Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020)
IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2

Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data

Pramod Bobade
Department of Computer Science and Engineering
National Institute of Technology, Karnataka
Surathkal, India
pramodbobade2@gmail.com

Vani M.
Department of Computer Science and Engineering
National Institute of Technology, Karnataka
Surathkal, India
vani.nitk@gmail.com

***Abstract*—Stress is a common part of everyday life that most people have to deal with on various occasions. However, having long-term stress, or a high degree of stress, will hinder our safety and disrupt our normal lives. Detecting mental stress earlier can prevent many health problems associated with stress. When a person gets stressed, there are notable shifts in various bio-signals like thermal, electrical, impedance, acoustic, optical, etc., by using such bio-signals stress levels can be identified. This paper proposes different machine learning and deep learning techniques for stress detection on individuals using multimodal dataset recorded from wearable physiological and motion sensors, which can prevent a person from various stress-related health problems. Data of sensor modalities like three-axis acceleration (ACC), electrocardiogram (ECG), blood volume pulse (BVP), body temperature (TEMP), respiration (RESP), electromyogram (EMG) and electrodermal activity (EDA) are for three physiological conditions - amusement, neutral and stress states, are taken from WESAD dataset. The accuracies for three-class (amusement vs. baseline vs. stress) and binary (stress vs. non-stress) classifications were evaluated and compared by using machine learning techniques like K-Nearest Neighbour, Linear Discriminant Analysis, Random Forest, Decision Tree, AdaBoost and Kernel Support Vector Machine. Besides, simple feed forward deep learning artificial neural network is introduced for these three-class and binary classifications. During the study, by using machine learning techniques, accuracies of up to 81.65% and 93.20% are achieved for three-class and binary classification problems respectively, and by using deep learning, the achieved accuracy is up to 84.32% and 95.21% respectively.**

***Index Terms*—photoplethysmography, stressors, accelerometer, dichotomous, sudomotor nerve activity, convex optimization**

I. Introduction

A person's standard of living is significantly affected by his emotional states such as stress and anxiety. According to S. Palmer [17], "Stress is often described as a complex psychological and behavioral state resulting from the perception of a significant imbalance between the demands placed on the individual and their perceived ability to meet these demands". Stress affects both mental and physical health by causing problems such as abnormalities in the cardiac rhythm i.e. arrhythmia and depression. According to the American Institute of Stress [14], 80% of workers feel stress on the job and nearly half say they need help in learning how to manage stress and 42% say their co-workers need such help. According to Health and Safety Executive (HSE), work-related stress, depression or anxiety accounted for 44% of all work-related ill health cases and 54% of all working days lost due to ill health in 2018/19 [15]. Studies carried on humans and animals suggest that stress may affect the immune system and enhance the possibility of cancer. These statistics and effects of stress on humans' health, calls for the system that can detect stress conditions so that stress can be relieved either by personalized interventions or by medications if needed.

Conventionally, psychological and physiological specialists decide stress condition of an individual using questionnaire based stress analysis. This approach carries lot of uncertainty and is unreliable as it depends entirely on the individuals' responses and the people will be timorous to answer the questionnaire.

Many research were undertaken to elicit and detect stress, based on a person's physiological parameters. A situation can be stressful when something psychological, such as persistent worry about losing a job, approaching work deadline, etc., happens. Such a stressful situation can trigger a course of stress hormones resulting in relevant physiological changes like pounding heart, the quickening of breath, tensed muscles, appearance of beads of sweat, etc. The body's physical ("fight-or-flight") response is a result of such physiological changes taking place. During these physiological changes, corresponding biosignals emanate from affected individuals. These biosignals help to detect stress by quantifying the physiological measures of an individual. Various physical sensors were employed for this purpose of automatic stress detection.

The objective of the proposed work is to automatically detect the stress condition of an individual by using the physiological data recorded during the stressful situations. Such a detection can help in monitoring stress to prevent dangerous stress-related diseases.Various machine learning and deep learning techniques are used for such stress detection and identifying a person as stressed or unstressed (also normal or amused or stressed). For achieving this objective, sequence of steps are carried out such as understanding the structure and format of the publicly available WESAD dataset, cleaning and transforming data to a set eligible to construct machine learning and deep learning classification methods, exploring and constructing various classification models and comparing them.

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.

II. Related work

In recent years, there are efforts to automate the prediction and detection of stress by machine learning models, which are trained using physiological responses to stress and emotional stimuli.

Philip Schmidt, et al. [1] had introduced WESAD dataset for the purpose of wearable affect and stress detection and made it available to the public. For collecting this data they had chosen 15 people and recorded the physiological data such as three-axis acceleration, electrocardiogram, blood volume pulse, body temperature, respiration, electromyogram and electrodermal activity by putting wearable devices - RespiBAN Professional and Empatica E4 on the chest and on the wrist respectively. They subjected the subjects to various stress conditions such as baseline, amusement, stress, meditation, etc. They had used and compared the performance of five machine learning algorithms for stress state detection: K-Nearest Neighbour (KNN), Linear Discriminant Analysis (LDA), Random Forest (RF), Decision Tree (DT), AdaBoost (AB). They achieved classification accuracies of up to 80.34% and 93.12% considering the three-class (amusement vs. baseline vs. stress) and binary (stress vs. non-stress) classification problems, respectively, by using common features and classical machine learning methods.

Jacqueline Wijsman, et al. [7] also measured physiological signals using wearable sensors for the purpose of detecting mental stress. They recorded ECG, skin conductance, respiration, and EMG of the participants, and calculated total of 19 physiological features from them. For further analysis, a subset of 9 features out of these 19 features was selected after studying correlations and normalizing feature values, which are then reduced to 7 features after using Principal component analysis. By using these features and different classifiers such as Linear Bayes Normal Classifier, Quadratic Bayes Normal Classifier, K-Nearest Neighbors Classifier, Fisher's Least Square Linear Classifier, an accuracy of 80% was obtained between stress and non-stress conditions of an individual. This experiment is almost similar to what [1] have performed except for the number of participants and the features extracted. They had used three different stressors in their study and compared their results to other papers on stress classification which used only one type of stressor.

New multimodal dataset named SWELL Knowledge Work (SWELL-KW) dataset was developed by Saskia Koldijk, et al. [6] for stress related research and user modeling. This dataset was collected using 25 people performing ordinary knowledge works like writing, reading, searching, etc. and by manipulating the working conditions using two stressors: time pressure and email interruptions. Recorded data includes information on body postures, facial expression, computer logging, skin conductance and heart rate. This dataset containing raw and preprocessed data with extracted features was made available to all. The dataset on working behavior and affect was assessed with validated questionnaires relevant to task load, mental effort, etc. In this study, none of the machine learning techniques was used for the benchmark, but it has introduced a new stress-related dataset to the literature.

The BioNomadix model BN-PPGED from Biopac, was a wearable device that was used to measure physiological responses in [4]. The participant wore the BN-PPGED on his non-dominant hand just like a wristband, with two electrodes located on two fingers measuring the signals of pulse plethysmograph (PPG) and electrodermal activity (EDA). Additionally, the PPG autocorrelation signal and the Heart Rate Variability (HRV) were extracted using AcqKnowledge software. Support vector machine (SVM) was used for the classification of a person as stressed or non-stressed, which resulted in an accuracy of 82%.

Addressing the employees report on stress at work, Saskia Koldijk, et al. [3] developed automatic classifiers to examine the relation between working conditions and mental stress-related conditions from sensor data: body postures, facial expression, computer logging and physiology(ECG and skin conductance). They found that the performance of the specialized model was just as well or better than a generic model in almost all cases when similar users are subgrouped, and models were trained on specific subgroups. In order to differentiate between a stressor and non-stressor working conditions, posture provides the most crucial information among the most useful modalities. Performance could be further improved by adding data about one's facial expressions. They got an accuracy of 90% employing SVM classifier.

Facial cues are another significant factor that can define a person's stress. Adding to this, G. Giannakakisa, et al. [8], developed a framework for detecting and analyzing emotional states of stress/anxiety via video-recorded facial cues. The investigated features were mouth activity, events related to eye, camera-based photoplethysmographic estimation of heart rate, and parameters of head action. Participants were made to sit 50 cm apart in front of a camera integrated computer monitor. Methods like Generalized Likelihood Ratio, Naive Bayes classifier, Support Vector Machines, K-nearest neighbors, and AdaBoost classifier were used and tested. In the social exposure process, the highest classification performance was achieved using the Adaboost classifier reaching an accuracy of 91.68%.

Among various available bio-signals, Md Fahim Rizwan, et al. [5] used ECG feature for the stress condition classification. Because of ECG feature extraction techniques and availability of various portable clinical grade recorders making its recording readily available, ECG was chosen as the primary candidate. There is no need for separate respiration measurement sensor system as a respiratory signal information can be detected from ECG using the EDR technique (i.e., ECG derived Respiration), making ECG more advantageous. Using RR interval, QT interval and EDR features with SVM classification method, the accuracy of almost 98.6% was achieved by them. But this result is not satisfactory because it had used only one signal i.e., ECG and had not considered other crucial signals from the body, which are also important for stress induction.

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SUMMARY TABLE OF REVIEWED ARTICLES WITH METHODOLOGY & PERFORMANCE EVALUATION.

Sr. No.	Title	Author	Methodology & Performance
1	Towards Mental Stress Detection Using Wearable Physiological Sensors.	Jacqueline Wijsman, Bernard Grundlehner, Hao Liu, Hermie Hermens	ECG, respiration, skin conductance, & EMG of the trapezius muscles was recorded. Accuracy of 80% by kNN(two class) achieved.
2	The SWELL Knowledge Work Dataset for Stress & User Modelling Research	Saskia Koldijk , Mark A. Neerincx, and Wessel Kraaij.	Introduce SWELL-KW data-set. Collected data by computer logging, face expression from camera recordings, body postures from a Kinect 3D sensor and heart rate (variability) & skin conductance from body sensors.
3	Stress Detection Using Wearable Physiological Sensors	Virginia Sandulescu, Sally Andrews, David Ellis, et.al.	Used a wrist worn device named BN-PPGED for data collection. Accuracy of 82% was achieved by using SVM.
4	Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection	Philip Schmidt, Attila Reiss, Robert Durichen, et.al.	Introduced the WESAD dataset. A benchmark is created on the dataset, using well-known features & standard machine learning methods. Accuracy of 80% (three class) and 93%(two class) was achieved.
5	Detecting Work Stress in Offices by Combining Unobtrusive Sensors	Saskia Koldijk , Mark A. Neerincx, and Wessel Kraaij.	Computer logging, facial expressions, posture of Employees. Accuracy of 90% using SVM.
6	Design of a Biosignal Based Stress Detection System using Machine Learning Techniques	Md Fahim Rizwan, Rayed Farhad, et.al.	ECG was selected as the primary candidate to stress detection based on RR interval, QT interval, etc. Accuracy of 98% using cubic SVM.
7	Stress and anxiety detection using facial cues from videos	G. Giannakakisa, M.Pediaditisa.	Used video-recorded facial cues and achieved accuracy of 91.68% for classification.
8	Automatic Stress Detection in working environments from smartphones' accelerometer data: A First Step	Enrique Garcia-Ceja, Venet Osmani and Oscar Mayora	Used data from the built-in smartphone accelerometer sensor to identify activity that corresponds with stress levels and achieved accuracy of 71%.
9	Continuous stress detection using a wrist device: In laboratory and real life.	M. Gjoreski, H. Gjoreski, and M. Gams.	Achieved 83% accuracy on a binary class problem using data provided from a commercial wrist device.
10	Emotion recognition based on physiological changes in music listening	Elisabeth Andre, Jonghwa Kim	Made users listen to 4 songs to record their emotional arousal and achieved a classification accuracy of 70%.
11	A Machine learning approach for stress detection using a wireless physical activity	B. Padmaja, V. V. Rama Prasad and K. V. N. Sunitha	Used data collected from FITBIT and achieved an accuracy of 62.14%.

Another work carried out for detection of the working environment was by Enrique Garcia-Ceja, et al. [2] who used data from the built-in smartphone accelerometer sensor to identify activity that corresponds with stress levels of the subjects. This sensor has been selected as privacy concerns are lesser as compared to location, audio, or video recording. Another reason for selecting this sensor is its suitability to get embedded in smaller wearable devices such as fitness trackers due to its low power consumption. Smartphones were provided to 30 subjects from two different organizations. Using similar users and user-specific models they reached an accuracy of 60% and 71%, respectively, only relying on data from a single smartphone's accelerometer, which in turn isn't sufficient for stress detection.

Some of the classical machine learning algorithms such as Random Forest were employed for stress classification, which in turn achieved 83% accuracy on a binary class ("No stress" vs. "Stress") problem [9] using data provided from a commercial wrist device. Elisabeth Andre, Jonghwa Kim [10] made users listen to 4 songs to record their emotional arousal and achieved a classification accuracy of 70% using EDMC (emotion-specific multilevel dichotomous classification), which is subject-independent. In a study [11] by Kurt Plarre, et al., stress levels were reported by participants by giving answers on a four-point scale (0=NO, 1=no, 2=yes, 3=YES) to questions like Cheerful?, Sad?, Stressed/Nervous?, Frustrated/Angry?, Happy? along with the physiological measurements captured by unobtrusive, wearable sensors and achieved an accuracy of 90% for binary classification.

Alexandros Zenonos, et al. [12] developed a mood identification system that was able to differentiate eight different kinds of states with their five intensity levels with subject-independent accuracy of 62.14%. In a study [13], Machine learning was applied to the data collected from the wireless activity tracker i.e., FITBIT. Various attributes, such as working hours, time in bed, minutes asleep, BMI, cardio, maximum heart rate, fat burn, etc. were used for stress detection. Data from Indian people working in IT and other sectors were taken into this system. Table I gives the summaries of reviewed articles.

Present study is an effort to analyze the biosignals through machine learning & deep learning models to detect stress related condition of an individual. WESAD dataset is used for this study, which is a multimodal physiological / biosignals dataset collected from the people through non-invasive techniques. Using machine learning algorithms like K-Nearest Neighbour (KNN), Linear Discriminant Analysis (LDA), Random Forest (RF), Decision Tree (DT), AdaBoost (AB) and SVM, and simple deep learning model, subjects are classified according to their stress conditions. This reliefs a doctor or

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.

psychiatrist from doing it manually and can be more reliable. After classification, if a person is found to have stressed, he can be given proper counseling or stress relief medications if required.

III. Methodology

A. Dataset and Features Extraction

WESAD is the dataset that is used for this study. This dataset was introduced and made publicly available by Attila Reiss, Philip Schmidt, et al. in 2018 [1]. This multimodal dataset is the collection of motion data and physiological features of 15 subjects from both a chest-worn device RespiBAN Professional and a wrist-worn device Empatica E4. Subjects were put into various study protocols such as preparation, baseline condition, amusement condition, stress condition, meditation, recovery, and their physiological stimuli were recorded. Reference [1] describes the details about sensor setup, sensor placement, and the procedure carried out to build this dataset, including which data are collected during which study protocol of the subject.

The RespiBAN measured ACC, RESP, ECG, EDA, EMG, and TEMP. All signals were sampled at 700 Hz. The E4 measured TEMP, EDA, ACC, and BVP sampled at 4 Hz, 4 Hz, 32 Hz, and 64 Hz, respectively. The dataset is organized so that each subject has a folder (SX, where X = subject ID). Each subject folder contains the following files:

- SX_readme.txt: includes information about the subject (SX) and information about data collection and data quality (if applicable).
- SX_quest.csv: includes all relevant information to obtain ground truth, including the protocol schedule for SX and answers to the self-report questionnaires.
- SX_respiban.txt: includes data from the RespiBAN device such as ECG, EDA, EMG, TEMP (°C), RESP, etc.
- SX_E4_Data.zip: includes data from the Empatica E4 device such as ACC, BVP, EDA, TEMP. This folder also has the following files:
  - ACC.csv: The 3 data columns refer to the three accelerometer channels.
  - BVP.csv: Data from a photoplethysmograph (PPG).
  - EDA.csv: Data is provided in μS.
  - TEMP.csv: Data is provided in °C.
- SX.pkl: includes synchronized data and labels.

A sliding window having a shift of 1 second was used for segmenting all sensor signals. Different modalities from the WESAD dataset were used for feature extraction and these features are displayed in Table I. These features are the subset of features described in [1].

Various statistical features, e.g., the standard deviation, mean, minimum, and maximum value, were computed on the raw ACC signal, for each axis separately (i ∈ x,y,z) along with summing up for all axes (3D) as absolute magnitudes. On raw ECG, BVP, RESP, and TEMP signals statistical features like the mean, standard deviation, minimum, and maximum value were computed. Besides, the peak frequency of BVP

TABLE II
LIST OF EXTRACTED FEATURES FROM WESAD DATASET.

Modality	Features	Description
ACC	$\mu_{ACC,i}$, $\sigma_{ACC,i}$ $min_{ACC,i}$, $max_{ACC,i}$ $i \in \{x, y, z, 3D\}$	Mean, standard deviation, minimum and maximum value for each axis separately and summed over all axes
ECG	μ_{ECG}, σ_{ECG} min_{ECG}, max_{ECG}	Mean, standard deviation, minimum and maximum value of the ECG
BVP	μ_{BVP}, σ_{BVP} min_{BVP}, max_{BVP}, f^{peak}_{BVP}	Mean, standard deviation, minimum, maximum and peak frequency of the BVP
EDA	μ_i, σ_i, min_i, max_i $i \in \{EDA, phasic, tonic, smna\}$	Mean, standard deviation, minimum, maximum value of the EDA signal, SCR/SCL and sparse SMNA driver of phasic component
EMG	μ_{EMG}, σ_{EMG} min_{EMG}, max_{EMG}, f^{peak}_{EMG}	Mean, standard deviation, minimum, maximum and peak frequency of the EMG
RESP	μ_{RESP}, σ_{RESP} min_{RESP}, max_{RESP}	Mean, standard deviation, minimum and maximum value of the RESP
TEMP	μ_{TEMP}, σ_{TEMP} min_{TEMP}, max_{TEMP}, δ_{TEMP}	Mean, standard deviation, minimum, maximum and slope of the TEMP

and slope of the signal δ_{TEMP} are used as features. The raw EDA signal was passed through a low pass filter with an order of 5 Hz, and then statistical features like standard deviation, mean, minimum, and maximum value were computed [1].

For calculating the statistical features mentioned above, let's suppose $W_i = \{x_1, x_2, ..., x_n\}$ is the window of raw data with x_1, x_2, etc., as n data points or measured signal values for a window of 1 second. The proposed work calculates the statistical features like mean, standard deviation, min and max for a particular window frame by using following mathematical expressions (1), (2), (3) and (4) respectively.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2} \tag{2}$$

$$min = u, \text{ where u is minimum among } x_1, x_2, ..., x_n \tag{3}$$

$$max = v, \text{ where v is maximum among } x_1, x_2, ..., x_n \tag{4}$$

Additionally, as the raw EDA signal consists of a phasic (skin conductance response (SCR)) and a tonic (referred to as skin conductance level (SCL)) components, they were separated. Once the SCL and SCR components were separated,

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.

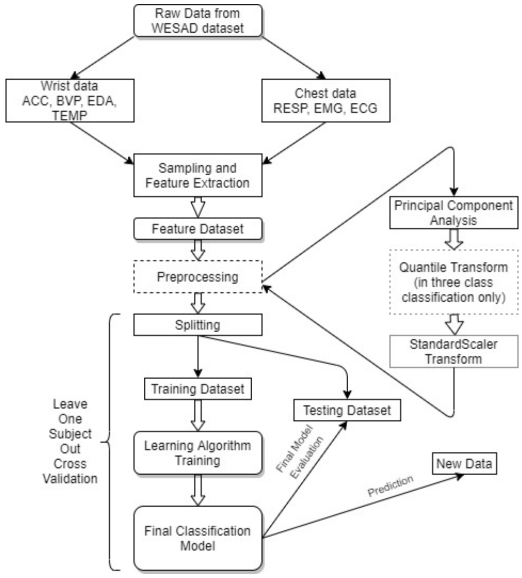

Fig. 1. Schematic flow diagram of Stress Detection Methodology.

```
Layer (type)            Output Shape  Param #
=============================================
dense_4 (Dense)         (None, 12)    252
_____________________________________________
dropout_2 (Dropout)     (None, 12)    0
_____________________________________________
dense_5 (Dense)         (None, 6)     78
_____________________________________________
dense_6 (Dense)         (None, 1)     7
=============================================
Total params: 337
Trainable params: 337
Non-trainable params: 0
```

```
Layer (type)            Output Shape  Param #
=============================================
dense_4 (Dense)         (None, 12)    252
_____________________________________________
dense_5 (Dense)         (None, 6)     78
_____________________________________________
dense_6 (Dense)         (None, 3)     21
=============================================
Total params: 351
Trainable params: 351
Non-trainable params: 0
```

Fig. 2. Summaries of deep learning architecture for binary (left) and three-class (right) classification.

the same features were computed as EDA. Also, sparse sudomotor nerve activity (SMNA) driver of the phasic component was derived, and the same features as EDA were computed by using a convex optimization approach to electrodermal activity processing (cvxEDA) [16]. In order to remove the DC component, the raw EMG signal was passed through a high pass filter with an order of 5 Hz. Then statistical features like standard deviation, mean, minimum, maximum, and peak frequency value were computed.

B. Preprocessings and Classification Algorithms

Six machine learning (Random Forest, Decision Tree, AdaBoost, k-Nearest Neighbour, Linear Discriminant Analysis and Kernel Support Vector Machine) and a deep learning artificial neural network (ANN) were used and their performance was compared. The extracted features, detailed above, are preprocessed according to suitability to the classification algorithms. Two types of classifications are used-three class and binary classification. Three class classification is defined as classifying an individual as amused, normal or stressed state, whereas, binary classification is defined as classifying an individual as either stressed or unstressed. Fig. 1 shows the schematic flow diagram of stress detection methodology.

For three-class classification problem both by all machine learning and deep learning algorithm, firstly, Principal Component Analysis (PCA) was applied with the number of components as 20 and full svd solver. To the data generated, Quantile Transformer method was used which transforms the features to follow an uniform or a normal distribution. Therefore, for a given feature, this transformation tends to spread out the most frequent values and reduces the impact of outliers. Then, Standard Scalar preprocessing was used which standardizes features by removing the mean and scaling to unit variance. For binary classification problem both by all machine learning and deep learning algorithm, firstly, Principal Component Analysis was applied with the number of components as 20 and full svd solver. Then, Standard Scalar preprocessing was used on the data generated after applying PCA as shown in Fig. 1. This research work has used Python's sci-kit learn implementation of the aforementioned machine learning classifiers and neural network library Keras for deep learning implementation.

For Decision Tree and Random Forest classifiers, 10 was set as the minimum number of samples for splitting a node, and maximum depth was set to 4 and 9, respectively, in three-class classification. In contrast, maximum depth was set to default (nodes are expanded until all leaves are pure or until all leaves contain less than the number of samples for splitting a node) values in case of binary classification. AB ensemble learner used Decision Tree as its base estimator, whose minimum number of samples for splitting a node was set as 5 and 10 in three-class and binary classification, respectively. In the k-Nearest Neighbour algorithm, the number of neighbors was set to 9 in both classification tasks. Moreover, a Support Vector Machine classifier was used for classification with *RBF (radial basis function)* kernel.

A simple neural network model was built for three-class and binary classification tasks with input, two hidden, and output layers. A dropout of 0.25 was added between two hidden layers in binary classification. In binary classification architecture, output has a single node with *sigmoid* as the activation function. In contrast, output has three nodes with a *softmax* activation function in the case of the three-class classification model. The summaries of both these models are as shown in Fig. 2.

Ground truths in the dataset are for three classes, namely amusement, baseline and, stress encoded as labels 0, 1, and 2 respectively. For using all the binary classification algorithms, classes amusement (label 0) and baseline (label 1) are combined to form a new class named non-stress (labeled as 0), and stress is another class (label changed to 1). F1-score (with macro averaging) and accuracy are used as evaluation metrics from sci-kit learn's metrics. During the study protocol, various conditions were carried out at different lengths, which makes a task that uses WESAD an unbalanced classification task and hence it is recommended to apply F1-score. The manner in which humans interpret and react to affective stimuli is very subject dependent. Personalization, therefore, becomes an

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.

important issue. Therefore, the *leave-one-subject-out (LOSO)* cross-validation procedure was used for evaluation of all the models, and the final accuracy is reported as the mean of all the testing accuracies when one subject is left for testing in LOSO cross-validation and trained on all other subjects. This makes the model generalize and perform better on previously unseen subjects' data making it subject independent.

IV. Results and discussion

The proposed work has recognized two classification tasks on the basis of the emotional states of a person for the detection of stress. First, a three-class classification task was defined: amusement vs. baseline vs. stress. Second, the amusement and baseline states were combined to non-stress class, and a binary classification task was defined: stress vs. non-stress. Table II shows the performance of given classifiers on both of these classifications.

When all the classifiers mentioned above are employed and their performance is compared, it becomes apparent from Table II that the DT classifier reached the lowest classification accuracy in both three-class and binary classification problems. Provided all the features as mentioned above and using the machine learning classifiers, the accuracy has reached up to 81.65% and up to 93.20% in the case of three-class and binary classification problems, respectively. Also, by using deep learning's simple artificial neural network classifier, accuracy has been reached up to 84.32% and up to 95.21% in the case of three-class and binary classification problems, respectively. Concluding from Table II, the DT had the overall worst performance, whereas kernel SVM had the best performance among all machine learning classifiers, and ANN gives the overall best performance among all classifiers. These results are better than those in the work of Philip Schmidt et al. [1], who presented accuracies of 80.34% on the three-class and 93.12% on the binary classification tasks.

TABLE III
Performance of all the classifiers in the three-class and binary classification tasks.

Techniques	three-class		binary	
	F1-score	Accuracy	F1-score	Accuracy
DT	53.73	68.16	84.92	87.59
RF	63.09	75.95	88.32	89.53
AB	67.55	78.19	89.88	91.06
LDA	64.66	74.83	87.60	90.15
kNN	66.76	74.71	84.63	87.92
SVM	73.57	81.65	92.31	93.20
ANN	78.71	84.32	94.24	95.21

V. Conclusion

The proposed research work has understood the structure and format of the publicly available WESAD dataset, cleaned and transformed data to a set eligible to construct machine learning and deep learning classification methods, explored and constructed various classification models and compared them. WESAD dataset contains data from multiple physiological modalities like three-axis acceleration (ACC), respiration (RESP), electrodermal activity (EDA), electrocardiogram (ECG), body temperature (TEMP), electromyogram (EMG) and blood volume pulse (BVP) which is not available in other datasets, which makes this work suitable for the detection of stress in human being. This model has achieved the accuracy of 84.32% and 95.21% on a three-class and a binary classification problems. As there were lesser subjects, caution must be taken while interpreting these results. However, our results show that generalization is possible as the LOSO evaluation scheme is used.

Further work can be done by taking self-reports of the subjects from the dataset into account, which were obtained using several organized questionnaires. The modalities such as facial cues, logging information, audio/video recordings, FITBIT data, etc. that are used in various studies separately can be merged with physiological data, and a new dataset can be introduced. Such a dataset can be more precisely used for stress detection as it will contain nearly all the features necessary for stress induction in human beings.

References

[1] Philip Schmidt, A. Reiss, R. Duerichen, Kristof Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable Stress and Affect Detection", International Conference on Multimodal Interaction 2018.
[2] Enrique Garcia-Ceja, Venet Osmani and Oscar Mayora, "Automatic Stress Detection in working environments from smartphones' accelerometer data: A First Step" , arXiv:1510.04221v1 [cs.HC] 14 Oct 2015.
[3] Saskia Koldijk , Mark A. Neerincx, and Wessel Kraaij, "Detecting Work Stress in offices by combining unobtrusive sensors", IEEE Transactions on affective computing 2018.
[4] Virginia Sandulescu, Sally Andrews, David Ellis, "Stress Detection using wearable physiological sensors", Springer International Publishing Switzerland 2015.
[5] Md Fahim Rizwan, Rayed Farhad, Farhan Mashuk, "Design of a biosignal based Stress Detection System using machine learning techniques", 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST).
[6] Saskia Koldijk, Maya Sappelli, Suzan Verberne, "The SWELL Knowledge Work dataset for stess and user modeling research", ICMI 2014.
[7] Jacqueline Wijsman, Bernard Grundlehner, Hao Liu, "Towards Mental Stress Detection using wearable physiological sensors", IEEE 2011.
[8] G. Giannakakisa, M.Pediaditisa, D.Manousos, "Stress and anxiety detection using facial cues from videos", Elsevier 2016.
[9] M. Gjoreski, H. Gjoreski, and M. Gams., "Continuous stress detection using a wrist device: In laboratory and real life." in UbiComp '16.
[10] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening", IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 12 (2008), 2067–2083.
[11] K. Plarre, A. Raij, and M. Scott. "Continuous inference of psychological stress from sensory measurements collected in the natural environment", in 10th International Conference on Information Processing in Sensor Networks (IPSN). 97–108.
[12] Alexandros Zenonos, Aftab Khan, and Mahesh Sooriyabandara, "HealthyOffice: Mood recognition at work using smartphones and wearable sensors", in PerCom Workshops 2016.
[13] B. Padmaja, V. V. Rama Prasad and K. V. N. Sunitha, "A Machine Learning approach for stress detection using a wireless physical activity tracker", International Journal of Machine Learning and Computing, Vol. 8, No. 1, February 2018.
[14] Global Organization for Stress on stress facts. http://www.gostress.com/stress-facts. Accessed: 2020-27-02.
[15] HSE on Work-related stress, anxiety or depression statistics in Great Britain, 2019. https://www.hse.gov.uk/statistics/causdis/stress.pdf. Accessed: 2020-27-02.

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.

[16] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi, IEEE"cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing", IEEE transactions on biomedical engineering, vol. 63, no. 4, april 2016.
[17] S. Palmer (1989). Occupational stress. The Health and Safety Practitioner, 7, (8), 16-18.

Authorized licensed use limited to: University of Liverpool. Downloaded on September 06,2020 at 04:35:54 UTC from IEEE Xplore. Restrictions apply.