Stress Detection through Machine Learning using HRV: A Systematic Review

Steven Fraters
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
s.j.fraters@student.utwente.nl

ABSTRACT

Stress has a significant impact on the health of humans, thus reliable detection methods are required so we can prevent the ever increasing stressors in life from harming us. One of the current stress detection methods is looking at the biomarker Heart Rate Variability (HRV) measured by wearables. HRV is the difference between time intervals of separate heart beats, which decreases when stress is introduced to the body hence making it viable as a biomarker in stress prediction. Therefore this systematic review aims to investigate machine learning models that have been used in stress prediction with through HRV. The databases selected for this systematic review are Scopus, PubMed and IEE Xplore. Applying the criteria of articles that use HRV and machine learning in stress prediction resulted in 19 articles being included in the analysis. Analysis shows that the most commonly used machine learning algorithms are random forest, k-nearest neighbor and support vector machines. The challenges most models faced are with the personal differences in HRV changes people experience, as this makes it hard to make a general model. The most accurate models were Convolutional Neural Networks (CNN) and thus are my recommendation for an algorithm in stress prediction through HRV. Further research into CNN seems promising for people to monitor, and possibly regulate, their stress levels.

KEYWORDS

Heart Rate Variability, Stress detection algorithms, wearable technology

1 INTRODUCTION

Stress is the human body's reaction to a challenging condition, or stressors, marked by great anxiety or duress [8]. The stress response is influenced by the intensity and chronicity of stressors. Prolonged exposure to stress can cause various negative health conditions such as hypertension, diabetes, and cognitive dysfunction [17]. This harmful effect of stress on the human body can be explained through the response to stressors, which is mediated by an interplay between the brain and the cardiovascular system, among others [22]. The brain determines what is stressful and is therefore an important player in this stress response [21]. The response also establishes objective physiological changes in the cardiovascular system such as altered heart rate variability (HRV) [13]. HRV describes the variations in the time intervals between sequential heartbeats and is used as a tool to assess psychological states, such as stress, depression, and anxiety [13][11]. In the case of stress, it is often observed that the heart rate rises drastically. This causes an increase in blood pressure, which is associated with reduced HRV. Therefore, HRV is widely used as a measurable indicator to investigate the stress response [13][11][25]. Currently, the use of machine learning in stress prediction is on the rise. Machine learning can be used to identify patterns in complex datasets, therefore being useful in building predictive models for specific health conditions such as stress [28]. Several studies are investigating the effects of training ML models on biomarker data, like HRV, to eventually predict stress [33]. Research that investigates stress detection through ML using HRV seems to yield promising results and may broaden the overall knowledge. Therefore eventually helping individuals to manage stress by changing their lifestyle or behaviors to keep them from serious health conditions [21]. Hence, this review aims to gain insight into machine learning and statistical algorithms that have been described to predict stress through HRV, together with their performance.

42nd Twente Student Conference on IT, January 31, 2025, University of Twente, the Netherlands

2 PROBLEM STATEMENT

Research that has been done into stress prediction with the use of biomarkers always needs to process the data and compare them to some baseline [15]. This is currently almost exclusively done with the use of machine learning or statistical algorithms. These algorithms can vary widely and thus it is useful to create an overview of which algorithms have been used and what their advantages and disadvantages are pertaining to stress prediction with HRV. To break this overview down into smaller parts the following research questions where chosen:

- **RQ1:** What machine learning or statistical algorithms have been used to predict stress using HRV data?
- **RQ2:** What are the challenges or limitations of these algorithms in predicting stress from HRV data?
- **RQ3:** Which algorithmic approaches (e.g., deep learning, support vector machines, signal processing methods) are most suitable for predicting stress using HRV data based on the available literature?

3 METHODOLOGY

A literature review will be done on the machine learning and statistical algorithms that are employed in predicting stress through HRV. To ensure that this literature review is systematic is to ensure that it is repeatable, transparent and transferable [5]. To accomplish this the PSALSAR method for literature review will be used [23]. This method builds on the framework of Search, Appraisal, Synthesis, and Analysis (SALSA), which is a methodology to determine the search protocols that the systematic literature review should follow [10] . The PSALSAR adds Protocol and Report to this framework where in the protocol the scope of the study is defined and in the report section the writing of the report is categorized. An overview of the PSALSAR can be seen in Table 1

Firstly, a research protocol for a systematic literature review helps minimize bias by doing exhaustive literature searches [23]. The most important part of the protocol is defining the research scope which is worked out in Section 2.

Secondly, the search strategy has to be defined. This systematic review is based on searches from PubMed, Scopus, and IEEE Xplore covering literature from 2022-2024 using the search string ''heart rate variability" AND stress AND ''Machine learning". Furthermore, articles must have a title that links stress prediction and heart measurement research to be included. The syntax used for the search was TITLE-ABS-KEY. Figure 1 shows that the initial results included 232 articles, of which 176 came from Scopus, 35 from PubMed and 21 from IEEE Xplore. On the PubMed and IEEE Xplore search the parameter of full text available was selected to limit the records not being able to be retrieved later on. These 232 records became 198 records at the moment of screening due to 34 duplicates being present in the search as can be seen in Figure 1.

Thirdly, the selected papers had to be appraised and filtered to fit into the scope of this review.This appraisal was done on title of the article and as the scope of the review is limited when the title induced doubt the record was not sought for retrieval. This left 52 papers which were deemed eligible based on their title. As shown in Figure 1 20 of these 52 records were not available and were therefore not considered in this review.

Fourthly, the synthesis of the selected records consisted of retrieving the models that were used to predict stress with HRV. This was done by finding the methods used to create the models in the records and putting into Google Sheets which record built which model. Next to the model it's accuracy, precision, recall and F1 score were also sought and put next to the model, as well as the challenges or shortcomings of the model.

Fifthly, the models, their performance and challenges were analyzed in order to determine the answer to the research questions. To be able to do this Table 3 was created. Next to that, the challenges of the models were written out in Section 4.

Lastly, the report phase consisted of putting together this review containing the methodology and the results.

4 RESULTS

After analyzing the articles selected in section 3, the algorithms that were described in the articles were summarized in Table 3 together with the accuracy, precision, recall and F1 score of the models. All

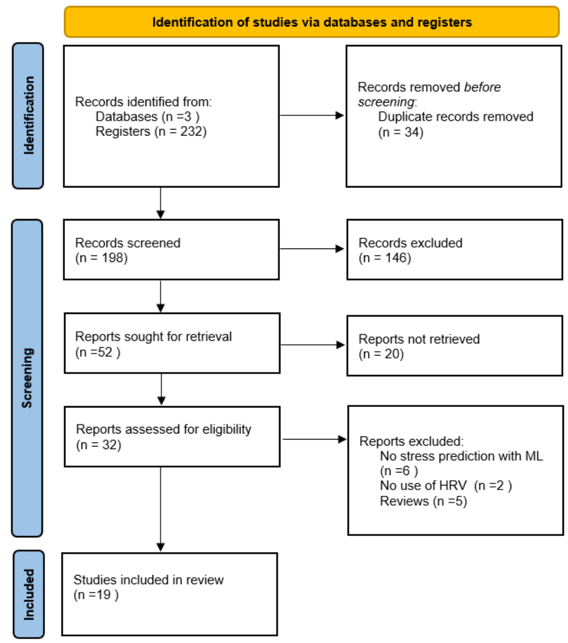

Figure 1: Flow diagram for the database search, adapted from Page et al. [26]

these scores are in a range from 0 to 1, if an article did not contain these performance metrics for their models the values were left blank. First, Mortensen et al. [24] developed a Convolutional Neural Network (CNN) that had to distinguish stress between no stress, interruption stress and time pressure stress. Notably, the performance of this model is outstanding, however a drawback of their CNN is that the version that uses all the features requires high computational power. The higher accuracy coming with a more complex model is also noticed in Hijry et al. [12] article where Extreme Gradient Boosting (XGBoost) achieved a higher performance over the more traditional models. Similarly to Mortensen et al. [24] Shikha et al. [30] used 3 classes whereas Hijry et al. [12] used a binary classification for stress. A lower performance in Cao et al. [6] compared to the previous models can be contributed to 4 out of 33 participants did not display difference in HRV and cortisol when stressed, thus making it harder for the models to accurately predict the level of stress present in the participants. Besides testing on the same dataset that the model was trained on Bahameish et al. [2] also decided to test their model on the Wearable Stress and Affect Detection (WESAD) [29] and SWELL-KW [19] datasets. Velmovitsky et al. [31] had low performance metrics, but this lower performance is due to the use of Apple Watches as the measurement devices, a shorter than standard Electro Cardiogram (ECG) measurement and a real life setting. Naegelin et al. [25] decided to approach the stress prediction from a smart office perspective where they also used data from the mouse and keyboard to aid the model. Interestingly, the model scored lower when HRV data was also used. Iovino et al. [14] investigated the difference between Linear Discriminant Analysis (LDA), Random Forest (RF), Support

Table 1: The framework PSALSAR used for systematic literature review

Steps	Outcomes	Methods
Protocol	Defined study scope	Only stress prediction with ML and HRV
Search	Define the search strategy Search studies	Searching strings Search databases
Appraisal	Selecting studies Quality assessment of studies	Defining inclusion and exclusion criteria Quality criteria
Synthesis	Extract data Categorize the data	Extraction template Categorize the data on the iterative definition and ready it for further analysis work
Analysis	Data analysis Result and discussion	Quantitative categories, description, and narrative analysis of the organized data Based on the analysis,show the state of the art and it's challenges
Report	Conclusion Report writing	Deriving conclusion and recommendation PRISMA methodology

Source: Modified from Mengist et al.[23]

Vector Machine (SVM) and k-Nearest Neighbor (KNN) models using automatic feature selection or physiologically based feature selection. The models all score similarly when using the automatic feature selection as shown in Table 3, but when the models were evaluated with the physiologically based features the SVM and KNN performed similarly to the automatic feature selection, whereas the RF and LDA decreased in performance. Although Liu et al. [20] developed an Adaptive Boosting (AdaBoost), RF, SVM and KNN with the performance stated in Table 3, the main focus was to look at ultra short term HRV features. With the results from standard HRV features the SVM emerged as the best performing and when they applied the ultra short term features with the SVM the performance only dropped 3 percentage points. Similarly to Bahameish et al. [2] Benchekroun et al. [4] focused on making there models generalizable by training them on another dataset then the testing dataset, because of this focus the performance is lower. Jeong et al. [16] focused their attention to patients that suffered from lung cancer, as this participant focus is different from the other articles the results are hard to compare with the rest of the results. With the aim of having the usually complex Graph Convolutional Network (GCN) model require less computational power Adarsh et al. [1] used pruning and quantization. The result is a high accuracy and a model that is 60-70 percent less complex. Unlike Adarsh et al. [1] Dobrokhvalov et al. [7] made the CNN models personal aiming for the highest accuracy. While PrabuShankar et al. [27] achieved really good results with traditional machine learning approaches, the dataset and if the classification is binary are not is not provided. Junqueira et al. ran two models where one was with HRV, skin temperature and blood pressure and the second model used only HRV and skin temperature. The second model performed better and is the one shown in Table 3. In contrast with all the previously mentioned models, Zawad et al. [34] did not only test the models separately, but rather combined where traditional machine learning methods were combined with an Artificial Neural Network (ANN). The combined algorithms all scored better in the performance category and the ANN in combination with Naive Bayes (NB) performed the best being able to provide a high accuracy within 0.80s. Gerasimova-Meigal et al. [9] encountered an issue regarding the collection of the perceived stress as 70 percent of their findings fell in 2-3 out of a range from 1 to 10. Banerjee et al. [3] encountered large differences in the effectivity of their models where their RF and DT outperformed the others significantly. Lastly, Vivisha et al. [32] did state one of their goals to be comparing the accuracy of multiple machine learning classification models, but focus more on integrating the Internet of Things (IoT) side. The part focus explains the lower achieved accuracies then some of the other models present in Table 3.

Table 2: Algortihm Abbreviations

Abbreviation	Meaning
CNN	Convolutional Neural Network
XGBoost	Extreme Gradient Boosting
LR	Logistic Regression
DT	Decision Trees
RF	Random Forest
KNN	K-Nearest Neighbor
SVM	Support Vector Machine
GB	Gradient Boosting
AdaBoost	Adaptive Boosting
NB	Naïve Bayes
LGBM	Light Gradient Boosting Machine
LDA	Linear Discriminant Analysis
LSTM	Long Short Term Memory
GCN	Graph Convolutional Network
MLP	MultiLayer Perceptron
NN	Neural Network
ANN	Artificial Neural Network
SVR	Support Vector Regression

Table 3: Algorithms used with their performance

Articles	Algorithms	Performance
Mortensen et al. [24]	CNN	acc = 0.99, prec = 1, recall = 1, F1 = 1
Hijry et al. [12]	XGBoost	acc = 0.997, prec = 0.997, recall = 0.997 , F1 = 0.997
	LR	acc = 0.642, prec = 0.64, recall = 0.67 , F1 = 0.56
	DT	acc = 0.708, prec = 0.70, recall = 0.70 , F1 = 0.64
	RF	acc =0.784 , prec = 0.78, recall = 0.71, F1 = 0.63
	KNN	acc = 0.922, prec = 0.92, recall = 0.92, F1 = 0.92
	SVM	acc = 0.734, prec = 0.73, recall = 0.71, F1 = 0.65
Shikha et al. [30]	XGBoost	acc = 0.922, prec = NA, recall = NA, F1 = 0.916
	GB	acc = 0.957, prec = NA, recall = NA, F1 = 0.957
	DT	acc =0.905 , prec = NA, recall = NA, F1 = 0.909
	RF	acc =0.934 , prec = NA, recall = NA, F1 = 0.924
	KNN	acc =0.846 , prec =NA , recall = NA, F1 = 0.856
	SVM	acc = 0.845, prec = NA, recall = NA, F1 = 0.832
Cao et al. [6]	XGBoost	acc = 0.722, prec = , recall = , F1 =
	AdaBoost	acc = 0.731 , prec = , recall = , F1 =
	RF	acc = 0.721, prec = , recall = , F1 =
	KNN	acc = 0.695, prec = , recall = , F1 =
	SVM	acc = 0.731, prec = , recall = , F1 =
Bahameish et al. [2]	LR	acc = 0.852, prec = 0.897, recall = 0.901, F1 = 0.872
	DT	acc = 0.859, prec = 0.905, recall = 0.895, F1 = 0.871
	KNN	acc = 0.816, prec = 0.852, recall = 0.875, F1 = 0.840
	NB	acc = 0.793, prec = 0.802, recall = 0.941, F1 = 0.844
	RF	acc = 0.855, prec = 0.848, recall = 0.967, F1 =0.892
	SVM	acc = 0.806, prec = 0.835, recall = 0.895, F1 = 0.806
Velmovitsky et al. [31]	RF	acc = 0.55, prec =0.55 , recall = 0.55, F1 = 0.55
	SVM	acc = 0.54, prec = 0.54, recall = 0.54, F1 = 0.54
Naegelin et al. [25]	SVM	acc = , prec = 0.528, recall = 0.545, F1 = 0.534
	RF	acc = , prec = 0.595, recall = 0.594, F1 = 0.589
	LightGBM	acc = , prec = 0.608, recall = 0.624, F1 = 0.612
Iovino et al.[14]	LDA	acc = 0.790, prec =0.691 , recall = 0.685, F1 = 0.685
	SVM	acc = 0.794, prec = 0.695, recall = 0.690, F1 = 0.692
	RF	acc = 0.799, prec = 0.699, recall = 0.685, F1 = 0.699
	KNN	acc = 0.788, prec = 0.686, recall = 0.682, F1 = 0.788
Liu et al. [20]	SVM	acc = 0.875, prec = , recall = 0.860, F1 = 0.897
	RF	acc = 0.810, prec = , recall = 0.830, F1 = 0.842
	KNN	acc = 0.850, prec = , recall = 0.893, F1 = 0.877
	Adaboost	acc = 0.827, prec = , recall = 0.862, F1 = 0.827
Benchekroun et al. [4]	LR	acc = , prec = 0.61, recall = 0.59, F1 = 0.57
	RF	acc = , prec = 0.63, recall = 0.62, F1 = 0.63
Jeong et al. [16]	LSTM	acc = 0.669, prec = 0.487, recall = 0.665, F1 = 0.539
	DT	acc = 0.638, prec = 0.683, recall = 0.683, F1 = 0.639
	RF	acc = 0.646, prec = 0.645, recall = 0.654, F1 = 0.625
	SVM	acc = 0.599, prec = 0.671, recall = 0.593, F1 = 0.606
	transformer	acc = 0.628, prec = 0.390, recall = 0.528, F1 = 0.412
Adarsh et al. [1]	GCN WESAD	acc = 0.978, prec = 0.944, recall = 0.962, F1 = 0.977
	GCN SWELL	acc = 0.945, prec = 0.935, recall = 0.922, F1 = 0.944
Dobrokhvalov et al. [7]	CNN	acc = 0.989, prec = , recall = , F1 =

Continuation of Table 3

Articles	Algorithms	Performance
PrabuShankar et al. [27]	LR	acc = 0.910, prec = 0.903, recall = 0.891, F1 = 0.897
	KNN	acc = 0.997, prec = 0.997, recall = 0.998, F1 = 0.997
	RF	acc = 0.959, prec = 0.946, recall = 0.956, F1 = 0.951
	DT	acc = 0.972, prec = 0.962, recall = 0.974, F1 = 0.968
Junqueira et al. [18]	NN	acc = 0.83, prec = 0.89, recall = 0.85, F1 = 0.900
Zawad et al. [34]	LR	acc = 0.637, prec = 0.632, recall = 0.637, F1 = 0.613
	NB	acc = 0.510, prec = 0.647, recall = 0.510, F1 = 0.454
	SVM	acc = 0.778, prec = 0.853, recall = 0.778, F1 = 0.752
	DT	acc = 0.884, prec = 0.888, recall = 0.884, F1 = 0.888
	RF	acc = 0.809, prec = 0.842, recall = 0.809, F1 = 0.797
	XGB	acc = 0.896, prec = 0.899, recall = 0.896, F1 = 0.895
	ANN + LR	acc = 0.955, prec = 0.955, recall = 0.955, F1 = 0.955
	ANN + NB	acc = 0.958, prec = 0.958, recall = 0.958, F1 = 0.958
	ANN + SVM	acc = 0.957, prec = 0.957, recall = 0.957, F1 = 0.957
	ANN + DT	acc = 0.952, prec = 0.952, recall = 0.952, F1 = 0.952
	ANN + RF	acc = 0.957, prec = 0.957, recall = 0.957, F1 = 0.957
	ANN + XGB	acc = 0.957, prec = 0.957, recall = 0.957, F1 = 0.957
Gerasimova-Meigal et al. [9]	RF	acc = 0.863
	CatBoost	acc = 0.853
	XGB	acc = 0.801
	LGBSM	acc = 0.829
	SVR	acc = 0.641
Banerjee et al. [3]	KNN	acc = 0.400 , prec = 0.405, recall = 0.400, F1 = 0.399
	RF	acc = 1 , prec = 1, recall = 1, F1 = 1
	SVM	acc = 0.427 , prec = 0.461, recall = 0.427, F1 = 0.417
	DT	acc = 0.985 , prec = 0.985, recall = 0.985, F1 = 0.985
Vividha et al. [32]	RF	acc = 0.735
	SVM	acc = 0.622
	KNN	acc = 0.730

5 DISCUSSION

This review has looked into the use of machine learning algorithms that have been used to predict stress with HRV features. Overall this amounted to 19 articles that were analyzed for the machine learning models and their performance. Among these 19 articles the RF, SVM and KNN were the most commonly used models with the RF being used 15 times, the SVM 12 times and lastly the KNN was used 10 times. Within machine learning there are 2 ways to provide data for the model. Firstly, there are supervised machine learning algorithms which need labeled data as input to be able to be trained. To obtain labeled data one should either collect and label their own dataset or use existing public datasets as input. Two very common public datasets that were used are the WESAD and SWELL-KW datasets[19, 29].Besides supervised algorithms there are also unsupervised algorithms, however these are used sparsely for stress prediction through HRV in the articles considered in this review.

An issue that all these models experience with HRV as a biomarker to predict stress is that differences in HRV are not equal between different people that are experiencing the same amount of stress or are having the same stressors applied to them [6]. Next to that, the dataset has to be labeled for the amount of stress the participants experienced at the time of the measurements. Currently this is performed through questionnaires answered by the participants, which allows the possibility for bias and differences between the stress people perceive to have and the stress that is visible in their body. Additionally, the change in HRV between being stressed and neutral is smaller than the HRV changes of being stress vs relaxed, which makes it harder for machine learning models to differentiate stressed and neutral versus stressed and relaxed [2, 30]. Besides, these issues with HRV as a feature there is also an issue with correctly measuring HRV. Velmovitsky et al. [31] found that using an Apple Watch instead of sensors directly on the skin near the heart caused lower performance of their model compared to other models. There is an inverse correlation between the intrusiveness of the measurement method and the accuracy of the measurement, which can become a challenge for wide scale development as people are less likely to adopt a more intrusive system.

Based on the challenges that these models face together with their performance, an argument can be made for CNN or GCN through the lens of absolute performance. When looking at a more deployable algorithm in edge devices the hybrid approach of Zawad et al. [34] seems to fit better as the time their most optimal model takes it just 0.8s.

Future research into stress prediction with machine learning and HRV features should focus on obtaining larger data samples to increase generalizability. Furthermore, the direction of personalizing the model holds promising results for accuracy. However, to be able to personalize the models there should be a reduction in the computational complexity, so this is also an area of interest for future research.

With more research into CNN or GCN the accurate prediction of stress can become more deployable and therefore widespread. The goal is that in the future it would be possible for end users to determine which activities stress them out and if possible adapt their lifestyle to reduce the amount of stress they endure. So further research in this area can lead to better personal regulation of stress.

REFERENCES

[1] V. Adarsh and G. R. Gangadharan. 2024. Mental stress detection from ultra-short heart rate variability using explainable graph convolutional network with network pruning and quantisation. *Machine Learning* 113, 8 (8 2024), 5467–5494. https://doi.org/10.1007/s10994-023-06504-9

[2] Mariam Bahameish, Tony Stockman, and Jesús Requena Carrión. 2024. Strategies for Reliable Stress Recognition: A Machine Learning Approach Using Heart Rate Variability Features. *Sensors* 24, 10 (5 2024). https://doi.org/10.3390/s24103210

[3] Jyoti Sekhar Banerjee, Mufti Mahmud, and David Brown. 2023. Heart Rate Variability-Based Mental Stress Detection: An Explainable Machine Learning Approach. *SN Computer Science* 4, 2 (3 2023). https://doi.org/10.1007/s42979-022-01605-z

[4] Mouna Benchekroun, Pedro Elkind Velmovitsky, Dan Istrate, Vincent Zalc, Plinio Pelegrini Morita, and Dominique Lenne. 2023. Cross Dataset Analysis for Generalizability of HRV-Based Stress Detection Models. *Sensors* 23, 4 (2 2023). https://doi.org/10.3390/s23041807

[5] Andrew Booth, Anthea Sutton, and Diana Papaioannou. 2016. Systematic Approaches to a Successful Literature Review. *Systematic approaches to a successful literature review* (5 2016), 272–300.

[6] Rui Cao, Amir M. Rahmani, and Karen L. Lindsay. 2022. Prenatal stress assessment using heart rate variability and salivary cortisol: A machine learning-based approach. *PLoS ONE* 17, 9 September (9 2022). https://doi.org/10.1371/journal.pone.0274298

[7] Maksim O. Dobrokhvalov and Anton Yu Filatov. 2023. Personalization of convolutional neural networks within the stress detection task using heart rate variability data. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics* 23, 6 (11 2023), 1178–1186. https://doi.org/10.17586/2226-1494-2023-23-6-1178-1186

[8] Shruti Gedam and Sanchita Paul. 2021. A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. , 84045–84066 pages. https://doi.org/10.1109/ACCESS.2021.3085502

[9] Liudmila Gerasimova-Meigal, Alexander Meigal, Vyacheslav Dimitrov, Maria Gerasimova, Anna Sklyarova, Nikolai Smirnov, and Vasilii Kostyukov. 2023. Assessment of Stress Level with Help of "Smart Clothing" Sensors, Heart Rate Variability-Based Markers and Machine Learning Algorithms †. *Engineering Proceedings* 58, 1 (2023). https://doi.org/10.3390/ecsa-10-16173

[10] Maria J. Grant and Andrew Booth. 2009. A typology of reviews: An analysis of 14 review types and associated methodologies. , 91–108 pages. https://doi.org/10.1111/j.1471-1842.2009.00848.x

[11] Yeaminul Haque, Rahat Shahriar Zawad, Chowdhury Saleh Ahmed Rony, Hasan Al Banna, Tapotosh Ghosh, M. Shamim Kaiser, and Mufti Mahmud. 2024. State-of-the-Art of Stress Prediction from Heart Rate Variability Using Artificial Intelligence. , 455–481 pages. https://doi.org/10.1007/s12559-023-10200-0

[12] Hassan Hijry, Syed Meesam Raza Naqvi, Kamran Javed, Omar H. Albalawi, Richard Olawoyin, Christophe Varnier, and Noureddine Zerhouni. 2024. Real Time Worker Stress Prediction in a Smart Factory Assembly Line. *IEEE Access* 12 (2024), 116238–116249. https://doi.org/10.1109/ACCESS.2024.3446875

[13] Sarah Immanuel, Meseret N. Teferra, Mathias Baumert, and Niranjan Bidargaddi. 2023. Heart Rate Variability for Evaluating Psychological Stress Changes in Healthy Adults: A Scoping Review. , 187–202 pages. https://doi.org/10.1159/000530376

[14] Marta Iovino, Ivan Lazic, Tatjana Loncar-Turukalo, Michal Javorka, Riccardo Pernice, and Luca Faes. 2024. Comparison of automatic and physiologically-based feature selection methods for classifying physiological stress using heart rate and pulse rate variability indices. *Physiological Measurement* 45, 11 (11 2024). https://doi.org/10.1088/1361-6579/ad9234

[15] Talha Iqbal, Adnan Elahi, Pau Redon, Patricia Vazquez, William Wijns, and Atif Shahzad. 2021. A review of biophysiological and biochemical indicators of stress for connected and preventive healthcare. https://doi.org/10.3390/diagnostics11030556

[16] Sangwoon Jeong, Hongryull Pyo, Won Park, and Youngyih Han. 2024. The Prediction of Stress in Radiation Therapy: Integrating Artificial Intelligence with Biological Signals. *Cancers* 16, 11 (6 2024). https://doi.org/10.3390/cancers16111964

[17] Alvaro A. Jiménez-Ocaña, Andrés Pantoja, Mario Andrés Valderrama, and Luis Felipe Giraldo. 2023. A Systematic Review of Technology-Aided Stress Management Systems: Automatic Measurement, Detection and Control. *IEEE Access* 11 (2023), 116109–116126. https://doi.org/10.1109/ACCESS.2023.3325763

[18] Luis Junqueira and Marta Pina. 2024. Development of an Intelligent System for Detection of Chronic Stress from Biological Signal Processing. In *IFMBE Proceedings*, Vol. 99. Springer Science and Business Media Deutschland GmbH, 115–126. https://doi.org/10.1007/978-3-031-49404-8{_}12

[19] Saskia Koldijk, Maya Sappelli, Suzan Verberne, Mark A. Neerincx, and Wessel Kraaij. 2014. The Swell knowledge work dataset for stress and user modeling research. In *ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction*. Association for Computing Machinery, Inc, 291–298. https://doi.org/10.1145/2663204.2663257

[20] Kun Liu, Yubo Jiao, Congcong Du, Xiaoming Zhang, Xiaoyu Chen, Fang Xu, and Chaozhe Jiang. 2023. Driver Stress Detection Using Ultra-Short-Term HRV Analysis under Real World Driving Conditions. *Entropy* 25, 2 (2 2023). https://doi.org/10.3390/e25020194

[21] Jeongok G. Logan and Debra J. Barksdale. 2008. Allostasis and allostatic load: Expanding the discourse on stress and cardiovascular disease. *Journal of Clinical Nursing* 17, 7B (7 2008), 201–208. https://doi.org/10.1111/j.1365-2702.2008.02347.x

[22] Bruce S Mcewen, Alfred E Mirsky, and Margaret Milliken Hatch. 2007. Physiology and Neurobiology of Stress and Adaptation: Central Role of the Brain. (2007). https://doi.org/10.1152/physrev.00041.2006.-The

[23] Wondimagegn Mengist, Teshome Soromessa, and Gudina Legese. 2020. Ecosystem services research in mountainous regions: A systematic literature review on current knowledge and research gaps. https://doi.org/10.1016/j.scitotenv.2019.134581

[24] Jon Andreas Mortensen, Martin Efremov Mollov, Ayan Chatterjee, Debasish Ghose, and Frank Y. Li. 2023. Multi-Class Stress Detection Through Heart Rate Variability: A Deep Neural Network Based Study. *IEEE Access* 11 (2023), 57470–57480. https://doi.org/10.1109/ACCESS.2023.3274478

[25] Mara Naegelin, Raphael P. Weibel, Jasmine I. Kerr, Victor R. Schinazi, Roberto La Marca, Florian von Wangenheim, Christoph Hoelscher, and Andrea Ferrario. 2023. An interpretable machine learning approach to multimodal stress detection in a simulated office environment. *Journal of Biomedical Informatics* 139 (3 2023). https://doi.org/10.1016/j.jbi.2023.104299

[26] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. https://doi.org/10.1136/bmj.n71

[27] K. C. Prabu Shankar and M. Hema. 2024. Detecting Mental Stress Using K-NN Classifier and IoT Devices. In *Lecture Notes in Networks and Systems*, Vol. 954 LNNS. Springer Science and Business Media Deutschland GmbH, 761–778. https://doi.org/10.1007/978-981-97-1724-8{_}66

[28] Darius Rountree-Harrison, Shlomo Berkovsky, and Maria Kangas. 2023. Heart and brain traumatic stress biomarker analysis with and without machine learning: A scoping review. , 27–49 pages. https://doi.org/10.1016/j.ijpsycho.2023.01.009

[29] Philip Schmidt, Attila Reiss, Robert Duerichen, and Kristof Van Laerhoven. 2018. Introducing WeSAD, a multimodal dataset for wearable stress and affect detection. In *ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction*. Association for Computing Machinery, Inc, 400–408. https://doi.org/10.1145/3242969.3242985

[30] Shikha, Dr Divyashikha Sethia, and S. Indu. 2024. Optimization of Wearable Biosensor Data for Stress Classification Using Machine Learning and Explainable AI. *IEEE Access* (2024). https://doi.org/10.1109/ACCESS.2024.3463742

[31] Pedro Elkind Velmovitsky, Paulo Alencar, Scott T. Leatherdale, Donald Cowan, and Plinio Pelegrini Morita. 2022. Using apple watch ECG data for heart rate variability monitoring and stress prediction: A pilot study. *Frontiers in Digital Health* 4 (12 2022). https://doi.org/10.3389/fdgth.2022.1058826

[32] Vividha, Drishti Agarwal, Paras Gupta, Soham Taneja, Preeti Nagrath, and Bhawna Gupta. 2022. Stress Prediction Using Machine Learning and IoT. In *Lecture Notes on Data Engineering and Communications Technologies*. Vol. 91. Springer Science and Business Media Deutschland GmbH, 615–624. https://doi.org/10.1007/978-981-16-6285-0{_}49

[33] Gideon Vos, Kelly Trinh, Zoltan Sarnyai, and Mostafa Rahimi Azghadi. 2023. Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review. https://doi.org/10.1016/j.ijmedinf.2023.105026

[34] Md Rahat Shahriar Zawad, Chowdhury Saleh Ahmed Rony, Md Yeaminul Haque, Md Hasan Al Banna, Mufti Mahmud, and M. Shamim Kaiser. 2023. A Hybrid Approach for Stress Prediction from Heart Rate Variability. In *Lecture Notes in Networks and Systems*, Vol. 519 LNNS. Springer Science and Business Media Deutschland GmbH, 111–121. https://doi.org/10.1007/978-981-19-5191-6{_}10