Performance analysis of Machine Learning techniques for classification of stress levels using PPG signals

Kushagra Awasthi
Dept. of Electronics and Communication
Ramaiah Institute of Technology
Bangalore, India
kushagra.awasthi33@gmail.com

Pranav Nanda
Dept. of Electronics and Communication
Ramaiah Institute of Technology
Bangalore, India
pranavnanda9@gmail.com

Dr. Suma K V
IEEE Senior Member
Dept. of Electronics and Communication
Ramaiah Institute of Technology
Bangalore, India
sumakv@msrit.edu

***Abstract*— Psychological stress has adverse effects on our autonomic nervous system. The main aim of this research work is to compare the performance and accuracy of various Machine Learning (ML) models used for classification of subjects as having average stress or high stress. Photoplethysmography (PPG) signals is recorded for all subjects after they are introduced to stress inducing stimulus. From the PPG signals Pulse Rate Variability (PRV) parameters are calculated and on the basis of these PRV parameters the subject is either classified as having average stress or high stress. A dataset is framed from the PRV parameters of the subjects and ML models namely Logistic Regression, Support Vector Machine (SVM), Decision Tree and Random Forest are trained and tested, for classification of subjects as average stress or high stress. The models are evaluated based on performance parameters such as confusion matrix, accuracy score, area under curve of Receiver Operating Characteristic (AUC-ROC) of the model, Mean Square Error of the model and other performance parameters. These ML models with good prediction accuracy can be used as an aid by the doctors for more accurate detection of subjects having high stress and thus, begin with the treatment at the earliest.**

Keywords—Pulse rate variability, logistic regression, Photoplethysmography, Stress, Random forest, SVM.

I. Introduction

The advent of machine learning (ML) has seen many applications in the healthcare industry. ML is widely being used in the realm of medical data analysis to enable healthcare professionals to not only predict ailments from existing conditions but to also draw patterns from patient history. Data can sometimes be overwhelming and recognizing a trend, more often than not, becomes challenging to the human mind. ML helps in the analysis of such data and makes hard-to-diagnose diseases a relatively simpler task. Pulse rate variability is one of the most reliable indicator of stress and Photoplethysmography (PPG) is a simple and low-cost optical technique which is efficient in the computation of PRV [1]. Various time, frequency and geometric domain parameters are obtained by using PPG signals called the PRV parameters. Some of these parameters are utilized by our machine learning models for analysis. Here we incorporate established algorithms such as logistic regression, SVM and Random forest to arrive at some inference from the PPG signal data. We aim to provide a mathematical framework backing up the viability of the PPG based diagnostic scheme for stress and its potential to be used alternatively to HRV (Heart rate variability), even during non-stationary conditions. It also provides an insight to the prominent parameters that could be kept in mind while arriving at a diagnosis for stress. The autonomic nervous system (ANS) controls involuntary activities such as heartbeat and has two main divisions vis-á-vis sympathetic and parasympathetic systems. The sympathetic system is triggered during stressful situations and it automatically raises the heart rate to facilitate expansion of the bronchi to help breathe easy. The para-sympathetic system on the other hand helps maintain the heart rate and brings the body back to the normal state. This understanding is crucial in coming to a statistical inference based on frequency and geometric domain parameters.

II. Literature Review

Suma P et al. [1] incorporate a device to compute PRV and PPG signal. The subject is shown a stressful audio-visual stimulus and PPG signal is recorded. Based on these signals, PRV parameters are calculated in time, frequency and geometric domains in order to assess the level of stress induced. PPG signal acquired from the device is compared with ECG signal for validation. It was found that there was significant decrease in SDNN (Standard Deviation of NN intervals) and RMSSD (Root Mean Square of Successive Differences) values and sufficient increase in SD2/SD1in the presence of stimulus which resulted in decrease in total PRV and indicates that the subject is stressed.

Arun Kumar M et al. [2] describe how Pulse Rate Variability (PRV) can be used for determining stress levels. PRV parameters give measurement similar to that of Heart Rate Variability which is measured by ECG. SDNN value greater than 50 is considered normal stress and less than 50 is abnormal stress. RMSSD value less than 10 indicates normal stress whereas RMSSD greater than 10 refers to abnormal stress. SD2 is considered as the non-linear analysis parameter, subjects having SD2 value greater than 64 are considered to be stressed.

Keeping in mind the long term goal of continuous blood pressure (BP) estimation, Elgendi et al [5] have compiled a review of several theoretical approaches used in BP measurement using PPG. Traditional techniques such as feature extraction of PPG signal followed by machine learning are mentioned. Owing to the advances in signal processing and machine learning, the scope of developing a wearable embedded device is also highlighted. One of the challenges in

978-1-7281-6828-9/20/$31.00 ©2020 IEEE

developing such a device is the design of an efficient filter which processes the PPG signal and makes it suitable for feature extraction. On a different note, the need of additional PPG features (such as PRV parameters) to analyze correlation with blood pressure is described, and our analysis on the PRV parameters would bear a positive effect on research in this area. Padmavathi Janardhanan et al. [9] have carried out a comparative study among classification algorithms to analyze heart data, breast cancer data and diabetes data. Naïve Bayes network, SVM and radial basis function network were used to analyze all three datasets and it was found that the SVM gave consistently better results than the other two algorithms. Various kernels like sigmoidal, polynomial and RBF kernels were used in the working of SVM. Accuracy of the SVM classifier for the three datasets were found to be 91%, 93%, and 86% respectively, which is significantly higher than the others. Joanne Peng et al. [10] convey a list of guidelines that are useful in tabulation and performance analysis of logistic regression models. A thorough introduction to the mathematical background required to implement a logistic regression model is given.

Arnu Pretorius et al. [11] have compiled a list of all novel random forest (RF) algorithms from 2000 to 2015. Information on performance measures such as ROC and area under the curve is provided. Statistical validation techniques such as t-test and kappa test are elaborated to help choose the right algorithm for the problem at hand. The source code used to develop RF was also given. Leo Breiman et al. [12] have given the complete theoretical background for RF and other important topics like feature selection at each node.

III. Methodology

Aim of this research work is comparative study of various machine learning models in classifying data as average or high stress based on the PRV parameters dataset. Photoplethysmography (PPG) was initially used to record the PRV parameters of 35 subjects. Based on the PRV parameters the subject was classified as Average Stress or High Stress. Table 1 shows an example of 5 subjects to illustrate how the dataset was framed.

A. *Data Acquisition*

The subjects who took part in the study comprised of healthy males and females in the age group of 18-24. To gauge the effect of stress in the months leading up to the study, the subjects were made to fill a Perceived Stress Scale (PSS) questionnaire. They were free of alcohol and caffeine for at least two hours prior to the study and general details like age, height and weight were recorded before commencing the data acquisition.

A pulse rate sensor interfaced with a microcontroller was used to acquire the PPG signal. Initially the subject was in a relaxed position for 15 minutes after which the PPG readings were noted from b to c as shown in Fig 1. A stress inducing audio visual stimulus was given to the subject from d to e, during which another set of PPG readings were noted. The recovery time mentioned from e to f is the time taken by the subject's PPG readings to return to normal.

The readings shown in Table 1 correspond to the duration from d to e (stressed state). Our analysis was carried out considering only the PPG readings corresponding to the stressed state. Using Label encoding, Average Stress was assigned dummy variable 0 and High Stress was assigned 1.

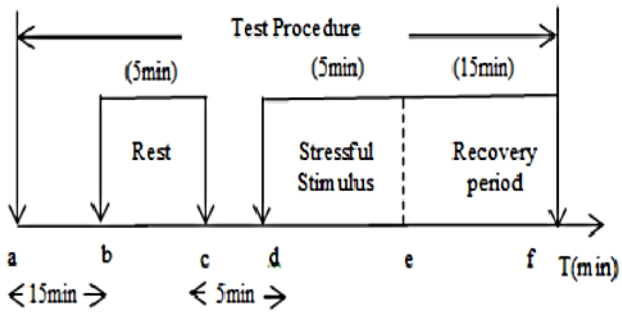

Fig 1. Timeline of test procedure [1]

B. *PRV Parameters*

The description and significance of parameters along with their unit and domain of relevance is listed below [1].

- BPM stands for beats per minute and measures pulse rate
- RR interval is the time elapsed between two successive normal R-waves of the PPG signal and is measured in milliseconds. The entire time domain analysis is computed based on RR interval.
- SDNN is the standard deviation of the RR intervals. It gives an idea of how well the heart responds to stress and estimates the overall pulse rate variability. Readings below the 20 millisecond mark is considered as an indication of long term stress related issues.
- RMSSD is the root mean square of successive differences in the RR interval. It is also a time domain parameter. It estimates high frequency variations in pulse rate for short term RR recordings. A value below 10 ms results in the development of cardiac disease.
- LF/HF is a frequency domain parameter. It is the ratio of low frequency to high frequency. It gives the ratio of

Subjects	pulse rate	RR Interval	SDNN	RMSSD	LF/HF	SD2/SD1	Recovery Time	Stress
1	87	590	38.68	27.38	1.04	2.74	4	0
2	76	626	49.26	34.40	1.02	2.85	3	0
3	82	654	53.01	56.91	1.03	2.78	5	1
4	89	582	38.45	26.52	1.05	2.41	6	1
5	80	694	37.39	38.62	1.04	2.26	6	1

Table 1. Dataset

overall balance between sympathetic and parasympathetic systems.

- SD2/SD1 is a geometric domain parameter which is the ratio of standard deviations and reflects sympathetic to parasympathetic activity.

For brevity, only the block diagram is shown in Fig 2, without the equations behind computing the parameters.

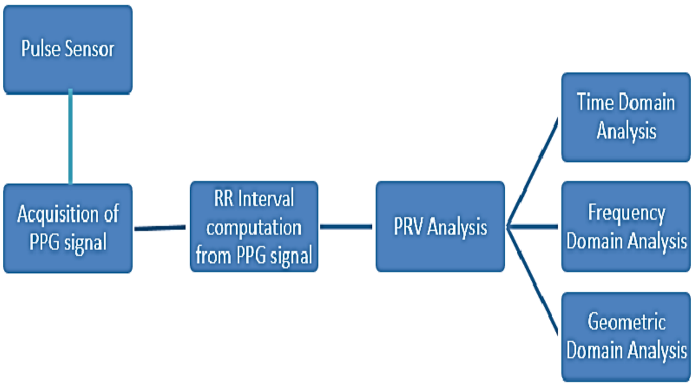

Fig 2. Framework for the PPG monitoring system [1]

IV. Machine Learning Models

In this section, the various types of machine learning models for classification of data and the result obtained from each is given.

A. Logistic Regression

Logistic Regression is a supervised learning classification algorithm. It uses the concept of probability for predictive analysis. Logistic Regression is different from Linear Regression as it uses sigmoid function as cost function, shown in equation 1.

$$f(x) = \frac{1}{1+e^{-(x)}} \qquad \ldots\ldots \quad (1)$$

Sigmoid Function

The Logistic Regression used for the classification for the given dataset used RR interval and Recovery time as significant features for classification. The training to test data ratio taken was 7:3. This Logistic Regression model gave an accuracy of 82% and AUC-ROC curve 93%.The confusion matrix and other performance parameters are given in the classification report is given in Fig. 3

```
Confusion Matrix :
[[7 0]
 [2 2]]
Accuracy Score is 0.8181818181818182
Classification Report :
              precision    recall  f1-score   support

           0       0.78      1.00      0.88         7
           1       1.00      0.50      0.67         4

    accuracy                           0.82        11
   macro avg       0.89      0.75      0.77        11
weighted avg       0.86      0.82      0.80        11

AUC-ROC: 0.9285714285714286
LOGLOSS Value is 6.279777526347397

R Squared = 0.2142857142857143
MAE = 0.18181818181818182
MSE = 0.18181818181818182
```

Fig. 3

B. Support Vector Machine (SVM)

SVM is also a supervised machine learning algorithm that can be used for classification as well as regression problems. To find the optimal boundary between the possible outputs it performs transformation data by using kernel trick technique. It uses more complex hinge loss function as cost function, shown in equation 2.

$$c(x, y, f(x)) = \begin{cases} 0, & \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases} \qquad \ldots\ldots \quad (2)$$

Hinge Loss Function

SVM model used for classification of dataset uses a linear kernel and RR interval and Recovery time were again the prominent features of the model. The training to test data ratio used was 8:2, this resulted in an accuracy of 86% and AUC-ROC value of 90%.The confusion matrix and other performance parameters for SVM model are given below in the classification report in Fig. 4

```
Confusion Matrix :
[[2 0]
 [1 4]]
Accuracy Score is 0.8571428571428571
Classification Report :
              precision    recall  f1-score   support

           0       0.67      1.00      0.80         2
           1       1.00      0.80      0.89         5

    accuracy                           0.86         7
   macro avg       0.83      0.90      0.84         7
weighted avg       0.90      0.86      0.86         7

AUC-ROC: 0.9
LOGLOSS Value is 4.93411091355867

R Squared = 0.30000000000000016
MAE = 0.14285714285714285
MSE = 0.14285714285714285
```

Fig. 4

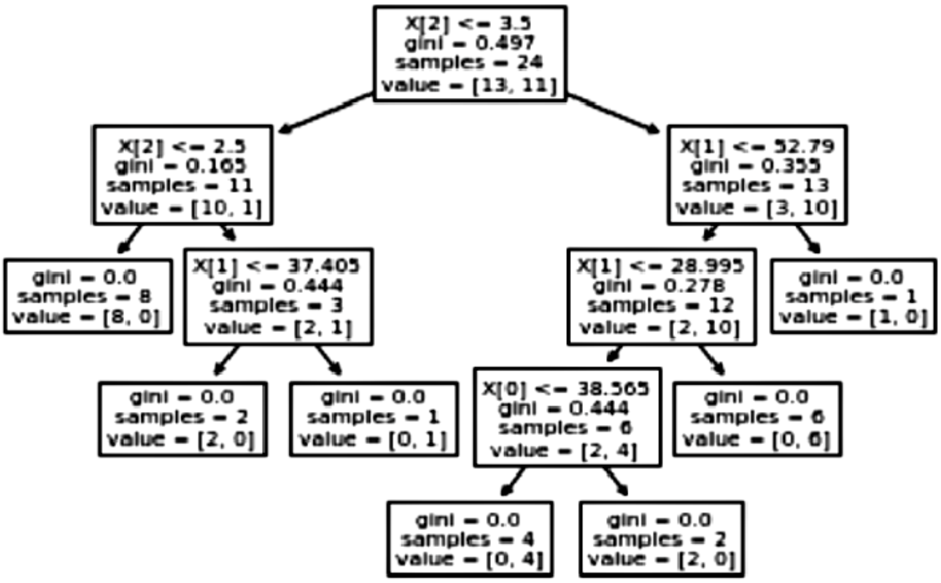

Fig. 5

C. *Decision Tree and Random Forest*

Decision Tree is a hierarchical model for supervised learning and is also a non-parametric model in the sense that we do not assume any parametric form for the class densities and the tree structure is not fixed priori but grows during learning depending on the complexity of inherent data. Whereas Random Forest is a model which is made by a combination of number of decision trees the only difference is that in a random forest model the decision trees are made by randomly sampling test data and while splitting of nodes random subsets of features are considered.

Fig. 5 represents the decision tree for the given dataset. SDNN, RMSSD and Recovery time are used as prominent features for the given model. The given decision tree uses gini impurity measure for classification. An accuracy of 82% is obtained from the given decision tree model and an AUC-ROC value of 80%.When a Random forest model is used, accuracy increases to 91% and the AUC-ROC value obtained is 96%. The classification reports for decision tree and random forest are given in Fig 6 and Fig 7 respectively

```
Confusion Matrix :
[[6 1]
 [1 3]]
Accuracy Score is 0.8181818181818182
Classification Report :
              precision    recall  f1-score   support

           0       0.86      0.86      0.86         7
           1       0.75      0.75      0.75         4

    accuracy                           0.82        11
   macro avg       0.80      0.80      0.80        11
weighted avg       0.82      0.82      0.82        11

AUC-ROC: 0.8035714285714286
LOGLOSS Value is 6.279850217022869
```

Fig. 6

```
Confusion Matrix :
[[7 0]
 [1 3]]
Accuracy Score is 0.9090909090909091
Classification Report :
              precision    recall  f1-score   support

           0       0.88      1.00      0.93         7
           1       1.00      0.75      0.86         4

    accuracy                           0.91        11
   macro avg       0.94      0.88      0.90        11
weighted avg       0.92      0.91      0.91        11

AUC-ROC: 0.9642857142857142
LOGLOSS Value is 3.1398887631736994
```

Fig. 7

V. RESULTS AND DISCUSSION

The results obtained by using various machine learning models for classification of average and high stress subjects can be analyzed based on its classification report. After analysis we will arrive at the conclusion that which machine learning model is more accurate in classifying the subjects.

A. *Analysis of Logistic Regression*

Logistic Regression model gave an accuracy of 82%. Sensitivity of the given model is 100% and Specificity is 75%. One of the most important performance metric in binary classification is area under the ROC curve, more is the area under the curve better is the model, AUC-ROC of the Logistic Regression model is 92% (Fig. 8) and also the mean square error is 0.18 which is very less and hence it is an add-on to the performance of the model.

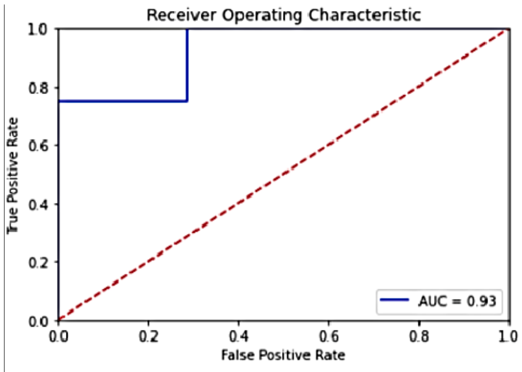

Fig. 8

B. Analysis of SVM Model

SVM model gives accuracy of 86% which is 4% more than that of logistic regression. The reason for better accuracy of SVM is that it is using a more complex cost function thereby increasing the classification accuracy. The model has specificity of 67% and high sensitivity of 100%. The area under the ROC curve is 90% (Fig. 9) and MSE of 0.14 indicates that performance of the SVM classifier is good.

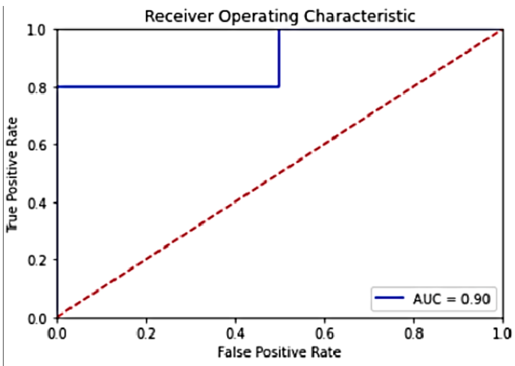

Fig. 9

C. Analysis of Decision Tree and Random Forest

Decision Tree classifier has an accuracy of 82% whereas in Random Forest the accuracy increase to 91% this increase in accuracy is because random forest is combination of different trees which were trained by random data and random subset of features. Random forest also dominates in the major performance measure i.e. area under the ROC curve. The AUC-ROC of decision tree is 80% (Fig.10) whereas for random forest it is 96% (Fig.11). Specificity and sensitivity for the decision tree is 86% and 75% respectively, the same metrics for random forest was found to be 88% and 100% respectively.

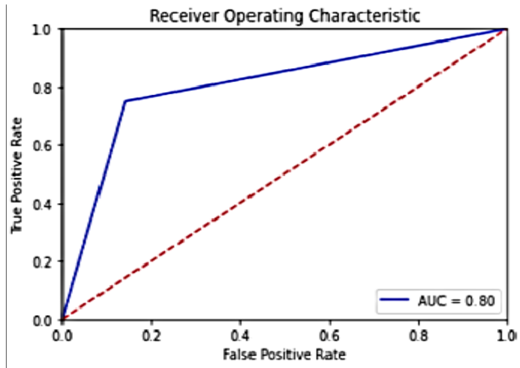

Fig. 10

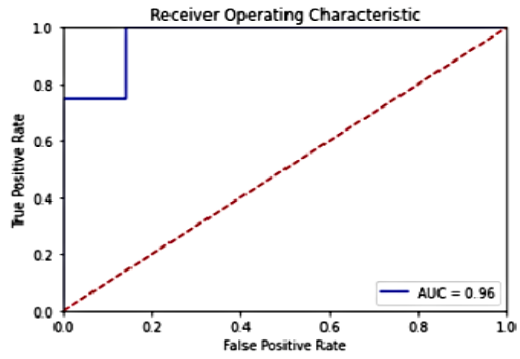

Fig. 11

Performance Metrics	Logistic Regression	Support Vector Machine	Decision Tree Classifier	Random Forest
Accuracy Score	0.82	0.86	0.82	0.91
AUC-ROC	0.93	0.90	0.80	0.96
Specificity	0.78	0.67	0.86	0.88
Sensitivity	1.00	1.00	0.75	1.0
Confusion Matrix Detail	Out of 11 test inputs 7 were correctly classified as average stress and 2 were falsely classified as high stress and 2 were correctly classified as high stress.	Out of 7 test inputs 2 were correctly classified as average stress and 4 were correctly classified as high stress and 1 was falsely classified as high stress	Out of 11 test inputs 6 were correctly classified as average stress and 1 was falsely classified as high stress, whereas, 3 were correctly classified as high stress and 1 was falsely classified as average stress.	Out of 11 test inputs 7 were correctly classified as average stress and 3 were correctly classified as high stress and only 1 was wrongly classified as average stress.
Root Mean Square Error	0.214	0.30	0.214	0.607

Table 2. Comparative study of ML models used

VI. Conclusion and Future Work

Random Forest being an ensemble learning technique is better at handling noise than SVM. Random forest provides us a great advantage in data pre-processing, it ignores near zero variance features and in this we do not have to normalize the features whereas, in SVM, feature scaling is required to obtain a better result. Selecting an appropriate kernel is also difficult task in SVM, as without having proper knowledge about the dataset we can't decide which kernel to be used. As we know according to the 'no free lunch' theorem of Wolpert and Macready that different algorithms give different optimization results with a particular dataset, therefore we need to test a given dataset with various models and based on the performance metrics we decide which algorithm is best for the given dataset. For the PPG dataset, we can clearly see from Table 2 that Random Forest classifier clearly emerges out as the best classifier followed by SVM and logistic regression respectively.

The development of portable single spot PPG monitoring devices can be used as an alternative measurement to HRV even during non-stationary conditions. PPG sensors can be placed on the fingertip and wrist which offers more flexibility to users. The development of wearable devices for monitoring biomedical signals suggests that it is possible to monitor pulse rate through PPG signals due to the simplicity of PPG waveforms. With the development of Android app to display PRV parameters, remotely monitoring the stress level becomes possible [1][2]. Furthermore, deep learning can be utilized as more data is collected and a custom filter can also be designed to process the PPG signal for traditional feature extraction [5].

Acknowledgment

We would like to express our gratitude to Suma P, Department of Electronics and Communication, Ramaiah Institute of technology for her continuous support in making this statistical analysis possible. We would also like to thank Dr. Arun Kumar M and Dr. Venkatesh D of the Physiology Department, Ramaiah Medical College for their valuable inputs in our work.

References

[1] P. Suma and K. V. Suma, "Pulse Rate Variability for Detection of Autonomic Tone of an Individual," 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2019, pp. 1172-1177.

[2] M, ARUN. (2016). CONCEPT OF MEASURING PULSE RATE VARIABILITY AS A TOOL FOR DETERMINATION OF STRESS IN HEALTHY INDIVIDUALS. Journal of evolution of research in Human Physiology. 2. 14-28.

[3] Lan, K., Raknim, P., Kao, W. *et al.* TOWARD HYPERTENSION PREDICTION BASED ON PPG-DERIVED HRV SIGNALS: A FEASIBILITY STUDY. *J Med Syst* 42, 103 (2018). https://doi.org/10.1007/s10916-018-0942-5

[4] M, ARUN & D, Venkatesh. (2015). NOVELTY OF UNFILTERED DATA IN CARDIO-RESPIRATORY WAVES RECORDING. Journal of Evolution of Research in Human Physiology. 1. 18-21.

[5] Elgendi, M., Fletcher, R., Liang, Y. *et al.* The use of photoplethysmography for assessing hypertension. *npj Digit. Med.* **2,** 60 (2019). https://doi.org/10.1038/s41746-019-0136-7

[6] Tharion, Elizabeth & Parthasarathy, Sangeetha & Neelakantan, Nithya. (2009). SHORT-TERM HEART RATE VARIABILITY MEASURES IN STUDENTS DURING EXAMINATIONS. The National medical journal of India. 22. 63-6.

[7] Susmaga R. (2004) CONFUSION MATRIX VISUALIZATION. In: Kłopotek M.A., Wierzchoń S.T., Trojanowski K. (eds) Intelligent Information Processing and Web Mining. Advances in Soft Computing, vol 25. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39985-8_12

[8] Matsumoto Y, Mori N, Mitajiri R, et al. STUDY OF MENTAL STRESS EVALUATION BASED ON ANALYSIS OF HEART RATE VARIABILITY. Journal of Life Support Engineering 2010;22(3):105–11.

[9] Marcus Vollmer.A ROBUST, SIMPLE AND RELIABLE MEASURE OF HEART RATE VARIABILITY USING RELATIVE RR INTERVALS. 2015 Computing in Cardiology Conference (CinC), pp. 609-612, 6 September 2015.

[10] Janardhanan, Padmavathi & Heena, L. & Sabika, Fathima. (2015). Effectiveness of Support Vector Machines in Medical Data mining. Journal of Communications Software and Systems. 11. 25-30. 10.24138/jcomss.v11i1.114.

[11] Peng, Joanne & So, Tak-Shing. (2002). Modeling strategies in logistic regression with SAS, SPSS, Systat, BMDP, Minitab, and STATA. Journal of Modern Applied Statistical Methods. 1. 147-156. 10.22237/jmasm/1020255720.

[12] Pretorius, Arnu & Bierman, Surette & Steel, Sarel. (2016). A meta-analysis of research in random forests for classification. 1-6. 10.1109/RoboMech.2016.7813171.

[13] LEO BREIMAN Statistics Department, University of California, Berkeley, CA 94720. Machine Learning, 45, 5–32, 2001, Kluwer Academic Publishers

[14] T. Haifley, "Linear logistic regression: an introduction," IEEE International Integrated Reliability Workshop Final Report, 2002., Lake Tahoe, CA, USA, 2002, pp. 184-187.

[15] Yujun Yang, Jianping Li and Yimei Yang, "The research of the fast SVM classifier method," 2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, 2015, pp. 121-124.

[16] J. K. Jaiswal and R. Samikannu, "Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression," 2017 World Congress on Computing and Communication Technologies (WCCCT), Tiruchirappalli, 2017, pp. 65-68.