Integrated Physiological Signal-based Biomarkers for Automatic Stress Detection

***Abstract*—Numerous mental intentions and life rates are responsible for dispensing mental stress. It's an essential purpose behind delivering numerous cardiovascular ailments. Identifying and addressing the effect of stress on the creation of automatic identification of different levels of mental stress thus provides a crucial path for progressive research to tackle stress. We investigated mental stress identification with the help of preparing the Electrocardiogram (ECG), Galvanic Skin Response (GSR) chronicles using Hjorth parameters, Autoregressive, Shannon entropy, and a few other features that were used in finding best features using Wrapper and Boruta-Shap function. The primary reason for this experiment was to evoke particular affective states in the participants.ECG, GSR recordings of 40 people while watching short videos was used from CLAS Dataset. Random Forest Classifier, Logistic Regression, and K-Nearest Neighbor Classifier (KNN) are all designed for stress detection and have obtained accuracy rates of 82.23%, 79.84%, and 78.48% respectively. These results can be used to make a device capable of identifying and measuring the stress levels experienced by individuals, so that stress can be better managed as short-term or long-term stress still poses a risk of harm to physiological and mental health.**

***Index Terms*—Electrocardiogram (ECG); Galvanic Skin Response (GSR); Stress detection; Wrapper; Boruta-Shap; Random Forest;**

I. INTRODUCTION

Stress is the human body's natural reaction to an external damaging force. Typical physiological reactions include, among other things, changes in skin temperature, pulse, pupil dilation, heart rate and electrodermal activity. [1][2][3]. Although modest amounts of stress can be helpful to the body, stress also has a detrimental impact on attention, memory, and decision-making[4][5]. Long-term stress has been related to a number of serious health effects such as anxiety, depression, and premature aging.

Physiological stress is a frequent illness that affects people in their later years of life. Stress is a factor that creates mental distress, whether physical, mental, or emotional, and is the body's response to a threat or demand caused by external factors such as interpersonal problems, pollution, physical environment, and internal factors that take into account lack of sleep, depression, expectations, and a vital factor for human health to take into account. It may influence habits and factors that raise the risk of heart disease: high blood pressure and cholesterol levels, smoking, physical inactivity, and overreaction. These psychological stress types can be listed as acute and chronic [6]. Acute stress is a type of short-term stress that does not necessarily endure long. This sort of stress is always necessary for our performance index. While chronic stress can cause numerous ongoing changes in our physiological parameters such as blood pressure, heart rate, body temperature, and so on[7]. Because stress may lead to lasting sickness, it is critical to identify it early on while also avoiding additional difficulties that may emerge as a result of physiological stress.

Chronic stress, on the other hand, is harmful to both physical and mental health[8]. As a result, a number of research have been carried out to assess the influence of acute stress on the development of automatic detection of various levels of mental stress. Han[9] suggested a stress monitoring system based on personal physiological signals such as the electrocardiogram (ECG), photoplethysmogram (PPG), and galvanic skin reaction to offer objective everyday healthcare (GSR). Their program measures tension with a precision of 81.82 percent in daily settings. Kalinkov, et al,[10] a standard experimental protocol was implemented through the evaluation that uses the MAHNOB-HCI data collection. Markova, et al,[11] present a restricted method of selection of attributes that use the Fisher separation characteristic assessment criterion followed by post-processing varietal reduction. The suggested method was demonstrated in an experimental setup for acute stress detection based on physiological signals. Markova, et al,[12] [13],their experimental evaluation said that the Benefits of Person-Independent Tags modeling Audio-video stimuli and person-built models specific tags auto-reported. Based on the tags themselves reported, they report that those are obtained with only a small extra effort relatively improved accuracy of HANV detection with up to 5%. They record test findings for various configurations of the binary detectors of negative emotions, a high degree of emotional anticipation, and negative-valence high-arousal states. Markova, et al,[14] present a three-step attribute selection process that builds on stages of assessment of person-independent and person-specific features. Their experimental setup focused on physiological signals was held to test the suggested approach on the ASCERTAIN database, adapted to high-arousal negative-valence detection. Nath, et al,[15] used two physiological signals Galvanic Skin Response (GSR) and Photoplethysmograph (PPG) to validate a stress identification model using cortisol as a stress biomarker. GSR and PPG signals were acquired from a total of 13 individuals during the trial, coupled with saliva samples taken at various time periods, with an overall accuracy of 92 percent. Zhang, et al,[16] used Electrocardiogram (ECG) signals to analyze the autonomic real-scene stress reactivity. Weak and stressful data sets were classified using vector supporting machine (SVM). While previous findings show that acute stress contributes to excessive motivating and parasympathetic stimulation deactivation[17][18], heart rate rises and falls; level fluctuation; the cumulative influence of long-term real-world non-verbal stress-inducing autonomic system events yet to be uncovered. Markova, et al,[19] present an implementation of the CLAS data set, a multimodal resource that was purposefully developed for research support and technology development (RTD) activities targeted at automated recognition of a certain particular state of mind.

Currently there are no specific way to detect or measure stress in real life. Mainly questionnaires are used for this purpose which can be flawed due to many memory biases and many other reasons. But by using physiological signals we can solve the problem and to classify these signals and extract features from them, machine learning is the popular way.Recently, machine learning has been used for many kinds of classification or regression problems. As per the literature survey for stress detection there, a lot of machine learning techniques were used for the detection of stress.This is the reason machine learning is being used in this paper for the detection of stress.

In this paper, our objective is the identification of mental stress with the guide of preparing the Electrocardiogram (ECG), Galvanic Skin Response (GSR) chronicles utilizing Hjorth Parameters, Autoregressive, Shannon entropy and few other features that were used in finding best features using Wrapper and Boruta-Shap function. After extracting features from ECG and GSR signals, we process our data for training our model. Our dataset was full of features as we merged two physiological signals. By using the wrapper method and Boruta-Shap algo, we decreased the dimension of features. Because, the Wrapper method determines the relationship between variables and finds a suitable subset of functions for the desired machine learning algorithm. Furthermore, Boruta-Shap can give not only a superior subset of features, but also the most accurate and consistent global feature rankings, which may also be utilized for model inference. As a result, we can train our model to identify stress using the optimal subset of characteristics.

No one has utilized as many features as we have, therefore the originality of our research is that we have used a mix of features. The rest of this paper is structured as follows. The approach was stated in Sec. 2. Sec.3 presents the experimental outcomes. Finally, in section 4, we came to a conclusion.

II. Methodology

A. Dataset

We used the CLAS data set. But we didn't use all of the info. We've just been using the short video experiment. Neurophysiological signals have been documented in this database using wearable sensors that enable independence because they use wireless technology. However, we have only used the pre-processed ECG and GSR.

CLAS is a data collection on individuals and groups for effects, personality and mood analysis. The data-set consists of profiles, scores of the participants, external annotations, neuro-physiological recordings and video recording of the participants. There are anonymized participants' data, personality profiles and mood (PANAS) in the profiles, ECG and GSR signals in neuro-physiological recordings and frontal HD, full-body, depth videos of a short experiment in video recording.

The signals were recorded in this database using wearable sensors, which, owing to their use of wireless technology, allow for independence. The ECG was recorded using three electrodes, two of which were placed as a reference on the right and left arm crooks and the third on the inner face of the left ankle. This setup provides for accurate detection of both heart beats and the full QRS ECG complex. The GSR signal was captured using two electrodes placed in the middle phalanges of the fingers.

In this short video experiment, 40 volunteers watched a compilation of 16 short effective film extracts. There were emotional videos chosen to evoke particular effective states in the participants(duration¡250s). Each participant was in individual settings and evaluated each video in valence, excitement, dominance, familiarity and liking, and selected fundamental feelings. Videos of this experiment were annotated externally by 3 annotators on the scales of valence and arousal.

1) ECG: Electrocardiography uses electrodes implanted on the skin to create a voltage versus time graph of the heart's electrical activity. Small electrical changes induced by heart muscle depolarization and repolarization are detected by these electrodes throughout each cardiac cycle. The quantity of electrical activity passing through the heart can provide information about the heart's relative size, allowing doctors to determine if it is overworked or bloated. Various cardiac abnormalities, such as cardiac rhythm disruptions, inadequate coronary artery blood flow, and electrolyte disturbances, as well as other changes in the usual ECG pattern, occur. An ECG has three main components: the P wave, which represents atrial depolarization; the QRS complex, which represents ventricular epolarization; and the T wave, which represents ventricular repolarization.

2) GSR: The galvanic skin response (GSR) is a term that describes variations in sweat gland activity which reflects the intensity level of a person's emotional state[21]. The time course of the signal is governed by two additive processes: the slow-changing tonic base level driver and the faster-varying phasic component. Variations in phasic activity may be seen in the continuous data stream because these bursts have a strong tilt to an identifiable peak and a sluggish fall in comparison to the baseline level.When there are significant changes in GSR activity due to responses to stimuli, an event-related skin conductance response takes place which is referred to as a reaction. GSR peaks, or emotional arousal to stimuli, are a type of response that might reveal information about emotional arousal. GSR activity peaks that are unrelated to the presentation of a stimulus are known as Non-Stimulus-locked Skin Conductance Responses. To offer quantitative data to emotional arousal research, skin conductance levels or the number of GSR peaks can be used. It is simpler to make fresh discoveries and get new insights into human nature when there is more data. Increased perspiration produces an increase in skin activity in a stressful situation because the sweat gland reacts to the SNS.

B. Classifier

In order to predict stress, we have used a range of different classifiers in our research. In particular, we have taken into consideration and used K-Nearest Neighbour(KNN), Logistic Regression(LR) and Random Forest(RF) to generate the predictions on stress.

1) Random Forest: The Decision Tree is a well-perceived and decoded computation, and so a single tree may not be adequate for the model to extract the highlights from it. Random Forest, on the other hand, is a ”Tree”-based computation that employs the features highlights of many Decision Trees for deciding[22][23][24].

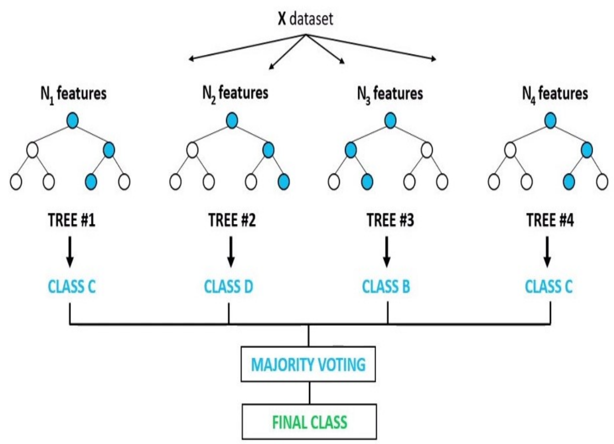

Fig. 1: Random Forest Classifier.

2) Logistic Regression: Logistic regression is one of the basic and common algorithms used to solve the problem of classification. Moreover, the logistic function, also called a sigmoid function. The sigmoid function is a numerical capacity used to map the anticipated values to probabilities[25][26]. It takes values between 0 and 1 be that as it may, never precisely at those cutoff points.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

In the event that we need to order if an email is a spam or not, on the off chance that we apply a Linear Regression model, we would get just consistent qualities somewhere in the range of 0 and 1, for example, 0.4, 0.7 and so forth. Then again, the Logistic Regression expands this straight relapse model by setting an edge at 0.5, subsequently the information point will be named spam if the yield esteem is more prominent than 0.5 and not spam if the yield esteem is lesser than 0.5. Along these lines, we can utilize Logistic Regression to characterization issues and get precise predictions.

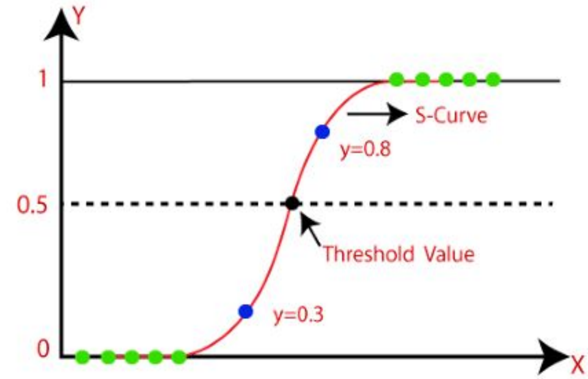

Fig. 2: Logistic Regression.

3) K-Nearest Neighbor: KNN Classifiers are instance-based classifiers that are widely utilized in medical applications. The KNN method uses a majority vote to give the mark to training examples that are closest to the feature space[27][28][29]. KNN captures the potential of comparability (also known as separation, nearness, or proximity) with some science we may have studied as children—determining the spacing between focus on a diagram.

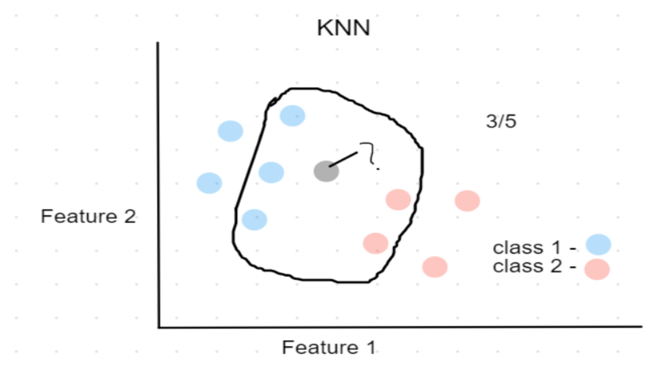

Fig. 3: K-Nearest Neighbor Classifier.

C. Feature Extraction

Highlight extraction is identified with dimensional decrease. During the time spent element extraction, it is required to contain applicable data from the information with the goal that the ideal errand can be performed by utilizing this decreased portrayal rather than the total data. At that point we utilized underlying strategies for Matlab and different Python libraries to extricate highlights. which are mentioned below: 1.mean heart rate (Statistical Feature, ECG) 2. mean RR (Statistical Feature, ECG) 3. Mean (Statistical Feature, ECG) 4. max NN interval (Statistical Feature, ECG) 5. min NN interval (Statistical Feature, ECG) 6. pNN50 (Statistical Feature, ECG) 7. SDNN (Statistical Feature, ECG) 8. RMSSD (Statistical Feature, ECG) 9. Standard deviation of the difference of successive NN intervals (Statistical Feature, ECG) 10. SD1(Short term variability) (Statistical Feature, ECG) 11. SD2(Long term variability) (Statistical Feature, ECG) 12. Power in (0-0.04) Hz (Frequency Domain Features,ECG) 13. Power in (0.04- 0.15) Hz (Frequency Domain Features, ECG) 14. Power in(0,15-0.4) Hz bands (Frequency Domain Features, ECG) 15. normalized powers in the three bands (Frequency Domain Features, ECG) 16. the power in the three bands in percent (Frequency Domain Features, ECG) 17. HRV (Frequency Domain Features) (Frequency Domain Features, ECG) 17. number of peaks (Peak's Statistical Features, GSR) 18. max and min amplitude of the peaks (Peak's Statistical Features, GSR) 19. mean conductance of the peaks (Peak's Statistical Features, GSR) 20. RMS, standard deviation and mean absolute value of the peaks (Peak's Statistical Features, GSR) 21. skewness and kurtosis of the peak's distribution (Peak's Statistical Features, GSR) 22. Mean resistance(Signal's Statistical Features, GSR) 23. First quartile (Signal's Statistical Features, GSR) 24. Second quartile (Signal's Statistical Features, GSR) 25. Third quartile (Signal's Statistical Features, GSR) 26. Interquartile range (Signal's Statistical Features, GSR) 27. Percentile 2.5 (Signal's Statistical Features, GSR) 28. Percentile 10 (Signal's Statistical Features, GSR) 29. Percentile 90 (Signal's Statistical Features, GSR) 30. Percentile 97.5 (Signal's Statistical Features, GSR) 31. Hjorth parameters 32.Autoregressive model 33. Shanon entropy 34. Wavelet

1) Wrapper Method: Wrapper techniques evaluate a subset of data by employing a machine learning algorithm that utilizes a search strategy to traverse the space of alternative subsets of features, evaluating each subset based on the quality of the output of a specific algorithm. The Wrapper method specifies the relationship between variables and identifies the best subset of functions for the intended machine learning algorithm. Wrapper approaches often outperform filter techniques in terms of prediction accuracy.

2) Boruta-Shap: Boruta-Shap is a wrapper feature selection technique that incorporates the Boruta feature selection process as well as the shapley values. This combination demonstrated that the original Permutation Importance technique outperforms in terms of both speed and consistency of the produced function subset. This function not only provides a superior subset of features, but it also provides the most reliable and consistent global feature rankings, which may also be utilized for model inference.

By wrapper method we get the best feature subset for logistic regression and by BarutaShap we get the best feature subset for random forest classifier as well as for k nearest neighbour(KNN). With the best feature subset we train our model to classify stress detection. Both of these techniques interact with the classifier.

III. Experiment and Result

Wrapper and Boruta-Shap have been used for dimension reduction from all features set. 'SEFeatures(5,12,17,23,47)', 'lfnu', 'WVARfeature(2,3)','HRfeature1 are the subset of best features that have been achieved through Wrapper. Shannon Entropy's features 5th and 12th value of the 1st block, 1st and 7th value of 2nd block, 15th value of the 3rd block, from 11 wavelet of variance 2nd and 3rd coefficient, low frequency normalization are the elements of the subset of the best features. Additionally, 'SD2', 'vlf percent', 'lf percent', 'WVARfeature[1,9]','SDNN' are the best feature subset accomplished by Boruta-Shap. From Wavelet of variance 1st and 9th value, percentage of very low frequency and low frequency, standard deviation of NN interval have been included in the subset of the best features. Using these features, we have got 82.27% highest accuracy through random forest classifier in Boruta-Shap and 79.84% highest accuracy through Logistic Regression in Wrapper.

TABLE I: Stress detection accuracy by different machine learning classifiers

	Random Forest classifier (Accuracy)	**Logistic Regression (Accuracy)**	**KNN (Accuracy)**
Boruta-Shap (parameters:'SD2','vlf percent','lf percent', 'WVARfeature[1,9]', 'SDNN')	82.27%	79.84 %	74.68 %
Wrapper (parameters: 'SEFeatures [5,12,17,23,47]' ,'lfnu', 'WVARfeature [2,3]','HRfeature1)	73.41%	79.74%	78.48 %

The most intuitive measure of success is accuracy and it is simply a proportion of correctly predicted observations to total observations. Moreover, The train accuracy is the precision of a model on models it was built on. The test accuracy is the precision of a model on models it hasn't seen. In the following table, we have shown training and testing accuracy. In Boruta-Shap, the training accuracy of Random Forest is 100% the highest, and also the testing accuracy of Random Forest is 82.27% the highest one. However, in Wrapper, the testing accuracy of Logistic Regression is 79.74% the highest one. On the other hand, KNN has lagged behind in training and testing both of the cases.

TABLE II: Training vs Testing accuracy of models.

	Random Forest Classifier		Logistic Regression		KNN	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
Boruta Shap	100	82.27	71.33	79.84	79.45	78.48
Wrapper	100	75.94	69.75	79.74	82.61	74.68

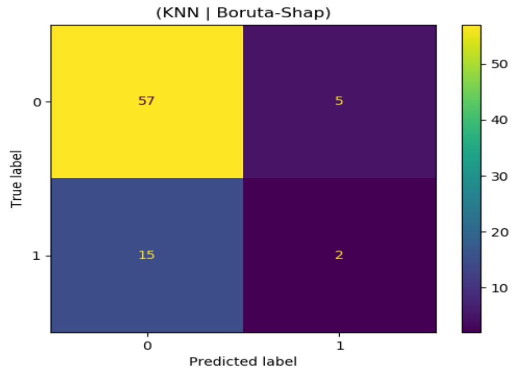

Fig. 4: Confusion matrix of KNN based on feature subset of Boruta-Shap.

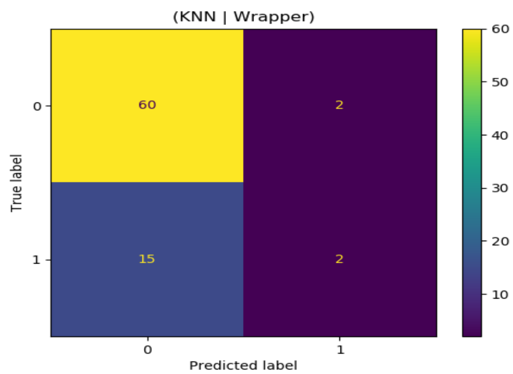

Fig. 5: Confusion matrix of KNN based on feature subset of Wrapper

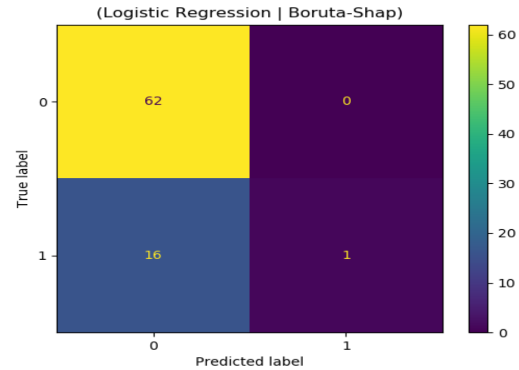

Fig. 6: Confusion matrix of logistic regression based on feature subset of Boruta-Shap.

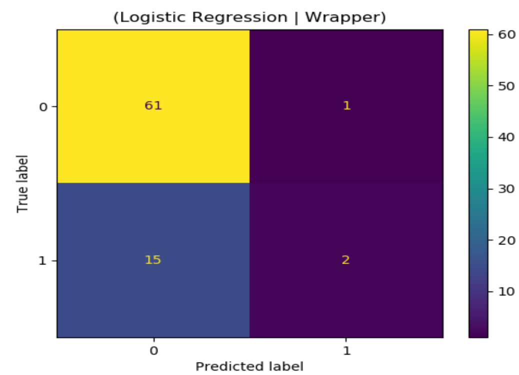

Fig. 7: Confusion matrix of Logistic Regression based on feature subset of Wrapper.

Clearly, we can infer that the Random Forest Classifier is a good model by watching all performance based scores as well as its accuracy. Our Proposed model best classifier had an accuracy 82.27% in detecting stress.

IV. Conclusion

In this paper, we investigated mental stress identifiers by utilizing Hjorth Parameters, Autoregressive, Shannon entropy and features which are extracted from ECG and GSR signals. We have used ECG and GSR recordings from CLAS dataset, while watching some emotional video clips by the participants. By implementing essential models, we were successful to identify stress from the dataset. We only used machine learning for the stress detection . The results of our classification indicate that our method and analysis offer a helpful identification of stress. In the future, other researchers may be able to use a deep learning approach and eliminate the manual feature extraction difficulty since deep learning can be utilized for automatic feature extraction.

References

[1] P Renaud and J.-P. Blondin, The stress ofstroop performance: physiological and emotionalresponses to color-word interference, task pacing,and pacing speed,Int. J. Psychophysiol.27(1997)87–97.

[2] C. Sandi, Stress and cognition,Wiley Interdiscip.Rev. Cogn. Sci.4(2013) 245–261.

[3] R.S. Stawski, M.J. Sliwinski and J.M. Smyth,Stress-related cognitive interference predicts cogni-tive function in old age,Psychol. Aging21(3) (2006)535–544

[4] J.Taelman, S. Vandeput, A. Spaepen andS. Van Huffel, Influence of mental stress on heartrate and heart rate variability, in4th European Conference of the International Federation for Medicaland Biological Engineering, eds. J. Vander Sloten,P. Verdonck, M. Nyssen and J. Haueisen, Vol.22 (Springer Berlin Heidelberg, 2009), pp. 1366–1369.

[5] M. Wikgren, M. Maripuu, T. Karlsson, K. Nordfj,J. Bergdahl, J. Hultdin, J. Del-Favero, G. Roos,L. Nilsson, R. Adolfsson and K. Norrback, Shorttelomeres in depression and the general populationare associated with a hypocortisolemic state,Biol.Psychiatry71(4) (2012) 294–230.

[6] R. Picard, Automating the recognition of stress and emotion: From lab to real-world impact", 2016. doi: 10.1109/MMUL.2016.38. 19

[7] Hjortskov, N., et al.: The effect of mental stress on heart rate variability and blood pressure during computer work. Eur. J. Appl. Physiol. 92(1–2), 84–89 (2004).

[8] Kim, P.(2017). MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence, 2017.

[9] HAN and H. JEONG, Objective stress monitoring based on wearable sensors in everyday settings", 2019. [Online]. Available: https://escholarship.org/uc/ item/16v1t1ft.

[10] K. Kalinkov, Markova V., and T. Ganchev,Front-end processing of physiological signals for the automated detection of high-arousal negative valence conditions", National Conference with Intern Participation (ELECTRONICA- 2019), 2019. doi: 10.1109/ELECTRONICA.2019.8825647. 20

[11] Markova,V. and Ganchev, T.(2018). Constrained Attribute Selection for

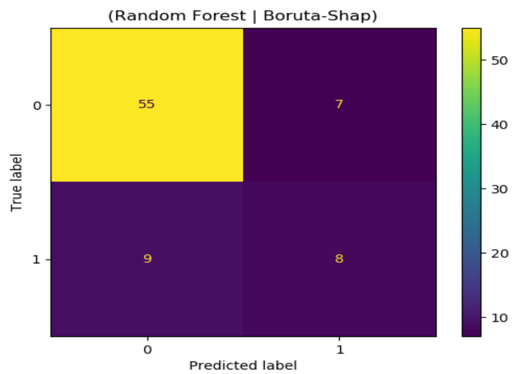

Fig. 8: Confusion matrix of Random Forest Classifier based on feature subset of Boruta-Shap.

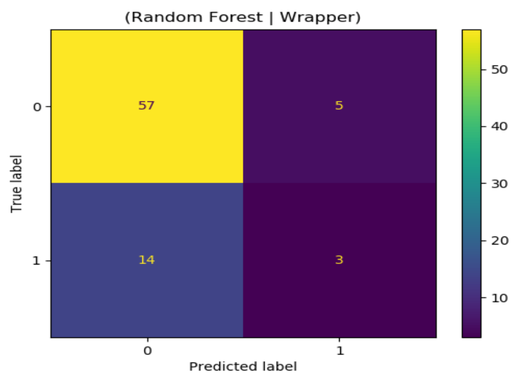

Fig. 9: Confusion matrix of Random Forest Classifier on feature subset of Wrapper.

Stress Detection Based on Physiological SignalsProceedings of the 2018 International Conference on Sensors, Signal and Image Processing,2018. doi: 10.1145/3290589.3290597

[12] V. Markova v. and T. Ganchev, Automated recognition of aect and stressevoked by audio-visual stimuli", in Seventh Balkan Conference on Lighting (BalkanLight), 2018. doi: 10.1109/BalkanLight.2018.8546887.

[13] T. Markova V.and Ganchev and K. Kalinkov, Detection of negative emotions and high-arousal negative-valence states on the move", 2018. doi: 10.1109/RTUWO.2018.85817888

[14] Markova,V. and Ganchev, T. Three-step attribute selection for stress detection based on physiological signals", in 2018 IEEE XXVII International Scienti_c Conference Electronics-ET, 2018. doi: 10.1109/ET.2018.8549658.

[15] K. Nath, H. Thapliyal, and A. Holt, Validating physiological stress detection model using cortisol as stress bio marker", 2020 IEEE International Conference on Consumer Electronics (ICCE), 2020. doi:10.13140/RG.2.2.24469.88807.

[16] Zhang. J. and W. W. et al, Recognition of real-scene stress in examination with heart rate features", International Conference on Intelligent Human-Machine Systems and Cybernetics, 2017. doi: 10.1109/IHMSC.2017.13.

[17] Sriramprakash S., Prasanna V.D., Murthy O.V.R,Stress detection in working people", Proc. Procedia Comput. Sci., 2017. doi: 10.1016/j.procs.2017.09.090.

[18] Zhang, J., Zong, C.(2015). Deep neural networks in machine translation: an overview. IEEE Intelligent System,2015. doi: 10.1109/MIS.2015.69

[19] Markova,V., Ganchev, T. and Kalinkov, K.(2019). CLAS: A Database for Cognitive Load, Affect and Stress Recognition.International Conference on Biomedical Innovations and Applications (BIA),2019. doi: 10.1109/BIA48344.2019.8967457

[20] B. Hjorth and A. Elema-Sch onander, "Eeg analysis based on time domain properties", Electroencephalography and Clinical Neurophysiology, 1970. doi: 10.1016/0013-4694(70)90143-4.

[21] B. Hjorth and A. Elema-Sch onander, Eeg analysis based on time domain properties, Electroencephalography and Clinical Neurophysiology, 1970. doi:10.1016/0013-4694(70)90143-4.

[22] Painsky A, Rosset S (2017). Cross-Validated Variable Selection in Tree-Based Methods Improves Predictive Performance. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39 (11): 2142–2153. arXiv:1512.03444. doi:10.1109/tpami.2016.2636831. PMID 28114007

[23] Davies, Alex; Ghahramani, Zoubin (2014). "The Random Forest Kernel and other kernels for big data from random partitions". arXiv:1402.4293

[24] Lin, Yi; Jeon, Yongho (2002). Random forests and adaptive nearest neighbors (Technical report). Technical Report No. 1055. University of Wisconsin. CiteSeerX 10.1.1.153.9168.

[25] Hyafil, Laurent; Rivest, RL (1976). "Constructing Optimal Binary Decision Trees is NP-complete". Information Processing Letters. 5 (1): 15–17. doi:10.1016/0020-0190(76)90095-8

[26] Engel, J. (1988). "Polytomous logistic regression". Statistica Neerlandica. 42 (4): 233–252. doi:10.1111/j.1467-9574.1988.tb01238.x.

[27] Belsley, David (1991). Conditioning diagnostics : collinearity and weak data in regression. New York: Wiley. ISBN 9780471528890.

[28] Kyurkchiev, Nikolay, and Svetoslav Markov. "Sigmoid functions: some approximation and modelling aspects". LAP LAMBERT Academic Publishing, Saarbrucken (2015).

[29] M. A. Chikh, M. Saidi and N. Settouti, Diagnosis of diabetes diseases using an artificial immune recogni-tion system2 (airs2) with fuzzy k-nearest neighbor,J. Med. Syst.36(5) (2012) 2721–2729