

PAPER • OPEN ACCESS

Mental Stress Detection Using Artificial Intelligence Models

To cite this article: Vishal Dham *et al* 2021 *J. Phys.: Conf. Ser.* **1950** 012047

View the [article online](#) for updates and enhancements.

You may also like

- [Real-time Mental Stress Detection Technique Using Neural Networks towards Wearable Health Monitor](#)
Nilava Mukherjee, Sumitra Mukhopadhyay and Rajarshi Gupta
- [Researching on Multiple Machine Learning for Anomaly Detection](#)
Yuanyuan Sun, Yongming Wang, Lili Guo et al.
- [Validation of a low intrusiveness heart rate sensor for stress assessment](#)
Tânia Pereira, Pedro R Almeida, João P S Cunha et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Extended abstract submission deadline: Dec 17, 2021

Connect. Engage. Champion. Empower. Accelerate.
Move science forward



Submit your abstract



Mental Stress Detection Using Artificial Intelligence Models

Vishal Dham¹, Kirtiman Rai¹ and Umang Soni^{1*}

¹Netaji Subhas University of Technology, New Delhi, India.

*corresponding author Email: umang.soni@nsut.ac.in

Abstract. Stress is a natural and common occurrence in humans. It leads to the release of hormones which help deal with the situation, but chronic stress affects our health and could lead to deleterious effects like depression, insomnia or headaches and therefore, early detection of stress becomes imperative to prevent such harmful consequences. This manuscript aims to automate the process of mental stress detection and help classify a stressed individual from a normal one through the use of physiological data collected from a wearable device. A publicly available dataset was used to evaluate our solution. Different Artificial Intelligence models like Artificial Neural Network (ANN), Hybrid of Artificial Neural Network and Support Vector Machine (ANN-SVM), Stacking Classifier and Radial Basis Function (RBF) Network were used, and their performance was compared using the accuracy of predicting correct stress state. During the study, Stacking Classifier gave the highest accuracy value of 99.92% while the RBF gave the least accuracy of 84.46% for three class classification of stress. The obtained results indicate the effectiveness of the proposed models in continuous monitoring of mental stress. The experimental results serve to demonstrate that the physiological signals can have a significant appositiveness in mental stress detection.

Keywords: Stress Detection, Artificial Intelligence, Machine Learning, Artificial Neural Network.

1. Introduction

Stress has become an inextricable part of our daily lives. It has come to be regarded as a noticeable concept in public health. The continuous presence of stress can lead to impairment of our health to a very large extent and can cause several inimical effects like high blood pressure, lack of sleep, increased risk of heart attack and other cardiovascular diseases. Since mental stress can become a pivotal cause of many diseases, a system that is capable of measuring stress levels from the signals generated through a wearable device can help create interesting real-world situations. Such a system has paramount importance when it comes to timely detection of stress to inform the user regarding the high stress level so that the user is made aware of the stress state to prevent further complications.

This study aims to use those physiological signals which proved to be effective indicators of mental stress by Schmidt et al. [13] and proposing some models which can be effective for the purpose. Such a stress detection system will have great applications in designing systems which are capable of detecting stress and taking preventive actions or issuing a warning quite well ahead in time so that the person can take precautionary measures. Stress detection can benefit researchers as they can obtain a more vivid and comprehensive view of how the users are affected by technology. This can also be useful to end-users who can subsume new applications according to their stress levels, in their work and personal lives.

2. Literature Review

The dataset that has been used in this paper is WESAD (Wearable Stress and Affect Detection), which is a multimodal dataset. This data set was obtained by experimenting on the 15 subjects by exposing them to different situations to evoke different affective states. The dataset contains three emotional states: neutral,



stress, and amusement. This dataset was created by Schmidt et al. [13] and was made publicly available for further exploration and analysis in the field of affect recognition.

A Reiss et al. [10] have used convolutional neural network on WESAD dataset for large scale heart rate estimation and compared the performance with other classical machine learning methods. Similarly, S. Chakraborty et al. [11] used multiple bio signals such as blood volume pulse (BVP), electrocardiography (ECG) and electromyography (EMG) and employed convolutional neural network architecture to classify bio signals to multiple state-of-mind categories. P Siirtola et al. [12] have utilized this dataset for finding the most important contributor for stress detection through the sensors in smartwatches. Schmidt et al. [13] have used several features based and convolutional neural network methods for multi-target affect detection by classifying different emotional states simultaneously using convolution neural network techniques. D Markovic et al. [14] have applied artificial neural network for patients' stress detection and used this in an IoT sensor data-based monitoring system collected from WESAD dataset. T Panure et al. [15] have used the WESAD dataset for collection of E4 device data and have implemented Artificial Neural Network, XGBoost and Support Vector Machine on the data for detecting stress.

Schmidt et al. [1] have used Decision Trees, Ada Boost, Random Forest, Linear Discriminant Analysis and k-nearest neighbors to benchmark three emotional states namely baseline, stress and amusement and also for binary classification of stress and non-stress states. The authors also performed a comparative analysis of the different machine learning algorithms that have been used by them. Haag et al. [2] have used Neural Network to recognize emotions using multiple signals from various biosensors. Wagner [3] have used Linear Discriminant Function, Neural Network and k-nearest neighbors to perform supervised classification of emotion using facial expression, speech and physiological signals. Similarly, Kim et al. [4] have used Linear Discriminant Analysis for emotion recognition by classifying four musical emotions. SVM has been used by Katsis et al. [5] for classifying the emotional states of car racing drivers through data collected by the biosensors. Sano and Picard [6] have employed Support Vector Machine and k-nearest neighbors to detect stress through data collected by wearable sensors and mobile phone usage. Jaques et al. [7] have used Support Vector Machine, Logistic Regression and Neural Network to predict the mood, stress level and health using personalized multitask learning and domain adaptation. Mozos et al. [8] have used k-Nearest Neighbor, Support Vector Machine and AdaBoost for automatic stress detection in social situations through the data captured by the sensor systems. Zhao et al. [9] have used Random Forest, Neural Network, Naive Bayes and based models for determining a person's emotion from RF signals reflected off the body. Till now Hybrid of Artificial Neural Network and Support Vector Machine (ANN-SVM), Stacking Classifier and Radial Basis Function (RBF) Network have not been tested in previous works.

3. Artificial Intelligence model overview

Various artificial intelligence models have been used before for stress and affect detection. A brief outline of the models that have been evaluated in this paper are presented in the forthcoming section.

3.1 Dataset Review

Data was collected by [13] by exposing 15 subjects to different situations and then recording the sensor output from the chest and wrist wearable device continuously. Target labels comprised of three classes - 'Neutral', 'Stress' and 'Amusement'. The neutral condition was recorded for 20 minutes by providing them with a table and reading materials to induce a neutral affective state. For the amusement condition, subjects were shown eleven funny video clips for 392 seconds. Stress was induced to each subject by making them speak in front of a panel without notifying them beforehand that they will be speaking publicly. After that, they were asked to count back from 2023 to 0 with steps of 17 without any mistake otherwise they had to start over from 2023. This session had a length of 5 minutes.

WESAD dataset consisted of data from two sources – Chest Wearable Device and Wrist Wearable Device. Due to variations in recordings from Wrist wearable, we have only used data from Chest wearable. It consisted of 8 features – Acceleration in three axes, Body Temperature, Respiration Level, EDA, ECG and EMG. Due to the huge amount of data, only 10% of randomly drawn samples of each subject were extracted for analysis totaling around 2.3 million rows of data to test with. Data was normalized to have mean centered at zero and unit standard deviation and then split in such a way that 70% data was used for training and 30% of the data was used for testing. All samplings were done in a stratified manner such that proportions of the target label were the same in both training and test set.

3.2 Artificial Neural Network

Artificial Neural Network (ANN) is a biologically inspired supervised learning algorithm which is analogues to biological neurons [20]. In biological neurons, electrical signals are received by dendrite from the axons of other neurons. While in the case of ANN, this role is performed by the numerical values which are mapped from input layers to output layers, passing through single or multiple hidden layers. Components of a hidden layer can be considered as a feature space which ANN has learned itself. While mapping from a particular layer 'l' to another layer 'l+1', a weighted sum of components for layer 'l' is first computed and then passed through an Activation Function. There is no theory behind the selection of an exact number of hidden neurons and layers. Hit and Trial method can be used to obtain optimal hyperparameters to minimize bias and variance. At least one hidden layer is required when dealing with non-linear data. The schematic layout of an artificial neural network algorithm is shown in Figure 1.

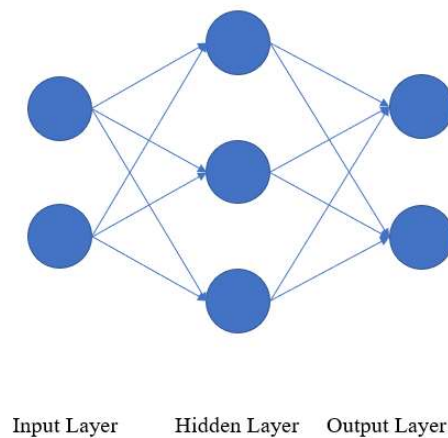


Figure 1. Illustrative representation of an artificial neural network

A single neuron is used in the output layer when regression has to be performed. For classification purposes, the number of target variables determine the number of neurons in the output layer whereas number of features or predictors in the dataset decides the number of neurons in the input layer. Activation functions are needed to capture non-linearity in the data [21] otherwise regardless of multiple hidden layers, the relationship between input and output when simplified will just be linear equation with high number of parameters. We have used Rectified Linear Unit (ReLU) and SoftMax activation function in our experimentation. ReLU function (denoted by $R(x)$) is a piecewise linear function that outputs zero for a negative value and returns the same value for positive input values. ReLU has faster convergence as

compared to sigmoid and tanh activation functions. Softmax (denoted by $S(x)$) is used to give class wise probability as it gives the sum of all the output as one thereby making it convenient for computing the class wise probability. Assuming K as the number of classes and x as the input, the expressions for ReLu and Softmax activation functions are as shown below:

$$R(x) = \max \{0, x\} \quad (1)$$

$$S(x)_i = \frac{e^{x_i}}{\sum_{i=0}^{K-1} e^{x_i}} \quad (2)$$

ANN is an optimization problem. We have trained our ANN using Stochastic Gradient Descent using mini batches. Backpropagation uses Chain rule of derivatives to compute gradients of our loss function with respect to weights. The cross-entropy loss function is used for this purpose. It is the most common loss function for classification problem. If we are predicting incorrect class with high confidence/probability, then our loss function will have a very high value and if we are predicting correct class with high confidence/probability then the output will be close to zero [23,24]. Gradient descent keeps updating weights in the direction of decreasing loss to find an optimal solution.

$$L(Y, \hat{Y}) = - \sum_{i=0}^{K-1} Y_i \log(\hat{Y}_i) \quad (3)$$

In the above expression, Y_i is the binary true value of i^{th} class and \hat{Y}_i is the predicted classwise probability for that class. The value that is fed into the gradient descent is the average of the losses of that mini batch. It is always better to normalize data before feeding it to ANN, but there are chances that variance is introduced during backpropagation [22]. Therefore, batch normalization is used in order to control the variance before feeding it to activation. γ and β are two new parameters that are learned while training our ANN. Mean and variance are estimated using the current batch.

$$W_j = \gamma_j \left[\frac{W_j - \mu_j}{\sqrt{\sigma_j}} \right] + \beta_j \quad (4)$$

We have used Keras framework with TensorFlow backend for training ANN containing one hidden layer with 32 neurons. Hidden layer is activated using ReLu and Output layer is activated using Softmax. Batch size of 4096 was selected and ANN was trained for 20 epochs.

3.3 Hybrid of Artificial Neural Network and Support Vector Machine

Support Vector Machine (SVM) is an algorithm which boundary function is a hyperplane that separates the classes by maximum margin [25]. Instead of just finding the plane which divides the classes, it finds the decision boundary with some width in between various classes [17]. Due to the size of the dataset, we have used a linear kernel instead of the Gaussian kernel.

This model was inspired from Hongbin et al. [16]. We have first used a pre-trained ANN model of the above-mentioned model and used the values of hidden layers as the inputs of our Linear-kernel SVM. But instead of using different SVM with different kernels for each subset of feature space, we have trained one

SVM model. Schematic layout of this hybrid algorithm is shown in Figure 2. With this hybrid model advantages of SVM can be used by training it on ANN learned features.

We have used Keras framework by first feeding the input to the hidden layer with 32 neurons and then using the same hidden layer units as inputs for SVM which was implemented using sci-kit learn library and then finally making final predictions.

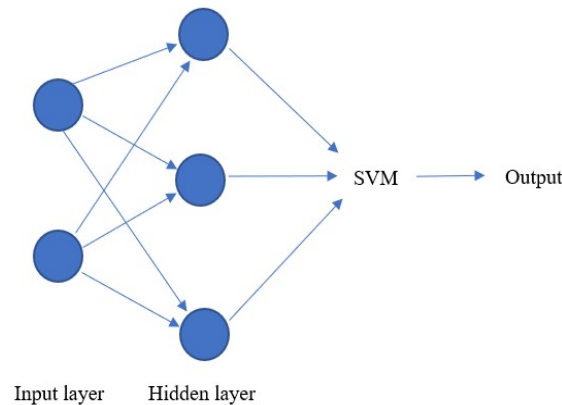


Figure 2. Illustrative representation of our Hybrid ANN-SVM

3.4 Stacking Classifier

Schmidt et al. [13] have used K-Nearest Neighbors ($k = 9$), Linear Discriminant Analysis, AdaBoost, Random Forrest, Decision Tree Classifier for their study in their paper. We implemented an ensemble of these models but instead of Voting or averaging the results we fed the output of all models to an ANN as final estimator. K-nearest neighbors (k-NN) algorithm is a supervised machine learning algorithm which can be used for both classifications as well as regression problems. In this algorithm, a prediction of the class to which the data point belongs is made depending upon the proximity of the data point to the nearest group [27]. This algorithm is an example of a "lazy learner" algorithm [26] because it does not model a function but instead, it refers to the training set. K-NN is also categorized as one of the non-parameterized algorithms.

Linear Discriminant Analysis (LDA) is another supervised machine learning algorithm which is based upon the concept of finding an optimal linear combination of variables (predictors) that can best separate two target classes. The main aim of this algorithm is to bring the features in higher dimension space onto a lower-dimensional space thereby reducing its dimensionality [28]. It drives a decision plane which minimises the within-class variance and maximises the between-class variance. LDA uses Bayes theorem and assumes that the data points in each class are normally distributed.

AdaBoost is a boosting algorithm that aims at converting a set of weak classifiers into a strong one. The final output of the classifier is the weighted sum of the outputs of the other learning algorithms [19]. AdaBoost is used to enhance the performance of decision trees on binary classification problems [29]. This algorithm is sensitive to noisy data and outliers. A decision tree is a tree-based model. The structure of a decision tree comprises of an internal node, branch and leaf nodes which represent a feature, decision rule and the outcome respectively. Root node is the topmost node in a decision tree [30]. The algorithm learns to classify based on attribute's value. The training set is iteratively split by the decision tree until each part comprises completely of the sample from one class [18]. Decision trees successfully unite a sequence of a basic test where every test compares a numeric value against a threshold value.

Random Forest Classifier is an ensemble algorithm which uses randomly selected samples from the training set to form a set of randomized decision trees. After this, it grosses the votes from different decision trees to predict the final class [31]. Random subset of features is used for each decision tree in the random forest. Firstly, all individual models were trained, and class-wise probabilities were computed. Since there are 5 models and 3 classes, the input of this ANN had 15 units as shown in Figure 3. ANN was trained using stochastic gradient descent with mini-batches for minimizing cross-entropy loss function. SoftMax activation function was used for the output layer.

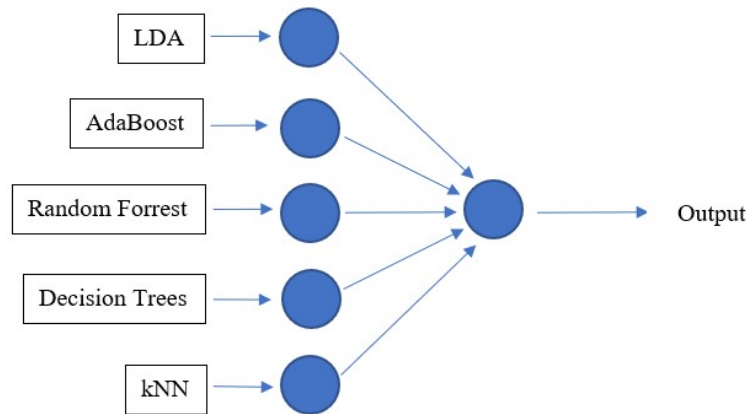


Figure 3. Illustrative representation of our Stacking Classifier

3.5 Radial Basis Function (RBF)

Radial Basis Function (RBF) Network has only one hidden layer [32]. Similar to Artificial Neural Networks, input layer consists of all the predictors. In RBF Network, Euclidean distance between input and a hidden node is computed and passed to Radial Basis Function (generally Gaussian) which acts as an activation function [33]. The output of this layer is a distribution having a mean and a standard deviation. If the input is close to the center then the activation shoots up, otherwise, if the input is far away, the activation will be low. For the output layer, the weighted sum is calculated of these activations to obtain class-wise probability which can then be used for classification.

$$h(x|l) = \|x - l\|_2 \quad (5)$$

$$G(x) = \frac{e^{\frac{-1(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \quad (6)$$

Values of these hidden nodes, which are compared with the input vector, are a random sample of data from the training set. The number of samples to be drawn depends on our choice of hidden neurons. These nodes can either be selected randomly or using a clustering algorithm like K-Means and then finding the center of each cluster. We used K-Means to find the centers (denoted by l). SoftMax activation function was used for the output layer. Similar to our previous approach, RBF Network was trained using Gradient Descent with 32 hidden units. The schematic diagram of an RBF network is shown in Figure 4.

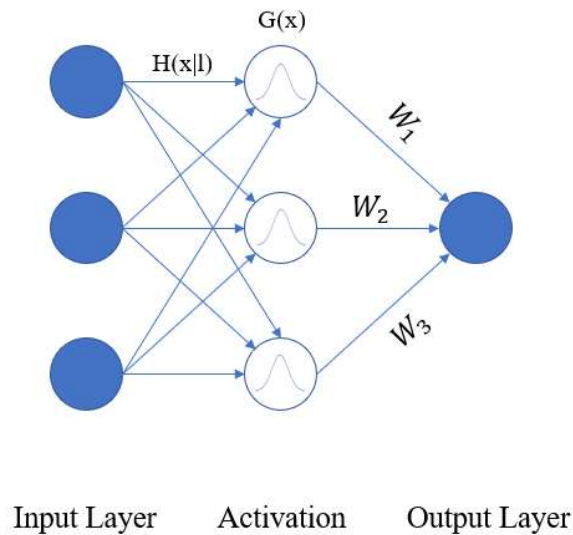


Figure 4. Illustrative representation of an RBF network

4. Results and discussions

The models that have been evaluated in this study are: Artificial Neural Network (ANN), Hybrid of Artificial Neural Network and Support Vector Machine (SVM), Stacking Classifier and Radial Basis Function (RBF). The accuracy in predicting the correct target label on the test data was the metric used for evaluation of the performance of all the models. All the previous papers on WESAD have used accuracy as their metric to judge the performance of their classification algorithms. Performance of the models that have been used in this paper is shown in Table 1.

Table 1. Performance of all the models

Models	Accuracy (in %)
Artificial Neural Network (ANN)	90.58
Artificial Neural Network-SVM	91.48
Stacking Classifier	99.92
Radial Basis Function (RBF)	84.46

It can be observed that among all the models, Stacking Classifier gave the best accuracy of 99.92%. Moreover, RBF network was the most expensive computationally, since apart from backpropagation, K-Means algorithm also had to be performed in order to find centers of different clusters. RBF network took the longest to train and yielded an accuracy of 84.46%, which was the lowest among all the models that have been tested in this paper. There was a slight increase in the performance by using our proposed ANN-SVM compared to ANN which shows that using learned features from hidden layers for hybrid of ANN-SVM proved to be effective in increasing performance of model for our study. The accuracy increased from 90.58 to 91.48 which indicates that ANN-SVM is able to better classify about 6,200 more data points

correctly. The comparison of all the tested models is shown in Figure 5 and the Confusion Matrix of Stacking Classifier is displayed in Table 2.

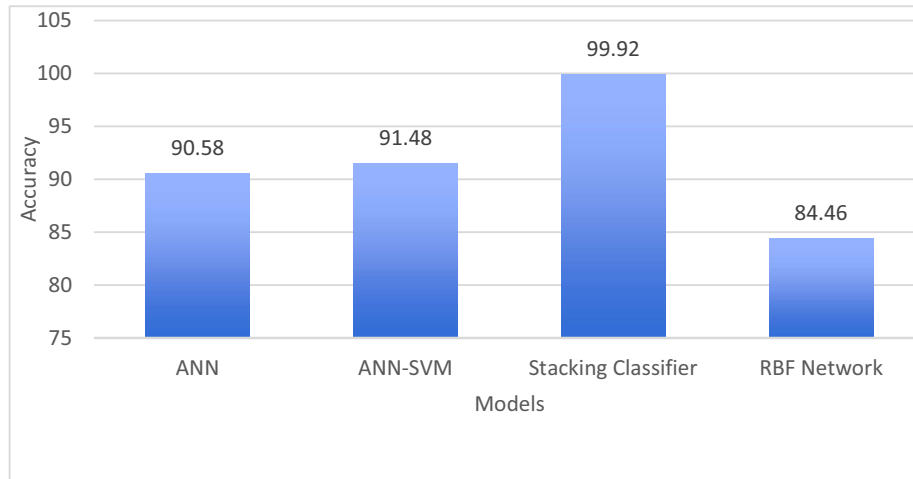


Figure 5. Comparison of all the tested model

Table 2. Confusion Matrix of Stacking Classifier

	Predicted 'Neutral'	Predicted 'Stress'	Predicted 'Amusement'
Actual 'Neutral'	369,761	59	11
Actual 'Stress'	88	209,066	132
Actual 'Amusement'	30116868	177	116,868

5. Conclusion

This paper serves to demonstrate the capability of deep neural networks and tree-based machine learning models for developing resilient and robust methods for stress detection with the help of physiological signals collected from wearable devices. In this paper, we propose four different neural network models that can be used for three class classification of stress: baseline, stress and amusement on the publicly available WESAD dataset. The algorithms that have been implemented in this paper are- ANN, hybrid of ANN-SVM, Stacking Classifier and RBF network. We introduced Hybrid of ANN-SVM, Stacking Classifier and RBF Network which have not been used for WESAD. The Stacking Classifier gave the most prominent results with a classification accuracy of 99.92% and RBF network gave the least accuracy at 84.46%.

Mental stress can be detected using various AI models. Keeping in mind the consequences of chronic stress, AI models can be deployed for building a continuous monitoring wearable system which detects stress level of a person and alerts if the stress level is continuously high for a certain threshold of time. Keeping the stress level in check is important for any individual whether they are a student or a working professional. Our proposed classification algorithms can make a practicable prediction by using the data from a wearable device.

In future, more work can be done by taking the self-reports of the subjects into account instead of using organized questionnaires. Further, features of WESAD dataset can be set as a reference in order to collect

more data in the future for further research. Also, AI models can be developed to more precisely classify the different affect level (like low, moderate and high stress). Different procedures like audio recordings, facial cues etc., that have been used distinctly can be merged with the physiological data to create a new dataset, which can detect stress more precisely as it will have nearly all the features that are required for inducing stress in human beings.

References

- [1] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018, October). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction* (pp. 400-408). ACM.
- [2] Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004, June). Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems* (pp. 36-48). Springer, Berlin, Heidelberg.
- [3] Wagner, J., Kim, J., & André, E. (2005, July). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *2005 IEEE international conference on multimedia and expo* (pp. 940-943). IEEE.
- [4] Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12), 2067-2083.
- [5] Katsis, C. D., Katertsidis, N., Ganiatsas, G., & Fotiadis, D. I. (2008). Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(3), 502-512.
- [6] Sano, A., & Picard, R. W. (2013, September). Stress recognition using wearable sensors and mobile phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 671-676). IEEE.
- [7] Jaques, N., Taylor, S., Sano, A., & Picard, R. (2017, September). Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. In *IJCAI 2017 Workshop on Artificial Intelligence in Affective Computing* (pp. 17-33).
- [8] Mozos, O. M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., & Ferrandez, J. M. (2017). Stress detection using wearable physiological and sociometric sensors. *International journal of neural systems*, 27(02), 1650041.
- [9] Zhao, M., Adib, F., & Katabi, D. (2016, October). Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (pp. 95-108). ACM.
- [10] Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks. *Sensors*, 19(14), 3079.
- [11] Chakraborty, S., Aich, S., Joo, M. I., Sain, M., & Kim, H. C. (2019). A Multichannel Convolutional Neural Network Architecture for the Detection of the State of Mind Using Physiological Signals from Wearable Devices. *Journal of healthcare engineering*, 2019.
- [12] Siirtola, P., Peltonen, E., Koskimäki, H., Mönttinen, H., Rönning, J., & Pirttikangas, S. (2019, June). Wrist-worn Wearable Sensors to Understand Insides of the Human Body: Data Quality and Quantity. In *The 5th ACM Workshop on Wearable Systems and Applications* (pp. 17-21).
- [13] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018, October). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction* (pp. 400-408). ACM.
- [14] Marković, D., Vujičić, D., Stojić, D., Jovanović, Ž., Pešović, U., & Randić, S. (2019, March). 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)* (pp. 1-6). IEEE.
- [15] Panure, T., & Sonawani, S. (2019, July). Novel Approach for Stress Detection Using Smartphone and E4 Device. In *International Conference on Sustainable Communication Networks and Application* (pp. 736-745). Springer, Cham.
- [16] Liu, H., Xie, D., & Wu, W. (2008). Soil water content forecasting by ANN and SVM hybrid architecture. *Environmental monitoring and assessment*, 143(1-3), 187.
- [17] Wang, W., Zheng, H., & Wu, Y. J. (2020). Prediction of fundraising outcomes for crowdfunding projects based on deep learning: a multimodel comparative study. *Soft Computing*, 1-19.
- [18] Khare, N., & Viswanathan, P. (2020). Decision Tree-Based Fraud Detection Mechanism by Analyzing Uncertain Data in Banking System. In *Emerging Research in Data Engineering Systems and Computer Communications* (pp. 79-90). Springer, Singapore.
- [19] Bahad, P., & Saxena, P. (2020). Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics. In *International Conference on Intelligent Computing and Smart Communication 2019* (pp. 235-244). Springer, Singapore.
- [20] Li, T., & Gong, Q. (2013). A Novel Artificial Neural Network Learning Algorithm. In *Proceedings of the 2012 International*

- Conference of Modern Computer Science and Applications* (pp. 83-88). Springer, Berlin, Heidelberg.
- [21] Velasco, L. C. P., Estoperez, N. R., Jayson, R. J. R., Sabijon, C. J. T., & Sayles, V. C. (2019, March). Performance Analysis of Artificial Neural Networks Training Algorithms and Activation Functions in Day-Ahead Base, Intermediate, and Peak Load Forecasting. In *Future of Information and Communication Conference* (pp. 284-298). Springer, Cham.
 - [22] Hong, J., & Liu, J. (2020). Cerebral Microbleeds Detection via Convolutional Neural Network with and Without Batch Normalization. In *Frontiers in Intelligent Computing: Theory and Applications* (pp. 152-162). Springer, Singapore.
 - [23] Ng, G. S., Chan, K. H., Erdogan, S. S., & Singh, H. (2000). Neural network learning using entropy cycle. *Knowledge and Information Systems*, 2(1), 53-72.
 - [24] Kline, D. M., & Berardi, V. L. (2005). Revisiting squared-error and cross-entropy functions for training neural network classifiers. *Neural Computing & Applications*, 14(4), 310-318.
 - [25] Saadna, Y., Behloul, A., & Mezzoudj, S. (2019). Speed limit sign detection and recognition system using SVM and MNIST datasets. *Neural Computing and Applications*, 1-11.
 - [26] Zhang, M. L., & Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7), 2038-2048.
 - [27] Kumar, K. S., & Bhavani, R. (2017). Human activity recognition in egocentric video using PNN, SVM, kNN and SVM+ kNN classifiers. *Cluster Computing*, 1-10.
 - [28] Yuan, Y., Zhao, K., & Lu, H. (2014, November). Multi-label linear discriminant analysis with locality consistency. In *International Conference on Neural Information Processing* (pp. 386-394). Springer, Cham.
 - [29] Badrinath, N., Gopinath, G., Ravichandran, K. S., & Soundhar, R. G. (2016). Estimation of automatic detection of erythematous diseases through adaboost and its hybrid classifiers. *Artificial Intelligence Review*, 45(4), 471-488.
 - [30] Panhalkar, A., & Doye, D. (2017). An outlook in some aspects of hybrid decision tree classification approach: a survey. In *Proceedings of the International Conference on Data Engineering and Communication Technology* (pp. 85-95). Springer, Singapore.
 - [31] Sandri, M., & Zuccolotto, P. (2006). Variable selection using random forests. In *Data analysis, classification and the forward search* (pp. 263-270). Springer, Berlin, Heidelberg.
 - [32] Wu, X. J., Jiang, G. C., Wang, X. J., Fang, N., Zhao, L., Ma, Y. M., & Luo, S. J. (2013). Prediction of reservoir sensitivity using RBF neural network with trainable radial basis function. *Neural Computing and Applications*, 22(5), 947-953.
 - [33] Sandoval, G., Vazquez, R. A., Garcia, P., & Ambrosio, J. (2014, June). Crop classification using different color spaces and rbf neural networks. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 598-609). Springer, Cham.