**Heart Rate Variability-Based Mental Stress Detection: An Explainable Machine Learning Approach**

Jyoti Sekhar Banerjee, Mufti Mahmud & David Brown

Abstract

Stress may be identified by examining changes in everyone's physiological reactions. Due to its usefulness and non-intrusive appearance, wearable devices have gained popularity in recent years. Sensors provide the possibility of continuous and real-time data gathering, which is useful for tracking one's own stress levels. Numerous studies have shown that emotional stress has an impact on heart rate variability (HRV). Through the collection of multimodal information from the wearable sensor, our framework is able to accurately classify HRV based users' stress levels using explainable machine learning (XML). Sometimes, ML algorithms are referred to as black boxes. XML is a model of ML that is designed to explain its objectives, decision-making, and reasoning to end users. End users may include users, data scientists, regulatory bodies, domain experts, executive board members, and managers who utilize machine learning with or without understanding or anybody whose choices are impacted by an ML model. The purpose of this work is to construct an XML-enabled, uniquely adaptable system for detecting stress in individuals. The results show promising qualitative and quantifiable visual representations that may provide the physician with more detailed knowledge from the outcomes offered by the learnt XAI models, hence improving their comprehension and decision making.

Introduction

Stress is an intensified physio-psychological condition of the human body that develops in reaction to a demanding situation or difficult occurrence. Stressors are the external elements that cause stress. A person's physical and mental health may be seriously impacted by long-term exposure to multiple stressors having an impact simultaneously, which may further result in serious illnesses varying from minor symptoms like trouble sleeping, weight loss or gain, headaches, to severe ones like heart attack [1].

In the past, physicians or researchers have relied on interviews and questionnaires to get information on people's emotional states. This strategy, however, may be invasive due to the fact that it disrupts on-going work or because the interviewer may have their own agenda in mind. Detection of emotional states and stress has been studied in a number of different ways, including through the observation of physiological reactions [2,3,4,5,6,7,8], via the use of audio-visual data [9, 10], or writing [11].

By analyzing physiological changes in the body, stress may be diagnosed. SNS, i.e., the sympathetic nervous system, initiates physiological reactions in response to stress by producing adrenaline and cortisol. These hormones may raise muscular tension, heart rate, sweat, and breathing. Then, these symptoms may be utilized to identify stress in an individual.

The majority of stress detection methods [3, 5, 7] employ generalized models. Although, each individual experiences physiological changes in a unique way. Although they provide a broad and individualized approach, Shi et al. [6] and Smets, et al. [12] only divide stress into two categories. From various machine learning [13] classifiers used in earlier papers, Random Forest (RF) [5, 14, 15], SVM [5, 16, 17], and decision trees [15,16,17,18] were found to be the most effective among all due to their better results as compared to others in stress detection [19,20,21,22].

To clarify how machine learning (ML) models create predictions, this research employed explainable machine learning (XML) [23,24,25], a novel branch of ML. Humans are able to comprehend machine learning (ML) [26, 27] in the form of explainable machine learning (XML). The explanation for its choices and behaviours is provided by XML. It helps people understand how machines arrive at their conclusions. The rising presence of ML in our daily lives has raised serious concerns among the general public, particularly in the medical field [28,29,30,31,32]. If we are going to put our faith in these institutions, we must be able to believe in them.

The contribution of the work is as follows:

1.

The purpose of this research is to propose and develop XML techniques for the categorization of stress levels based on multimodal data.

2.

The outcomes show promise in terms of both qualitative and quantitative visualizations, which may provide the clinician with more detailed information to aid in their comprehension and decision making of the outcomes provided by the intelligent XML models.

3.

As far as the authors are aware, no previous research article has used an explainable machine learning technique to identify mental stress.

Following is the continuation of this paper: the reason for utilizing explainable ML and why it is superior to conventional ML and HRV datasets for stress detection are discussed in “Research Methodology”. The experimental outcomes of our classification method are then discussed in “Results and Discussion”. “Comparative Study of Several Mental Stress Detection Systems” presents the comparative study and finally, “Conclusion” concludes our investigation.

## Research Methodology

This section discusses the reason behind using explainable ML and why it is better than traditional ML and HRV datasets for stress detection. XML is depicted as an ML system that provides an explanation of the assumptions made during the prediction. Simply, Explainable machine learning refers to ML that people can comprehend. XML enhances the accountability, fairness, trustworthiness, and transparency of the ML techniques. The main driving force behind XML is that more consumers are beginning to doubt the conclusions given by ML. Before relying on the forecasts and making judgments, they want to know how they were created. Figure 1 demonstrates the concept of the traditional and explainable machine learning approach. Our proposed framework also follows the same explainable ML approach that utilizes the WESAD dataset and SHAP interpretability model to detect mental stress.

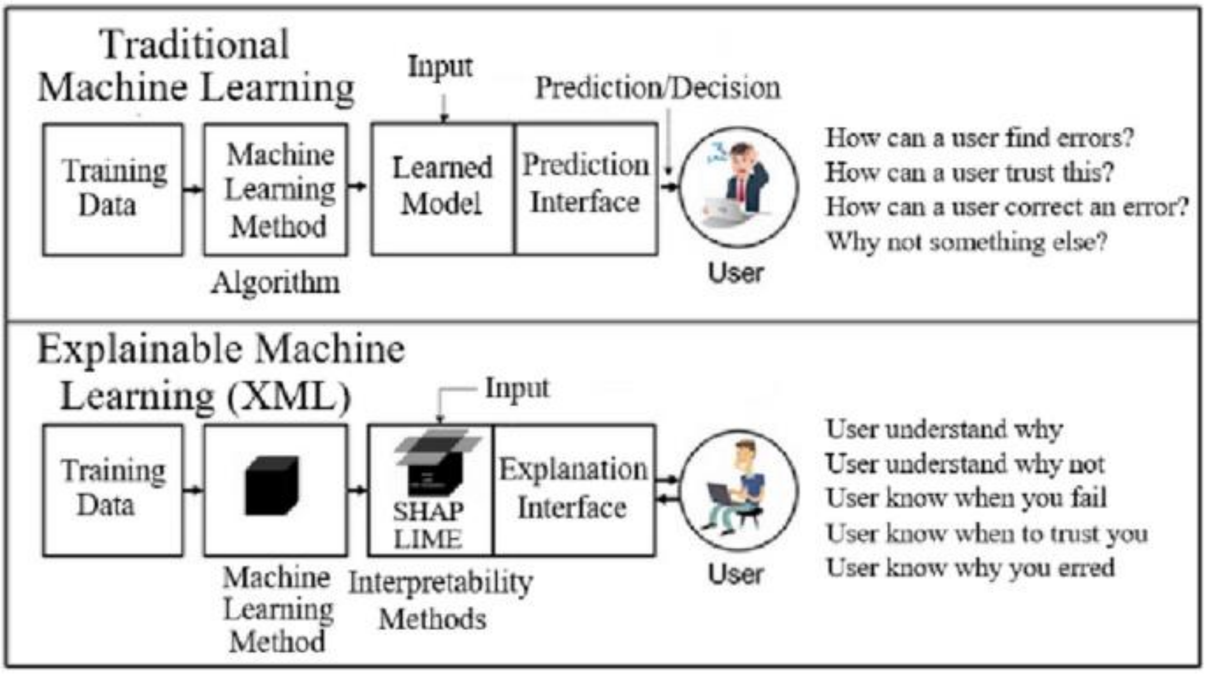



figure 1: The concept of the traditional and explainable machine learning approach

The two physiological signs that are most often employed to evaluate stress levels are heart rate variability (HRV) and electrodermal activity (EDA). We solely evaluated heart rate variability (HRV) for stress detection using XML in this study since important research contributions [1, 3, 11, 12] give justification for using HRV as a psychological stress indicator.

Dataset Used

The dataset complied for this investigation is WESAD. Attila Reiss and Philip Schmidt, et al. first presented and made this dataset accessible to public in the year 2018 [3]. This multimodal dataset assembles mobility information and physiological characteristics from 15 people using the wrist sensors (EDA data) and chest-worn (HRV data) sensors Empatica E4 and RespiBAN Professional, respectively. The chosen HRV characteristics for this research are RMSSD, SDSD, SDRR_RMSSD, pNNx, SD1, SD2, RELATIVE_ RR, VLF, LF, HF and LF/HF [33, 34].

Correlated features have been discarded and remaining features been considered for deriving the output (mental stress). The resulting dataset has been summarized in the following Table 1.

| | Signal | # Of sample | # Of features | # Of classes |
|---|---|---|---|---|
| Original WESAD | HRV | 158,920 | 40 | 3 |

| | Signal | # Of sample | # Of features | # Of classes |
|---|---|---|---|---|
| Considered WESAD | | 108,520 | 20 | 3 |

Table 1 Summary of the down sampled dataset

Results and Discussion

Simulation findings have been provided here for the purpose of empirical results. Various development frameworks have been used for simulations to detect stress, such as Python and Orange. Like MATLAB, Orange is also professional software and specially designed for Machine Learning applications.

Performance Measures of Different ML Models

Experiments based on the WESAD dataset were undertaken utilizing four multiclass classification algorithms: K-nearest neighbor (KNN), decision tree, random forest, and support-vector machines (SVM) (see Fig. 2). The effectiveness metrics for the four classification methods are shown in Table 2. Those are the results received from Orange.

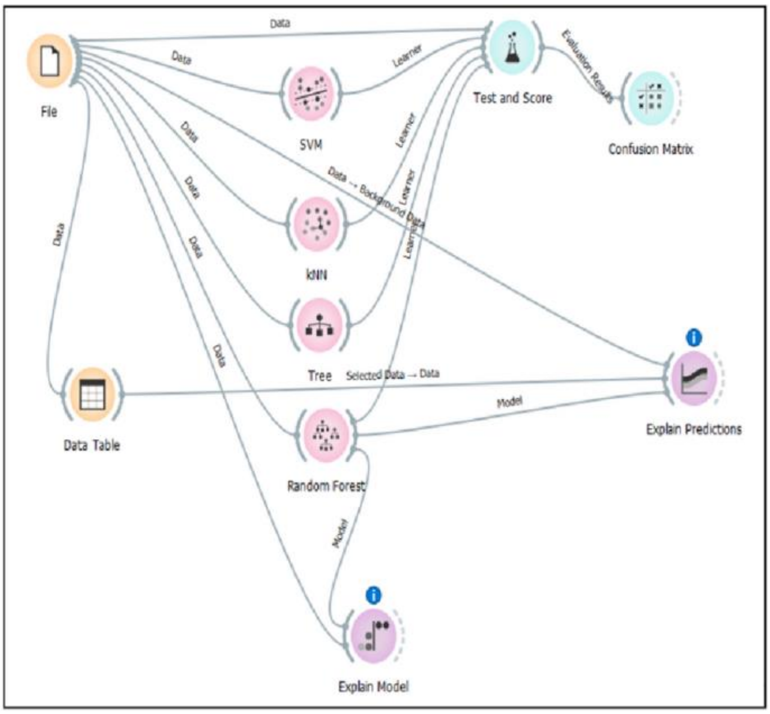



figure 2: Implementation of ML algorithms along with explainable blocks using Orange

Table 2 Comparison of efficiency measures of the different ML Algorithms

| Model | AUC | CA | F1 | Precision | Recall |
|---|---|---|---|---|---|
| kNN | 3.426 | 0.400 | 0.399 | 0.405 | 0.400 |

| Model | AUC | CA | F1 | Precision | Recall |
|---|---|---|---|---|---|
| Tree | 5.949 | 0.985 | 0.985 | 0.985 | 0.985 |
| SVM | 3.688 | 0.427 | 0.417 | 0.461 | 0.427 |
| Random Forest | 5.974 | 1.000 | 1.000 | 1.000 | 1.000 |

Using Table 2, we can observe that the Random Forest classifier outperforms all other ML models in terms of efficiency.

Figure 3 shows the confusion matrix for the prediction results using Random Forest, SVM, KNN, and decision tree. Orange and Weka are the tech frameworks used to simulate these.

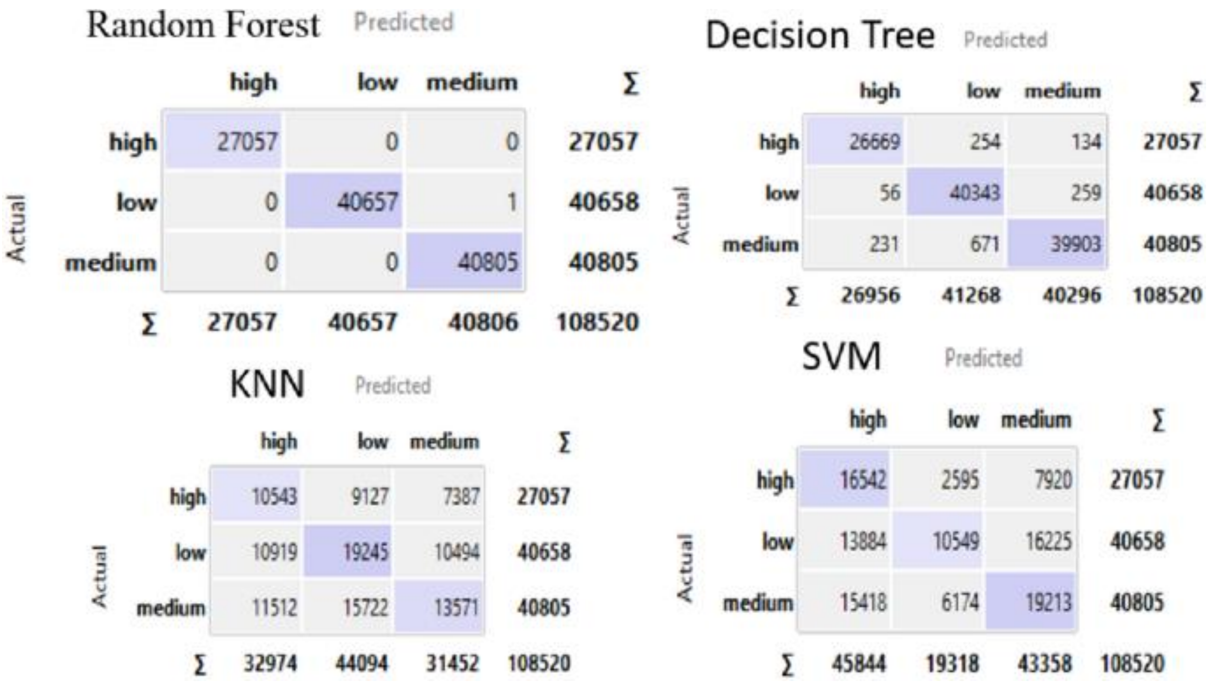
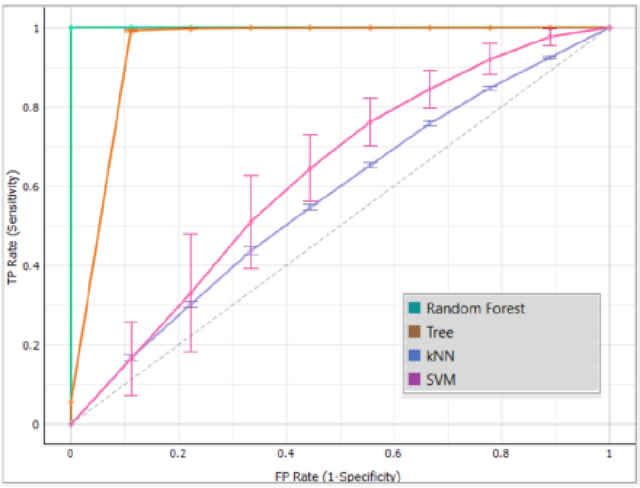
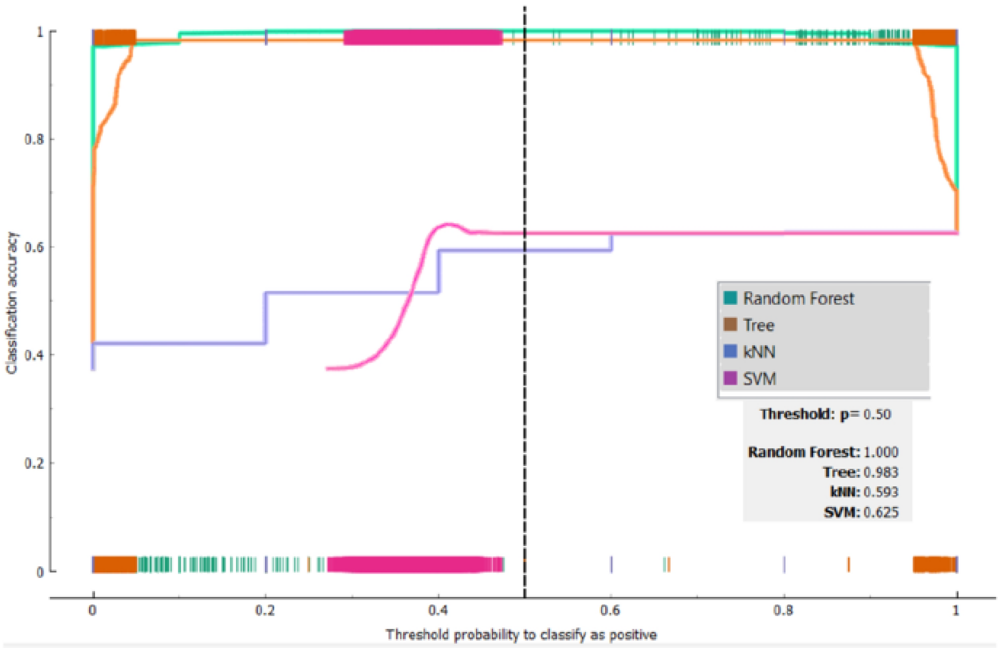



figure 3: Confusion matrix of the different ML algorithms

Receiver operational characteristic (ROC) analysis, which is the graphical tool for evaluating the output of a classifier, of four ML algorithms, is shown in Fig. 4a and the Classification accuracy plot, i.e., the proportion of correctly classified observations is shown in Fig. 4b.




**(a)**




**(b)**

figure 4: a ROC analysis using RF, SVM, KNN and Tree (DT) algorithms. b Classification accuracy plot of the different ML algorithms

XML Based Model Predictions

SHAP Global Explainability

The computational complexity of random forest, SVM, AdaBoost, and XGBoost models makes it challenging for the common users to grasp how predictions are formed. A game theoretical post-hoc interpretation approach known as Kernel SHAP was used to increase model explainability for these types of models. For example, SHAP gives both a global explanation of the model's structure (global explainability) as well as a particular prediction (specific explainability or local explainability).

SHAP Feature Importance

Identifying the significance of the features is the primary objective of a model (see Figs. 5 and 6). Characteristics having high absolute Shapely values are considered as the most important. Importance or significance of the features is basically the mean as well as standard deviation of impurity reduction accumulated inside each tree.

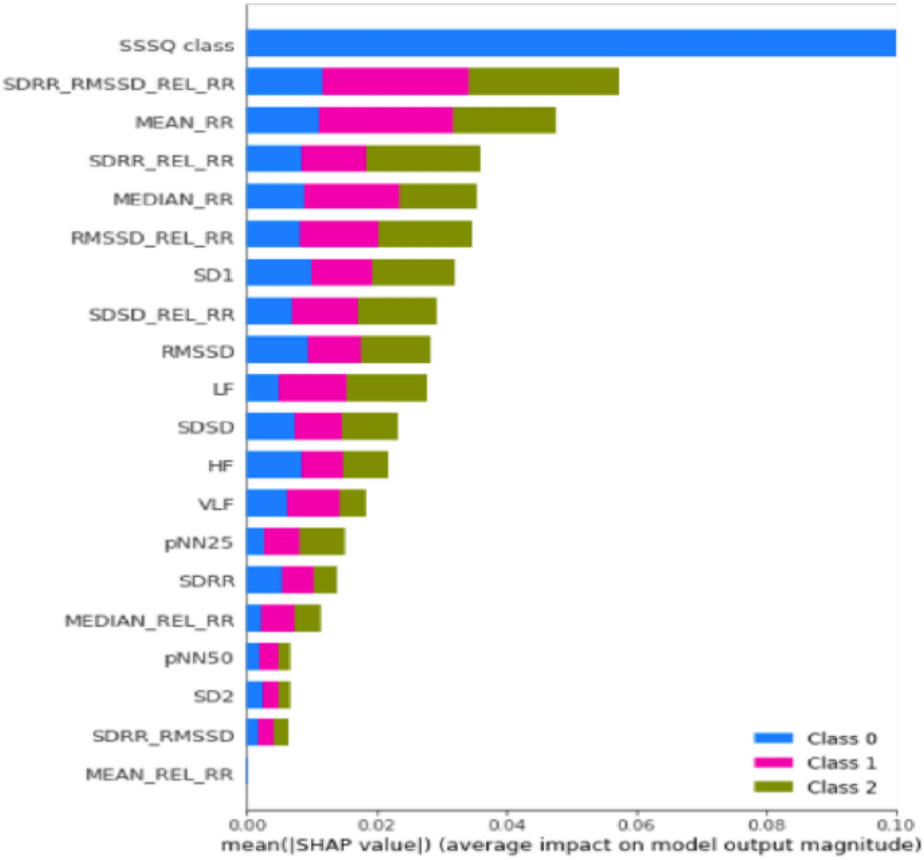



figure 5: Average impact on model output magnitude (Mean SHAP value)

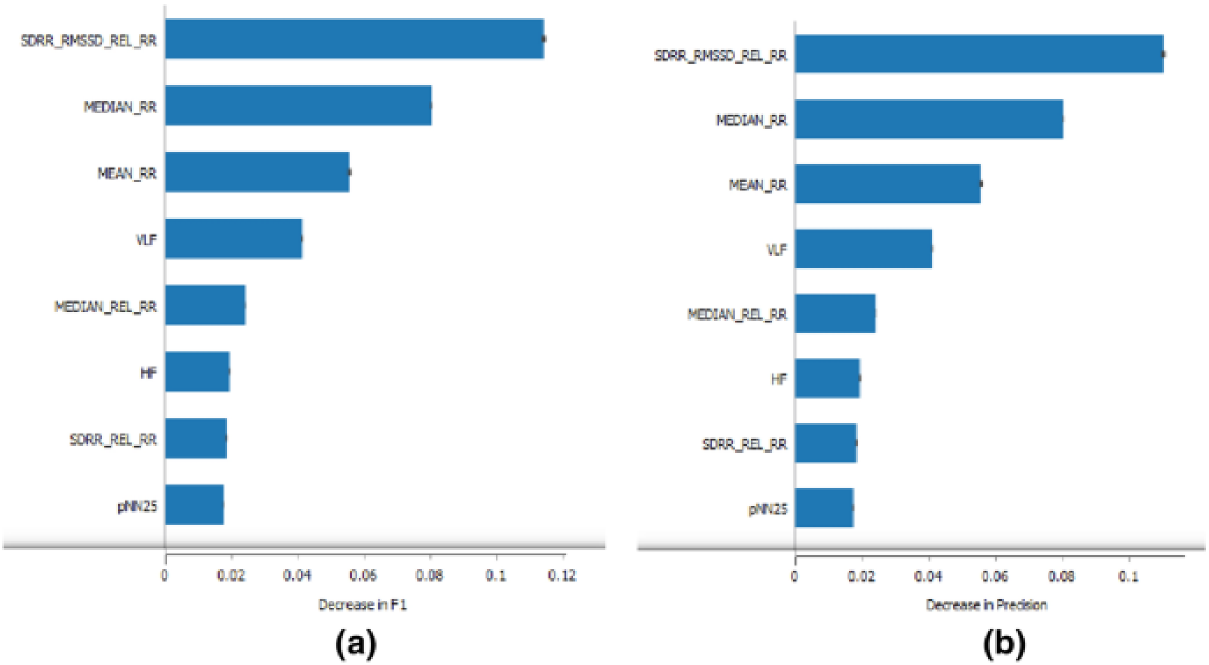



Figure 6: a, b Feature Importance on model output magnitude

SHAP Summary Plot

The model's feature significance and feature impacts are combined to create the summary plot. A Shapely value that corresponds to one instance per feature is shown by each point over the figure. The feature and X axis values that correspond to each instance are used to determine the location on the Y axis. From the Fig. 7, it is clear that SDRR_RMSSD_REL_RR feature has the highest range of Shapely value and hence can be considered as the most significant parameter. Color is used to represent the feature's value, which ranges from Low to High. We can see how the distribution of the Shapley values for each feature is distributed since the points which got overlapped are jittered in the direction of Y-axis. The attributes are listed according to their importance.

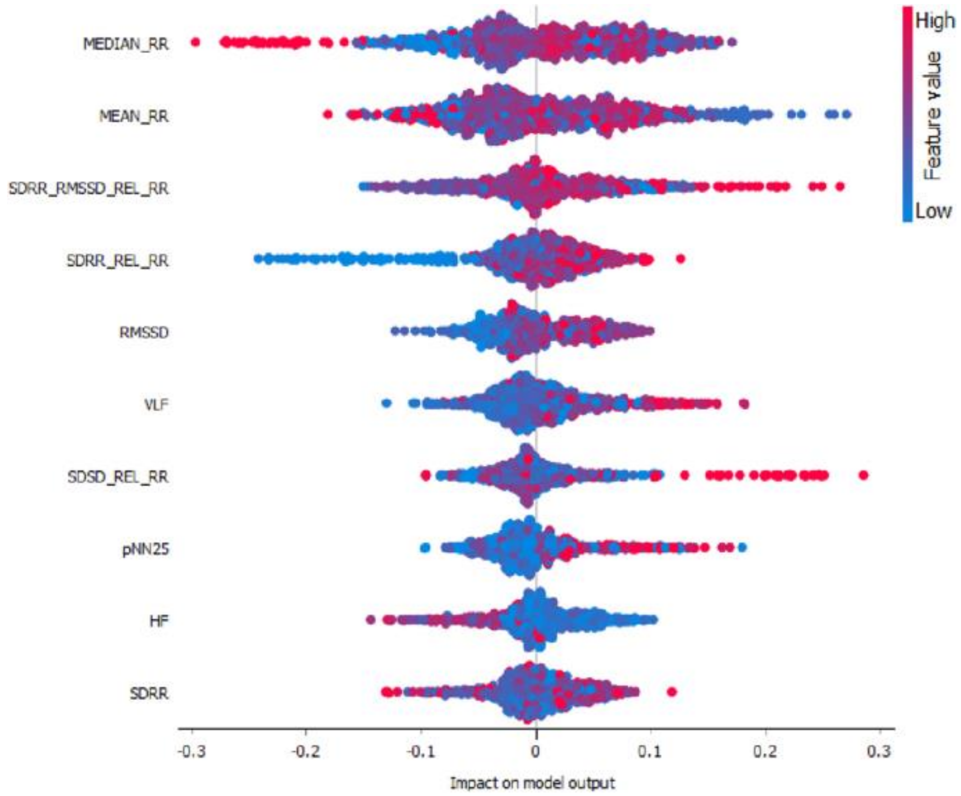



Figure 7: Summary Plot

SHAP Dependence Plot

The partial dependence plot basically reflects the minimal dependence of only a few parameters on the prediction based outcome of any ML model. This plot can also exhibit the nature (like linear, monotonic or even complex) of the relationship between a feature and the target. This method is a global approach, which considers all the instances, comes to a conclusion on the global relationship between a feature and the prediction-based result. A dependency plot (see Fig. 8) is nothing more than a scatter plot that shows how one characteristic affects the predictions the model produces. Every dot marks a unique prediction (row wise) out from dataset. A dataset's real value is shown along the X-axis.

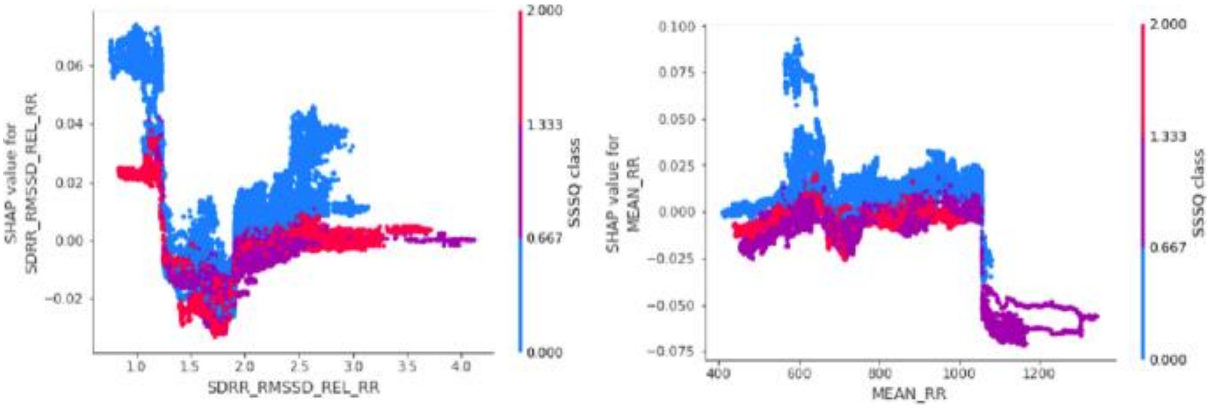



figure 8: SHAP partial dependance plot

SHAP Force Plot

This graph (SHAP Force Plot) shows us how easily one model prediction can be explained (see Figs. 9 and 10). Further, this Force plot is a way for inaccuracy assessment and explanation of a particular case prediction; i.e., both the cases where the prediction is accurate and occasions where it is inaccurate can be analyzed. This also provides a notion of which features are driving to the incorrect forecast. From the graph the average value or base value is obtained, and each feature’s influence towards deriving the prediction is shown as well. Features plotted in red lead to the prediction of greater than average (Base) value, most likely default value. On the other side, features plotted in blue color, lead to the prediction of smaller than average (Base).

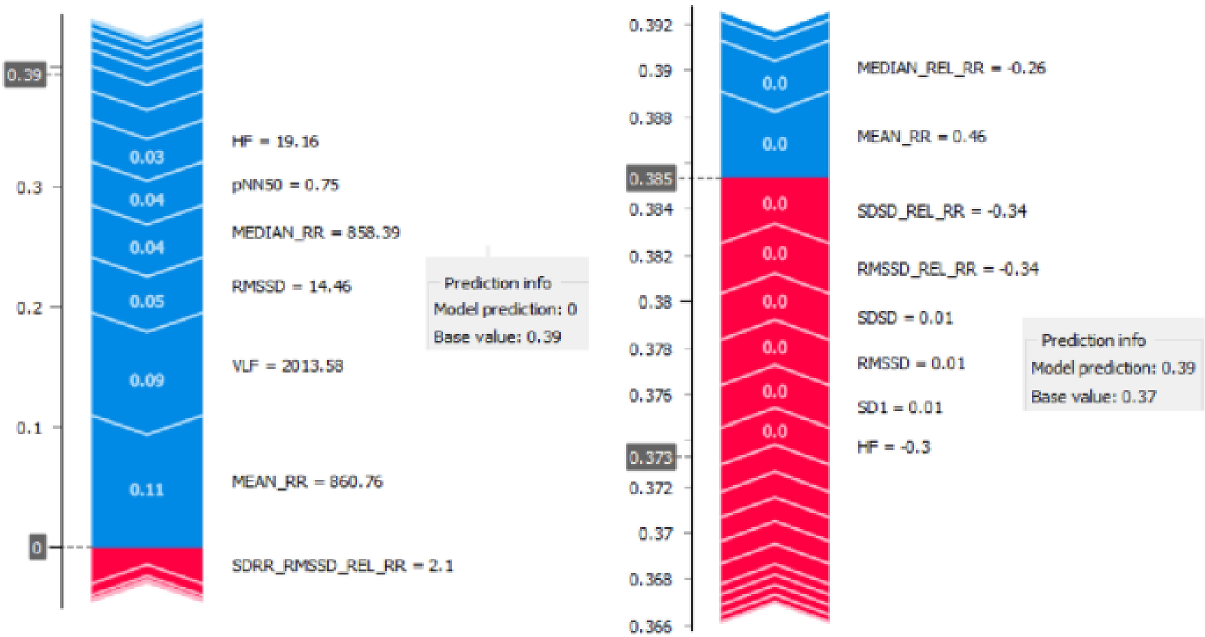



Figure 9: Prediction using RF algorithm (left) and prediction using decision tree algorithm (right)

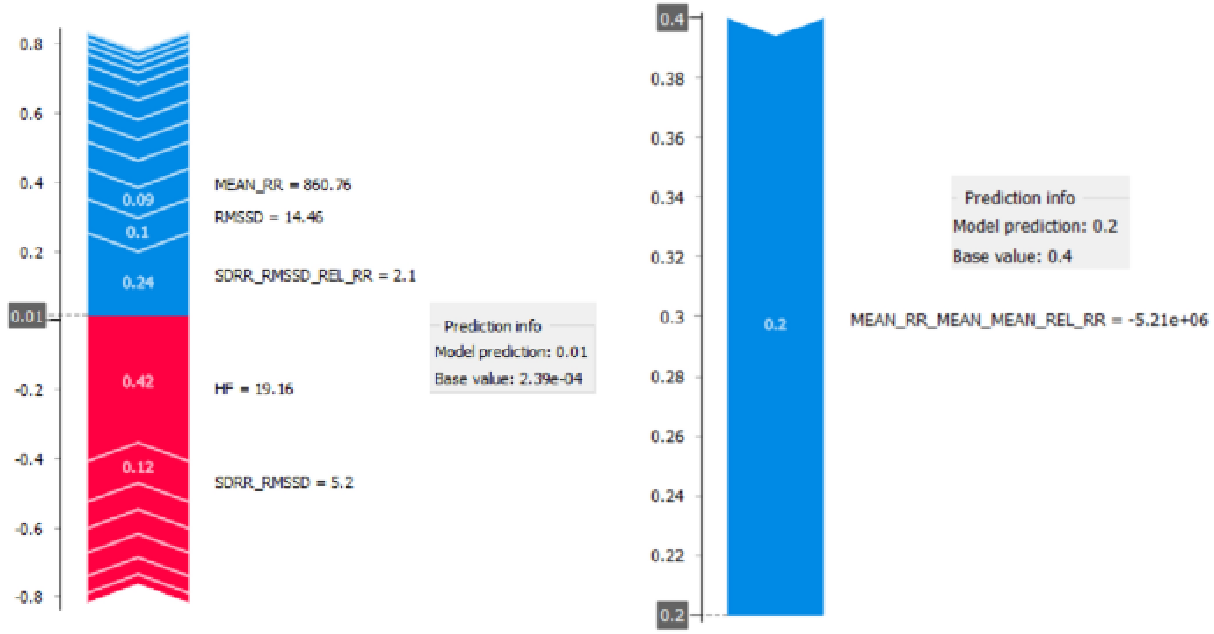



Figure 10: Prediction using SVM algorithm (left) and prediction using KNN algorithm (right)

SHAP Decision Plot

Effectively the same output information is depicted through the Decision Plot as the Force Plot. Here the model's average (base) value is marked through the grey vertical line. As usual, a red colored line indicates whether the output value has drifted higher or lower relative to the prediction's average value. Whenever, many features are to be considered for analysis purpose, Decision Plot plays important role. And, in case of more number of predictors, the information appears to be very condensed in the Force plot. For reference, feature values are presented beside prediction line. The prediction line illustrates how the SHAP values add up from the base value at the bottom of the plot to the ultimate score of the model at the top. These decision plots are easy for interpretation. Placing decision plots together can aid in the identification of outliers depending on their SHAP scores (see Fig. 11).

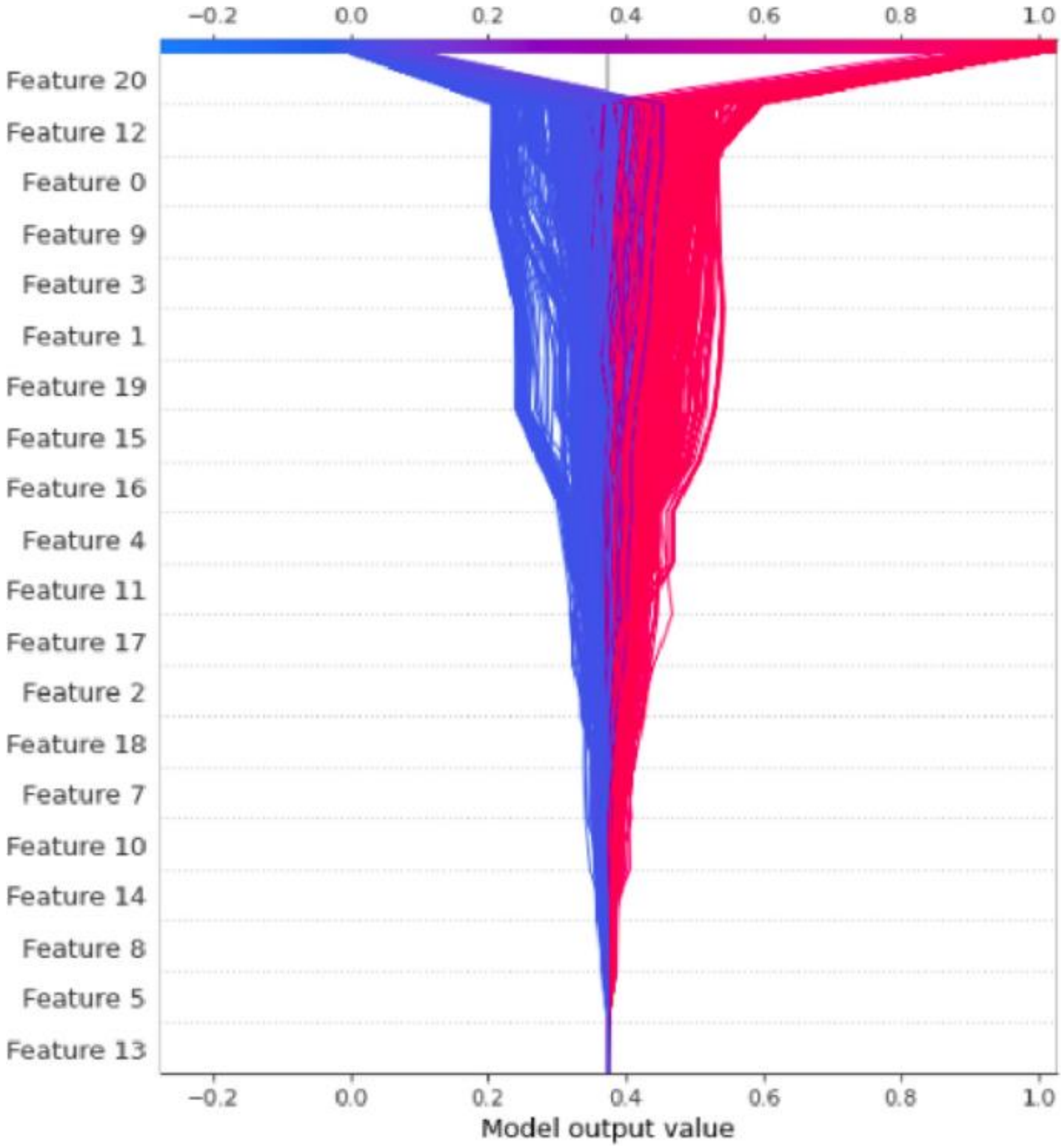



Figure 11: SHAP decision plot

SHAP Results Summary

In the field of machine learning model interpretation, the SHAP framework has shown to be a significant development. SHAP combines a number of current approaches to provide an approach that is both theoretically solid and understandable for explaining predictions for any model. The size and direction (positive or negative) of a feature’s impact on a prediction are quantified by SHAP values. Hence, using SHAP Machine learning models can be explained better with the importance of features.

Comparative Study of Several Mental Stress Detection Systems

In this part, the authors comprehensively compared several mental stress detection methods concerning the explainability of the systems (see Table 3). The chart makes it obvious that our approach outperforms other techniques for detecting mental stress.

Table 3 Comparison of efficiency measures of the different ML Algorithms

| Reference | Year | Applied software applications | Wearables data used | Applied ML Algorithms | Explainability Features Available |
|---|---|---|---|---|---|
| Heyat et al. [35] | 2022 | MATLAB, Anaconda | Yes | Decision Tree, Naive Bayes, Random Forest, and Logistic Regression | No |
| AlShorman et al. [36] | 2022 | MATLAB | Yes | Support Vector Machine and Naive Bayes | No |
| Can et al. [20] | 2019 | MATLAB | Yes | PCA, LDA, SVM, k-NN, Logistic Regression, Random Forest, Multilayer Perceptron | No |
| Indikawati et al. [19] | 2020 | Apache Spark | Yes | Logistic Regression, Decision Tree, and Random Forest | No |
| Proposed | 2022 | MATLAB, Orange, Weka | Yes | k-NN, Decision tree, SVM, Random Forest | Yes |

Conclusion

Explainable ML improves the ML systems to a greater extent to explain its objectives, decision-making, and reasoning to end users. In this paper, the authors used multimodal HRV data from the WESAD dataset that was collected by the wearable sensors. We have adopted a set of algorithms like SVM, KNN, Decision Tree and Random Forest in this research work. We have used several ML algorithms to divide people's stress level into three categories: Not Stressed, Medium Stressed, and Highly Stressed. The Random Forest algorithm outperforms all other models that were tested. By analyzing the system's output, the Random Forest method has achieved 83% accuracy in performing individualized stress detection. Instead of using a single parameter HRV, a combination of HRV and EDA can be applied to detect stress using XML for improving interpretability.

Data availability

The data generated and analyzed during the current study are available from the corresponding author on reasonable request.

## References


McEwen BS, Stellar E. Stress and the individual: mechanisms leading to disease. Arch Intern Med. 1993;153(18):2093–101.

Xu Q, Nwe TL, Guan C. Cluster-based analysis for personalized stress evaluation using physiological signals. IEEE J Biomed Health Inform. 2014;19(1):275–81.

Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM international conference on multimodal interaction, 2018, pp. 400–8.

Keshan N, Parimi PV, Bichindaritz I. Machine learning for stress detection from ECG signals in automobile drivers. In: 2015 IEEE International conference on big data (Big Data), IEEE, 2015, pp. 2661–9.

Gjoreski M, Luštrek M, Gams M, Gjoreski H. Monitoring stress with a wrist device using context. J Biomed Inform. 2017;73:159–70.

Shi Y, Nguyen MH, Blitz P, French B, Fisk S, De la Torre F, Kumar S. Personalized stress detection from physiological measurements. In: International Symposium on Quality of Life Technology, 2010, pp. 28–9.

Marković D, Vujičić D, Stojić D, Jovanović Ž, Pešović U, Ranđić S. Monitoring system based on iot sensor data with complex event processing and artificial neural networks for patients stress detection. In: 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), IEEE, 2019, pp. 1–6.

Chakraborty A, Banerjee JS, Bhadra R, Dutta A, Ganguly S, Das D, Kundu S, Mahmud M, Saha G. A Framework of Intelligent Mental Health Monitoring in Smart Cities and Societies. IETE Journal of Research. (2023)(Press).

Mirsamadi S, Barsoum E, Zhang C. Automatic speech emotion recognition using recurrent neural networks with local attention. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 2227–31.

Tzirakis P, Zhang J, Schuller BW. End-to-end speech emotion recognition using deep neural networks. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2018, pp. 5089–93.

Gangemi A, Presutti V, Recupero DR. Frame-based detection of opinion holders and topics: a model and a tool. IEEE Comput Intell Mag. 2014;9(1):20–30.

Smets E, Casale P, Großekathöfer U, Lamichhane B, Raedt WD, Bogaerts K, Hoof CV. Comparison of machine learning techniques for psychophysiological stress detection. In: International symposium on pervasive computing paradigms for mental health. Cham: Springer; 2015. p. 13–22.

Banerjee J, Maiti S, Chakraborty S, Dutta S, Chakraborty A, Banerjee JS. Impact of machine learning in various network security applications. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, (2019). p. 276–281.

Can YS, Chalabianloo N, Ekiz D, Fernandez-Alvarez J, Riva G, Ersoy C. Personal stress-level clustering and decision-level smoothing to enhance the performance of ambulatory stress detection with smartwatches. IEEE Access. 2020;8:38146–63.

Delmastro F, Di Martino F, Dolciotti C. Cognitive training and stress detection in mci frail older people through wearable sensors and machine learning. IEEE Access. 2020;8:65573–90.

Sano A, Picard RW. Stress recognition using wearable sensors and mobile phones. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013, pp. 671–6.

Sun FT, Kuo C, Cheng HT, Buthpitiya S, Collins P, Griss M. Activity-aware mental stress detection using physiological sensors. In: International conference on mobile computing, applications, and services. Cham: Springer; 2010. p. 282–301.

Airij AG, Sudirman R, Sheikh UU. GSM and GPS based real-time remote physiological signals monitoring and stress levels classification. In: 2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS), IEEE, 2018, pp. 130–5.

Indikawati FI, Winiarti S. Stress detection from multimodal wearable sensor data. IOP Conf Ser. 2020;771(1):012028.

Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. Sensors. 2019;19(8):1849.

Nkurikiyeyezu K, Yokokubo A, Lopez G. The effect of person-specific biometrics in improving generic stress predictive models. 2019. arXiv preprint arXiv:1910.01770.

Gedam S, Paul S. A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access. 2021;9:84045–66.

Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60.

Ye Q, Xia J, Yang G. Explainable AI for COVID-19 CT classifiers: an initial comparison study. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2021, pp. 521–6.

Wang M, Zheng K, Yang Y, Wang X. An explainable machine learning framework for intrusion detection systems. IEEE Access. 2020;8:73127–41.

Mandal JK, Misra S, Banerjee JS, Nayak S. Applications of Machine intelligence in Engineering: Proceedings of 2nd Global Conference on Artificial Intelligence and Applications (GCAIA, 2021). Jaipur, India: CRC Press; 2022. p. 8–10

Saha P, Guhathakurata S, Saha S, Chakraborty A, Banerjee JS. Application of machine learning in app-based cab booking system: a survey on Indian scenario. In Applications of Artificial Intelligence in Engineering. Singapore: Springer: 2021. p. 483–497.

Guhathakurata S, Kundu S, Chakraborty A, Banerjee JS. A novel approach to predict COVID-19 using support vector machine. In Data Science for COVID-19. Academic Press; 2021. p. 351–364.

Pandey I, Dutta HS, Banerjee JS. WBAN: a smart approach to next generation e-healthcare system. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019. p. 344–349

Guhathakurata S, Saha S, Kundu S, Chakraborty A, Banerjee JS. South Asian Countries are less fatal concerning COVID-19: a fact-finding procedure integrating machine learning & multiple criteria decision-making (MCDM) technique. Journal of The Institution of Engineers (India): Series B, 2021;102(6):1249–1263. https://doi.org/10.1007/s40031-021-00547-z

Guhathakurata S, Saha S, Kundu S, Chakraborty A, Banerjee JS. A new approach to predict COVID-19 using artificial neural networks. In Cyber-Physical Systems. Academic Press; 2022. pp. 139–160

Guhathakurata, S., Saha, S., Kundu, S., Chakraborty, A., & Banerjee, J. S.: South Asian countries are less fatal concerning COVID-19: a hybrid approach using machine learning and M-AHP. In Computational Intelligence Techniques for combating COVID-19 (pp. 1-26). Springer, Cham, (2021).

Costin R, Rotariu C, Pasarica A. Mental stress detection using heart rate variability and morphologic variability of EeG signals. In: 2012 International Conference and Exposition on Electrical and Power Engineering, IEEE, 2012, pp. 591–6).

Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig. 2018;15(3):235.

Bin Heyat MB, Akhtar F, Abbas SJ, Al-Sarem M, Alqarafi A, Stalin A, Wu K. Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal. Biosensors. 2022;12(6):427.

AlShorman O, Masadeh M, Heyat MBB, Akhtar F, Almahasneh H, Ashraf GM, Alexiou A. Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection. J Integr Neurosci. 2022;21(1):20.

Author information

Jyoti Sekhar Banerjee: Department of Computer Science and Engineering (AI & ML), Bengal Institute of Technology, Kolkata, India

Jyoti Sekhar Banerjee, Mufti Mahmud & David Brown: Department of Computer Science, Nottingham Trent University, Nottingham, UK, NG11 8NS