.amrita.edu)

Electrocardiogram (ECG) was taken as the biosignal to detect stress. ECG [4][5] represents electrical activity of human heart. ECG is a composite of 5 waves - P, Q, R, S and T. Unique ECG features chosen are EDR (ECG Derived Respiration), QT interval and RR interval. Supervised machine learning method i.e. SVM was used for building the model. Physionet's "drivedb" database was used for the purpose.

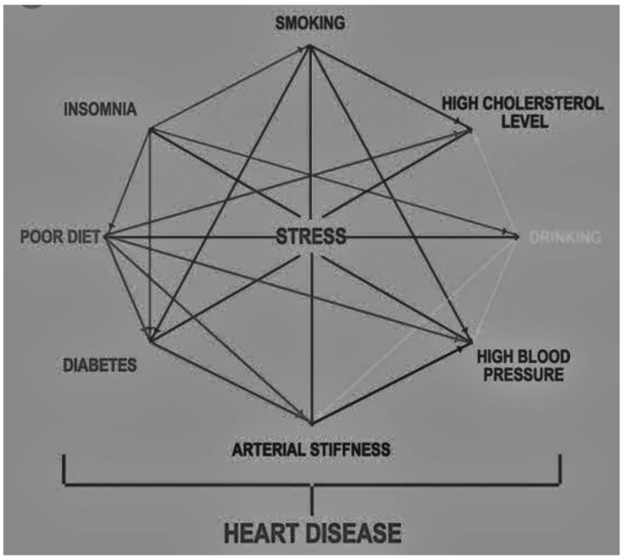

Fig. 1. Effects of Stress

The rest of the paper is organized as below. The related works and data modules are described in Section II and Section III respectively. The methodology is explained in Section IV. Section V discussed about the results. At last, Section VI concludes the paper with conclusion of the work.

II. Related Works

The paper [6] discussed SVM model to detect stress using ECG as the parameter. QT, EDR and RR were the input features used. The dataset was from Automobile drivers' database. It classifies the output into two categories i.e. Stressed or Not Stressed. The model was first trained with SVM models like Linear, Quadratic, Cubic with default kernel function.

978-1-7281-4988-2/20/$31.00 ©2020 IEEE

For training and validation, the Classification Learner app was used which was from MATLAB's Machine Statistics and Machine Learning Toolbox. Accuracy was measured using confusion matrix in MATLAB to find the best SVM model. To find the relevance of the features different combination of features were also used in training. The conclusion drawn was that Cubic SVM model showed higher accuracy rate than other models. Then for further analysis, kernel functions were tuned to improve the performance. In an appropriate feature space, the kernel functions return the inner product between two points and it maps the nonlinear separable dataset into linearly separable one. Here the kernels used for testing were Linear, Quadratic, Cubic and Gaussian. It was found that Gaussian kernel shows higher accuracy level which is shown with the help of Confusion Matrix in Fig. 2. So, from their study it was obvious that Cubic SVM model with a Gaussian Kernel surpassed the other SVM model in accuracy.

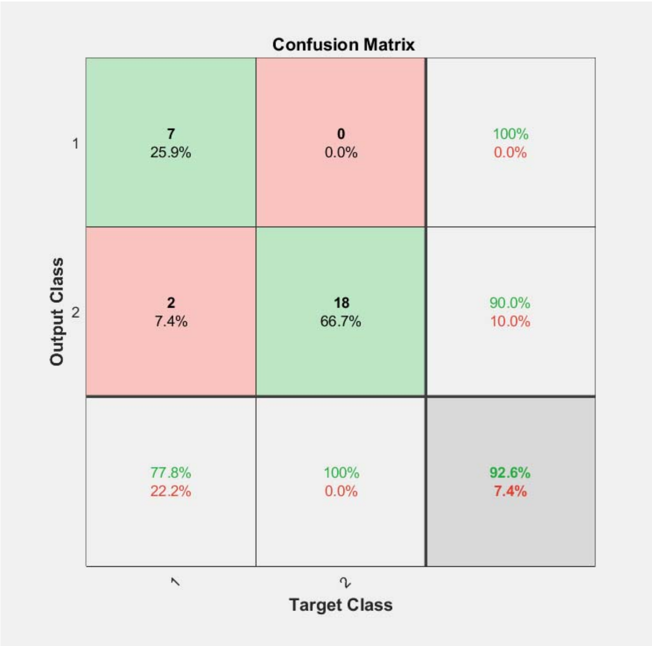

Fig. 2. Confusion Matrix of Gaussian Kernel function used in Cubic SVM

III. Data for Modelling

For designing the model, test study was directed and substantiated for stress detection using database "drivedb" [Stress Recognition in Automobile Drivers] which was taken from the website Physionet. During the process, drivers' physiologic reactions like Electrocardiogram (ECG), Electromyogram (EMG), skin conductivity and respiration were recorded in two different environments. Based on the protocol, the first data were recorded during regular route of the vehicles. Then the drivers were taken in different road conditions with variable traffic conditions where there is possibility of various stress level changes. A total of 27 subject's data were considered. The ECG signal was used for the study.

The people driving in resting state (normal conditions) were considered as "Not Stressed" and the people driving in overwhelming rush hour gridlock courses were considered as "Stressed".

The three main Components of ECG which was taken for the study were (i) QT interval: It is taken from the beginning of QRS complex till the finish of T wave. (ii) RR interval: the time interval between QRS complexes which is in the range 0.6-1.2 seconds (iii) EDR: known as ECG Derived Respiration. The components and intervals of ECG signals were shown in Fig. 3. Technique used to derive Respiratory signal is from ECG signals.

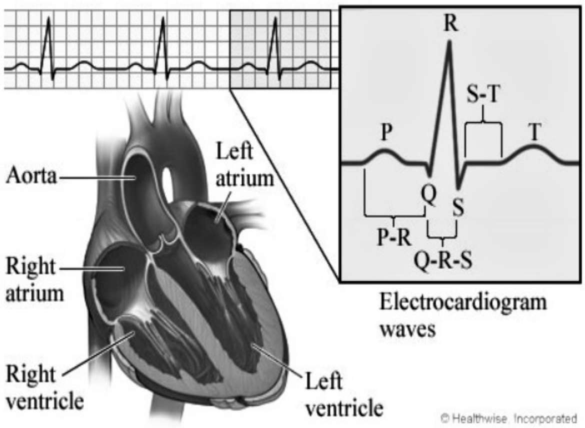

Fig. 3. Components and Intervals of ECG signals

A. *Preprocessing*

It was done to remove the unwanted signals [7]. Two methods have been adopted here. The first one was Baseline Correction [8]. Its for the elimination of negative peaks in the ECG signals. Next process was Wavelet Decomposition, it is for smoothening the signals. Here four levels of decomposition were made.

B. *Feature Extraction*

The second module was Feature Extraction [9][10]. From the ECG signals, localisation of P, Q, R, S, T waves were done to get QT, RR and EDR intervals. Once all the 3 features were extracted,make it into a feature vector [11]. The characteristics of a normal ECG signal is shown below in Fig. 4. The sample dataset is shown in Fig. 5.

Component	Characteristics
Heart Rate	60-100 bpm
PR Interval	0.12–0.20 sec
QRS Interval	0.06–0.10 sec
QT Interval	Less than half of the R-R interval
ST segment	0.08 sec

Fig. 4. Characteristics of a normal Electrocardiogram signal

QT	RR	EDR	DECISION
0.527714285714286	1.37271428571429	0.0126820609275441	Stressed
0.532695652173913	1.66817391304348	0.0521638578301796	Not Stressed
0.489629629629630	1.42281481481481	0.0376711210964073	Not Stressed
0.625333333333333	2.10044444444444	0.0575170098576174	Stressed
0.539040000000000	1.55056000000000	0.0774177590212802	Stressed

Fig. 5. Sample dataset

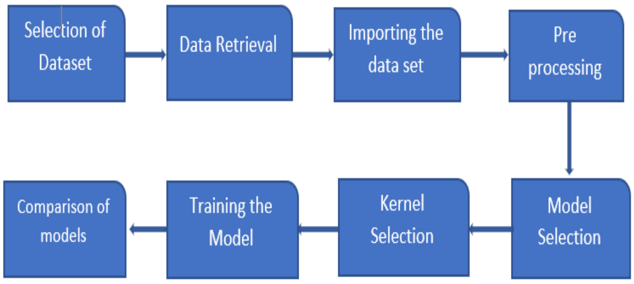

Fig. 6.Workflow of the Proposed model

IV. METHODOLOGY

Stress or Distress happens when a person faces mental strain due to external or internal factors. Stress affects an individual mentally, physically, socially and in many aspects. So, giving timely guidance to the patients is of great significance. For that a more accurate model for detecting stress is necessary for helping doctors in consulting.

A. Performance Measures

The evaluation metric used in this study is Accuracy which is calculated using the help of Confusion Matrix. The terms in that is defined as:

1. **True Positive (TP):** It detects the condition when the actual and predicted value is of positive category.

2. **True Negative (TN):** It denotes the condition when the actual and predicted value is of negative category.

3. **False Negative (FN):** It denotes the condition when the actual is of positive category and predicted value is of negative category.

4. **False Positive (FP):** It denotes the condition when the actual is of negative category and predicted value is of positive category.

5. **Accuracy:** It denotes the proportion of correctly predicted observations to the total number of samples.

$$\text{Accuracy} = \frac{(TN+TP)}{(TN+TP+FN+FP)}$$

6. **Sensitivity (Recall/True Positive Rate):** When the output is really true, and how often it is correctly predicted as true result.

B. Proposed System

Initially, the model was trained using Cubic SVM with Gaussian Kernel. For a better model, here we have used Tree Optimised SVM which is a combination of Decision Tree and SVM algorithms. The flow of the proposed system is shown in Fig. 6.

Decision Trees are a kind of Supervised Machine Learning algorithm used for predictive modelling where the data is continuously split according to a several input parameters [12]. The classification is done using various QRS detection algorithms and other functions in MATLAB. In the Tree based SVM, we have the training data and the testing data as the inputs. In testing data, each level (or stage) has the prediction probability. When we give one test data, it is compared with the first training sample, it will give some weight in the first process, similarly the second test data is compared with second training sample in the second process, it will also generate a weight. So, like a tree, first stage is the first weight. From the first weight, next nearest weights are calculated. Weight is the comparison result value (i.e. The test data when compared with trained). Highest weight was taken as the tree point. Probability tree point changes based on the weight of comparison result. Based on this the SVM does the classification. Tree optimised SVM is tested on Linear, Quadratic and Cubic SVM models to find better performance among it [13]. The bio signal-based Stress detection system using Tree optimised Cubic SVM [14] shows much accuracy than the already existing models as shown in Table I. Thus, Elapsed time and Sensitivity values are showing improvements in our model.

TABLE I
MODEL PERFORMANCE

MODEL TYPE	ACCURACY
Cubic SVM with Gaussian	92.6%
Tree Optimised SVM	96.3%

V. PERFORMANCE RESULTS

Tree Optimised Cubic SVM is providing the better accuracy which is demonstrated in Table II. The main enhancement of this Tree Optimised SVM model is that it shows improvement in Sensitivity and Elapsed. It is able to generate an accuracy of 96.3% which is shown in Fig. 7. This model will be able to do better detection of stress.

TABLE II
PERFORMANCE OF THE PROPOSED SYSTEM MODEL

MODEL TYPE	KERNEL	ACCURACY
Tree Optimised SVM	Linear	88.9%
	Quadratic	92.6%
	Cubic	96.3%

Objective function is the optimisation of SVM model hyper parameters. For every iteration, the results of objective function are shown in these Fig. 8. Objective function has to below for efficient SVM model and observed objective should be same as estimated objective value.

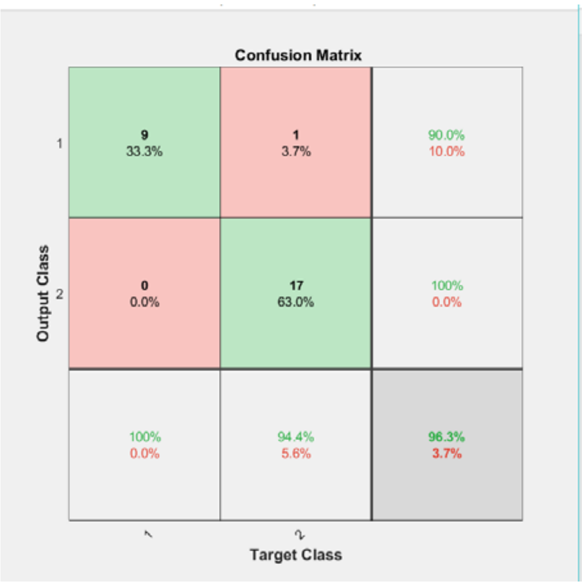

Fig. 7. Confusion Matrix of Tree Optimised cubic SVM

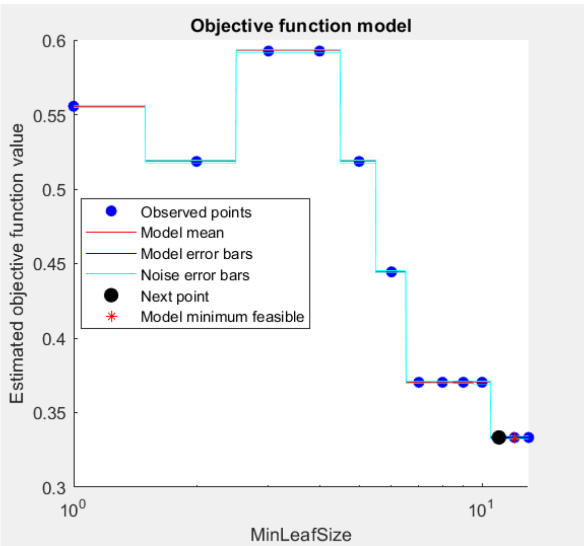

Fig. 8. Objective function model

Estimated objective function value was found to be 0.33336 and Estimated function evaluation time was 0. 06634.It has been shown in Fig. 9.These are the optimisation function values. As objective value become less,it denotes a better optimised SVM model.

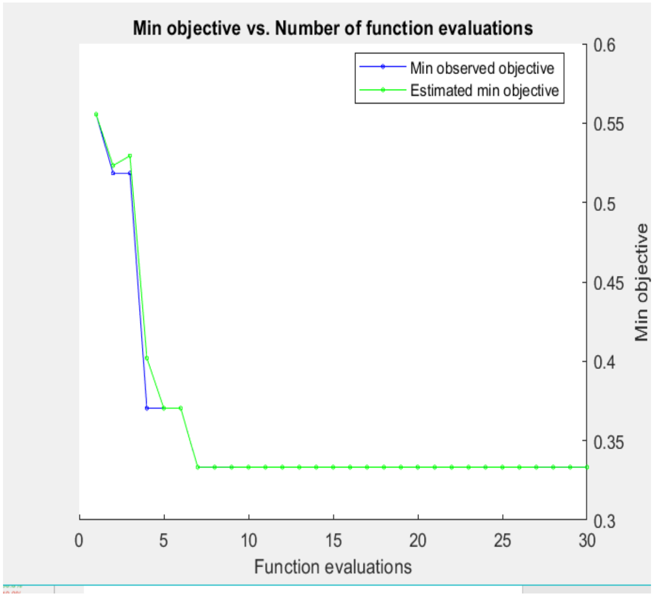

Fig. 9. Minimum Objective versus Number of function Evaluations

VI. CONCLUSION

With our proposed model we attained a better performance than earlier findings. Our model with Tree optimised Cubic SVM shows more accuracy in identifying stress when compared to already existing models. With our accurate model we can take remedial measures to reduce health risks.

REFERENCES

[1] Yoshihiro J. Akashi, David S. Goldstein, Giuseppe Barbaro, and Takashi Ueyama "Takotsubo Cardiomyopathy-A New Form of Acute, Reversible Heart Failure" on 16 Dec 2008 in Circulation. 2008;118:2754–2762

[2] Alexander R Lyon, Paul SC Rees, Sanjay Prasad, Philip A Poole-Wilson & Sian E Harding "Stress (Takotsubo) cardiomyopathy—a novel pathophysiological hypothesis to explain catecholamine-induced acute myocardial stunning" in *Nature Clinical Practice Cardiovascular Medicine* volume 5, pages 22–29(2008) [**4**]Joseph A. Boscarino, Ph.D., M.P.H., Jeani Chang, M.P.H. "Electrocardiogram abnormalities among men with stress-related psychiatric disorders: Implications for coronary heart disease and clinical research" in *Annals of Behavioral Medicine*, Volume 21, Issue 3, September 1999, Pages 227–234

[3] Y.J. Akashi, K. Nakazawa, M. Sakakibara, F. Miyake, H. Koike, K. Sasaka "The clinical features of *takotsubo* cardiomyopathy" in *QJM: An International Journal of Medicine*, Volume 96, Issue 8, August 2003, Pages 563–573

[4] Joseph A. Boscarino, Ph.D., M.P.H., Jeani Chang, M.P.H. "Electrocardiogram abnormalities among men with stress-related psychiatric disorders: Implications for coronary heart disease and clinical research" in *Annals of Behavioral Medicine*, Volume 21, Issue 3, September 1999, Pages 227–234

[5] Koji Matsuoka, MD Setsuya Okubo, MD Eitaro Fujii, MD Fumiya Uchida, BA Atsunobu Kasai, MD Toshikazu Aoki, MD Katsutoshi Makino, MD Chikaya Omichi, MD Naoki Fujimoto, MD Satoshi Ohta,

MD Toshiki Sawai, MD Takeshi Nakano Evaluation of the arrhythmogenecity of Stress-Induced "Takotsubo cardiomyopathy" from the time course of the 12-lead surface electrocardiogram in VOLUME 92, ISSUE 2, P230-233, JULY 15, 2003

[6] Md Fahim Rizwan,Rayed Farhad,Farhan Mashuk,Fakhrul Islam,Mohammad Hasan Imam "Design of a Biosignal Based Stress Detection System Using Machine Learning Techniques" in International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST) 2019

[7] Haritha C, Ganesan. M, Sumesh E.P," A survey on modern trends in ECG noise removal techniques", 2016 IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016.

[8] Zhi-Min Zhang,Shan Chen, Yi-Zeng Liang "Baseline correction using adaptive iteratively reweighted penalized least squares" in Analyst 1 May 2010, issue 5, Page 809 to 1156

[9] S. Z. Mahmoodabadi1(MSc), A. Ahmadian1 (Phd), M. D. Abolhasani1(Phd) "ECG feature extraction using daubechies wavelets" in Proceedings of the Fifth IASTED International Conference VISUALIZATION, IMAGING, AND IMAGE PROCESSING September 7-9, 2005, Benidorm, Spain

[10] S.Z. Mahmoodabadi ; A. Ahmadian ; M.D. Abolhasani ; M. Eslami ; J.H. Bidgoli "ECG Feature Extraction Based on Multiresolution Wavelet Transform" in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference

[11] Adeena K.D ,Remya.R, "Extraction of relevant dataset for support vector machine training: A comparison", International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015

[12] Anjali T, Rakesh N, and Akshay Krishna M P," A Novel based Decision tree for Content Based Image Retrieval: An Optimal Classification Approach" in 7th IEEE International Conference on Communication and Signal Processing, ICCSP 2018.

[13] Sukanya M.V, SathyadevanS, Unmesha Sreeveni," Benchmarking support vector machines implementation using multiple techniques", Advances in Intelligent Systems and ComputingVolume 320, 2015, Pages 227-238

[14] Binoy B. Nair,V.P Mohandas,N.R. Sakthivel "A Genetic Algorithm Optimized Decision TreeSVM based Stock Market Trend Prediction System" in International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2981-2988