LEVERAGING SENSOR DATA TO CLASSIFY STRESS

Submitted by:

APARNA KASIVISWANATHAN

Supervisor:

DR. AHMED H. ZAHRAN

Second Reader:

DR. ROSANE MINGHIM

MSc Data Science and Analytics

School of Computer Science & Information Technology
University College, Cork

September 2, 2022

Abstract

Stress is a major issue faced by all human beings at some point in their life. Thus, it becomes important to identify and manage stress and related issues.

This research project uses a publicly available stress-related dataset to classify the state of being according to measurements from various sensors. Using various machine learning algorithms like decision trees, random forests, K nearest neighbor algorithm, and so on, one of the goals is to find the best model that can classify the state of being of an individual based on measurements from sensors and quote the most important variables that have been used in the model with the best performance.

The paper is divided into sections that give details about background, dataset, preprocessing, feature extraction, model training, model evaluation, challenges faced and others.

Dataset preprocessing becomes an important part of this research paper as the dataset is enormous and the frequencies of the sensors vary to a great extent which makes the proportion of measurements for various sensors very different from each other. The methodology used in preprocessing the dataset also becomes an important milestone after which machine learning models are implemented. Several statistical measures are extracted from the raw data in order to create a dataset that can be used to train and test statistical models.

The classification problem is divided into 2 types: a three-class classification problem and a five-class classification problem. Models are trained on the data using all the features and the performance of the model is assessed. These models are then modified using dimensionality reduction techniques and scaling and results are obtained from these models as well.

Finally, suggestions have been made that involve exploring other techniques along with using other kinds of preprocessing which might be more efficient or less time-consuming.

Declaration

I confirm that, except where indicated through the proper use of citations and references, this is my original work and that I have not submitted it for any other course or degree.

Signed: ____________________

Aparna Kasiviswanathan
September 2, 2022

Acknowledgements

Without the assistance of numerous people, this thesis project would not have been feasible. Many thanks to Dr. Ahmed H. Zahran, my supervisor, who guided me at each step of the project and assisted in clearing out the muddle. Also grateful to Dr. Rosane Minghim my second reader for taking her time and be my paper's second reader.

I am grateful to the University College Cork for giving me this opportunity. Finally, I would like to express my gratitude to my parents, and the many friends who supported and loved me throughout this protracted process.

Dedication

I would like to dedicate this paper to my friends and family who have supported me throughout the completion of this project and whose words of motivation inspired me to take up this topic.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Brief overview

This research paper uses the publicly available Wearable Stress and Affect Detection (WESAD) dataset from the UCI Machine Learning repository (Schmidt et al. 2018) in order to classify the mental state given a set of parameters. All machine learning techniques used are supervised ML algorithms as the label (state of being) is already given in the dataset.

The dataset has been preprocessed and required features have been extracted and compressed using various statistical measures. This paper tries to build an effective machine learning model that can perform well in classifying the states of being such as amusement, stress, meditative state, baseline state, and so on. The models are compared using metrics such as accuracy and F1 score in order to present a picture of which model works the best for this dataset and tries to achieve the thesis objective.

1.2 Thesis Statement

The purpose of this thesis paper is to use the available dataset to classify various human emotions and try to build an effective model that can be employed to effectively classify various labels. The project also dives into some parameters that play an important role in the classification of human emotions which significantly improves the performance in terms of accuracy and F1 score.

Two kinds of classifications have been presented in this paper, one is a five-class problem (baseline, transient, amusement, stress and meditation) and the other is a three class problem (baseline, amusement, stress). In addition to this, two kinds of window sizes have also been used to preprocess the dataset. Decision tree, random forest, and K nearest neighbour algorithm have been built on these datasets along with techniques such as Principal Component Analysis (PCA), scaling etc. and the best algorithm among these has been found.

1.3 Thesis Structure

This section explains in detail the structure of the thesis paper, i.e., how the paper is organized into chapters and the content inside each chapter.

There is an abstract of what the research paper is about followed by a detailed explanation of the overview of this research paper, a thesis statement, and a detailed structure of the thesis (here).

Chapter 2 is dedicated to the background of this paper. This chapter throws light on the term 'stress', its definition, the impact of stress on the human body, an organ in the human body that plays an important role when met with a stressor, and so on. The second subsection of this chapter illustrates the available datasets in this domain and other research papers analyzing and classifying stress and stress-related issues.

The next chapter titled 'Data Analysis' explains how the WESAD dataset was collected, and provides detailed information about all the folders and files contained, parameters that were measured, frequencies of each sensor, and so on. The chapter then goes on to explain the challenges faced with the dataset, followed by the preprocessing stage which consists information about the statistical measures used to preprocess the dataset, techniques used, and preparation of the final dataset used for all training purposes. This subsection ends with a feature extraction process explaining the features used for employing machine learning algorithms.

Chapter 4 named 'Models', gives information about the classification models used for training and testing the dataset, other techniques used with the model to enhance the model, followed by a detailed description of the comparison of models.

The final chapter, titled 'Conclusion' gives information about the results obtained after training all the different models and the scope for further work that could be done using the WESAD dataset.

Chapter 2

Background

2.1 Stress

One of the most common words used today is "stress". According to Healthline (*Everything to Know About Stress: Causes, Prevention, and More* n.d.), stress is a situation that triggers a particular biological response. When a human brain perceives a major challenge, threat, lack of control, or any other stressors, the human body releases hormones and chemicals to cope with the stressor(s).

Using the word 'stress' to describe the mildest challenging situations to severely problematic situations blurs the line between the medical term 'stress' and the meaning that individuals in society have attached to the word 'stress' (Koolhaas et al. 2011). Stress is the human body's response to pressure which can be triggered by different scenarios. It is often triggered when one experiences something new, unexpected that threatens one's sense of self, or when one feels that they have little control over a situation. One cannot always say that stress is bad. For example, stress is helpful for helping individuals meet deadlines. However, stress becomes an issue if it is triggered by something that might be as small as crossing the road.

In order to learn about stress, it is important to know the role of the nervous system in a human body. The sympathetic nervous system (or SNS) and the parasympathetic nervous (or PNS) system which is a part of the autonomic nervous system play a major role in physical reaction to stress. When a stressor is encountered and the body becomes stressed, the individual responds by switching to a flight or fight mode (McCorry 2007). Flight mode is the mode where the individuals body tries to escape or flee from the stress whereas fight mode is the mode where the individual's body tries to fight the stress. Later on, a freeze and fawn mode were also added to the bodily responses to stress. The sympathetic nervous system governs the flight, fight, freeze or fawn mode. The SNS causes the release of stress hormones like adrenaline and cortisol and makes the heart beat faster. Once the stressor disappears or the crisis is over, the body returns to a relaxed or unstressed state and this is facilitated by the parasympathetic nervous system. During this time the body feels safe and other processes like rest and digestion will happen (*Fight, Flight, or Freeze: How We Respond to Threats* n.d.).

According to the American Psychological Association (*Stress effects on the body* n.d.), the musculoskeletal system which refers to the muscles in the human body, respiratory system, cardiovascular health which refers to heart health, gastrointestinal, nervous, and reproductive systems are affected by stress.

In many datasets that are available in this domain, it can be found that there is at least one measurement that is related to the heart. This is because the heart and blood vessels work together to provide nourishment and oxygen to the body and hence the activity of the heart and blood vessels is also related to an individual's bodily response to stress (*Stress effects on the body* n.d.). For example, momentary stress or short term stress, known as acute stress such as meeting a deadline or being stuck in an elevator causes an increase in heart rate, and in the process it also releases hormones such as adrenaline and cortisol which are the messengers for these effects. Once the acute stress episode has passed, the body returns to its normal state. There is also another state of constantly being stressed which is known as chronic stress and it can be identified from the consistent increase in heart rate, and elevated blood pressure among others. This can take a serious toll on the body and thereby increasing the risk for hypertension, heart failure, heart attack, or even stroke (*What Is the Difference Between Chronic and Acute Stress? - AFC Urgent Care* n.d.).

Stress cannot be avoided completely as it is a part of life. However, it can be managed efficiently so that it doesn't harm the body. Modifications to the existing lifestyle can be a game changer when it comes to stress management. Exercising regularly, eating homemade food, and having good social and personal life contributes to reducing stress. Other factors such as environmental conditions, cleanliness, and satisfaction with the government are also major factors in determining the stress level of a person.

Research conducted on employees showed that the employees that are stressed showed reduced productivity and higher costs for healthcare ranging from depression to heart disease. According to Forbes (*Workplace Stress Responsible For Up To $190B In Annual U.S. Healthcare Costs* n.d.), the consequences of stress-related issues accounted for up to 190 billion US dollars a year for US businesses. In many countries, companies are investing in scientifically proven mediation techniques after realizing that meditation helps workers to manage emotions, changing the wiring of the brain slowly and leading to healthier work productivity. In a study conducted in 2016 (Monique Tello 2016), it was proved that meditation had a long-lasting effect on stress when compared to vacation. One of the states of being that the study tries to identify is the meditative state.

Headspace is a well-known app that provides guided, unguided, and other meditation courses and techniques. A study conducted in the year 2018 (*Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial | SpringerLink* n.d.) revealed that 8 weeks of meditation resulted in a forty six percent decrease in stress levels and thirty-one percent reduction in negative feelings. The study concluded with the saying: "Brief mindfulness training has a beneficial impact on several aspects of psychosocial well-being."

2.2 Stress Data Analysis

There is a lot of data available in the field of classification and identification of stress. In addition to this, nowadays it can be seen that people use wrist watches to track their heart rate, pulse, and other variables and this can be used to identify if an individual is stressed or not. The motivation behind this project is to use statistical analysis to explore different parameters that contribute in identifying the state of being of an individual and use machine learning models to classify the state of being. Some of the parameters used in this study to identify stress are recorded by placing sensors in the heart area as stress can immensely affect an individual's heart health. Different machine learning models have been employed and their performances have been captured to classify the different states of being.

2.2.1 SWELL Dataset

One of the datasets available in this field is the SWELL knowledge work (SWELL-KW) dataset (Koldijk et al. 2014). The study involved 25 subjects of which 8 were female and 17 were male. The average age of the participants was 25 with a standard deviation of 3.25. The participants were given standard knowledge tasks, and the dataset was gathered (writing reports, making presentations, reading e-mail, searching for information). The subjects' working environment was modified by adding stressors like time constraints and email interruptions. Some of the parameters measured include heart rate variability and skin conductance. In addition to this, computer logging, camera recordings of face expressions, Kinect 3D sensor data on body postures were also recorded. The dataset which was made accessible includes preprocessed data along with raw data. With the help of validated questionnaires, the participants' subjective perceptions of workload, mental effort, emotion, and perceived stress were evaluated. The resulting dataset on working behavior and affect makes a significant contribution to a number of study areas, including user modeling, context-aware technologies, and work psychology.

The experiment involved a neutral state where the subject worked on normal working tasks with a maximum duration of 45 minutes. After this, time pressure was introduced in order to manipulate the working conditions. In the final stage of the data collection process, another stressor was introduced which involved interruptions such as emails that sometimes required a reply and sometimes not, some that were relevant to the task and some that were not, etc.

The fully preprocessed data, aggregated every minute for all 25 participants have the following characteristics present in it: 12 features for computer interaction, 40 for facial expression, 88 for body posture, and 3 for physiology. The settings under which the data was gathered were marked in the feature dataset. An average of 45 minutes of working under normal circumstances, 45 minutes of working with email interruptions, and an average of 30 minutes of working under time constraints were included for each participant (Koldijk et al. 2014).

A research paper titled 'The SWELL Knowledge Work Dataset for Stress and User Modeling Research' (Koldijk et al. 2014) gives in-depth detailed descriptions of the data collection, preprocessing, feature engineering, and so on can be found in this research study.

2.2.2 OSMI Dataset

OSMI stands for Open Sourcing Mental Illness, an organisation promoting awareness about mental health specifically in tech (*Research :: Open Sourcing Mental Health - Changing how we talk about mental health in the tech community - Stronger Than Fear* n.d.). The OSMI data tell that people in the tech industry face more mental health problems than others (*Open Sourcing Mental Illness Wants to Break the Mental Health Stigma* n.d. There are survey datasets available on their website for various years.

The 2017 dataset had 750 responses from a variety of employees in different technology-related departments and the responses comprised personal as well as professional aspects in order to give a comprehensive view of the environment surrounding the individual. The dataset has 750 responses from various individuals and 68 features that represent their personal as well as professional lives.

The research paper titled "Machine Learning Techniques for Stress Prediction in Working Employees" (Reddy, Thota, and Dharun 2018) uses the 2017 survey dataset to determine whether a person can develop mental health issues or not. The research found that gender, family history, and provision of mental health benefits to employees by the company (or employer) had a huge impact in deciding the development of mental health issues. In addition to this, it also confirms that employees in the field of technology were more susceptible to stress than employees who did not work in the field of technology. The research paper suggests companies use the results to frame better HR policies.

2.2.3 Driver Stress Research

The paper titled "Machine Learning for Stress Detection from ECG Signals in Automobile Drivers" (Keshan, Parimi, and Bichindaritz 2015) uses the dataset related to drivers and stress involved while driving. The dataset was obtained from MIT-BIH PysioNet Multi-parameter Database (*PhysioBank Databases* n.d.). It has the data of 17 subjects who were drivers with parameters such as ECG, EMG, foot galvanic skin response, hand galvanic skin response, intermittent heart rate, marker, respiration and time stamp. Just like the WESAD dataset, all the data in the driver stress detection paper have also been collected using wearable sensors.

The dataset was compiled by Healey, and Picard 2005, and the subjects drove from MIT East Garage to River Street Bridge, 3 cities and 2 highways. The measurements involved observations of initial and final rest states and stress during driving. Following this, the dataset was split to low stress level, moderate stress level and high stress level. The low stress level was during the initial and final rest state, the moderate stress was while driving through highway and the high stress was while driving through the cities. One of the assumptions made here was that the stress is exclusively because of traffic conditions and no other personal or professional or any other reasons. The 'Machine Learning for Stress Detection from ECG Signals in Automobile Drivers' (Keshan, Parimi, and Bichindaritz 2015) uses 10 of the drivers' data. Label 0 has been taken as the rest state, label 1 as the moderate stress state and label 2 as the high stress state.

Various algorithms such as Naive Bayes, Logistic Regression, Multilayer Perceptron, Support-Vector Machine, 1-Nearest Neighbor, K-Nearest Neighbors, Decision Tree, Random Forest and others have been trained on the final dataset consisting of 67 rows and 14 parameters in the "Machine Learning for Stress Detection from ECG Signals in Automobile Drivers" (Keshan, Parimi, and Bichindaritz 2015) paper. In addition to this, various techniques like leave one out cross validation, 2 fold cross validation and 10 fold cross validation have also been used. Two kinds of splits have been used for the training and testing set. One is the 75 percent split where 75 % of the data is used of training and 25 % of the data is used for testing and the other split is the 90 % split where 90 % of the data is used for training and 10 % of the data is used for testing.

After all training and testing it was concluded that decision trees has the best performance in terms of accuracy when combined with the leave one out cross validation technique.

Chapter 3

Data Analysis

3.1 Dataset Explained

The dataset used for the purpose of this work is the publicly available WESAD data set from the UCI Machine Learning Repository (*UCI Machine Learning Repository: WESAD (Wearable Stress and Affect Detection) Data Set* n.d.). This dataset was introduced and was made available by Attila Reiss, Philip Schmidt, et al. in 2018 (Schmidt et al. 2018). The WESAD dataset is a multimodal dataset.

The dataset is organised in such a way that each subject has a folder name S2, S3, S4 and so on (SX where X = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17) comprised of the following files :

- SX_readme.txt : This consists of the information about the subject and information about data collection
- SX_quest.csv : This consists of answers given by the subjects to a questionnaire among other information
- SX_respiban.txt : This consists of the data from the chest worn RespiBAN device
- SX_E4_Data.zip : This consists of the data from the wrist worn Empatica E4 device
- SX.pkl : This is a pickle file containing information about synchronised data and labels

The dataset used for this study is the synchronized data in the pickle format which is essentially a nested dictionary. This nested dictionary contains information about the subject ID, signals which include all the raw data taken from the chest-worn device and the wrist-worn device, and label which denotes the label of the current condition which includes stress, amusement, transient, baseline, meditation and other. The dataset consists of physiological and motion data recorded from a chest-worn device and a wrist-worn device. The study involved 17 subjects but due to sensor malfunction, the data of 2 subjects are missing. This makes the total number of subjects whose data is available 15. Among these 15 subjects, 12 were male and 3 were female. The chest-worn device is RespiBAN Professional and the wrist-worn device is Empatica E4. The study was conducted on graduate students excluding students who were pregnant or had mental health issues or cardiovascular diseases or chronic diseases or were heavy smokers. The average age of the subjects was between 25 and 30 years.

The kinds of situations that the subjects had to go through were amusement, meditation, and stress in addition to the baseline condition. For the baseline condition, the data was recorded for 20 minutes while the subjects read a neutral magazine. The amusement condition was induced by making the participants watch funny video clips and had a total length of nearly 6 minutes. For the stressful situation, the participants were exposed to the well-known Trier Social Stress Test which had a total duration of about 10 minutes excluding the rest period. After this, participants went through a meditative state where they were given audio with instructions for the meditation and this lasted for about 7 minutes. In addition to this, subjects were also given questionnaires to fill. To summarise, all the procedures took about 2 hours to get completed.

Measurements from the different sensors are taken and a label is attached to each set of measurements.

Variables explained (as given in the readme.pdf file of the dataset):

Wrist worn device data variables :

- TEMP - Sampled at 4 Hz. Data from temperature sensor expressed degrees on the Celsius (°C) scale.
- EDA - Sampled at 4 Hz. Data from the electrodermal activity sensor expressed as microsiemens
- BVP - Sampled at 64 Hz. Data from photoplethysmograph.
- ACC - Sampled at 32 Hz. Data from 3-axis accelerometer sensor.

Chest worn device data variables :

- ECG (mV): Electrocardiogram measurement
- EDA : Electrodermal activity measurement
- EMG (mV): Electromyography measurement
- TEMP (°C): Temperature
- RESPIRATION

All the chest-worn device measurements which include ECG, EMG, EDA, TEMP, and RESP were sampled at 700 Hz. In the case of the wrist-worn device, BVP was sampled at 64 Hz, three axis ACC measurements were taken at 32 Hz frequency and EDA and temperature were measured at 4 Hz. Due to this, the data for chest-worn devices is much larger than the data for wrist-worn devices. Labels are also taken at the frequency of the fastest sensor, i.e., 700 Hz. Labels are represented with the help of numbers. 0 means that the condition is not defined or transient, 1 is used to represent the baseline condition. Label 2 is attached to the stress condition, label 3 refers to the amusement state and label 4 is used to represent the meditation state. There are also 3 other labels namely 5,6,7 which are ignored during the data analysis and model-building process.

The .pkl file which has been used for data analysis and model training purposes has been parsed using Python using the joblib and pickle packages. As per the Joblib module documentation (*Joblib: running Python functions as pipeline jobs â joblib 1.2.0.dev0 documentation* n.d.): "Joblib is a set of tools to provide lightweight pipelining in Python." The pickle package is primarily used for serialising and de-serializing a Python object. During the pickling process, the Python object hierarchy is converted to a byte stream and during the unpickling process the reverse happens, i.e., a byte stream is converted back into an object hierarchy (*Python pickling: What it is and how to use it securely | Synopsys* n.d.). The python object which is stored in the form of a pickle file in the case of the WESAD dataset is a nested dictionary. The dictionary is parsed using for loop and each measurement is stored in a list object or data frame object.

3.2 Challenges Faced

One of the very first challenges confronted was the high volume of the dataset. Due to this, it took a lot of time for parsing the nested dictionary and extracting the features. This was tackled successfully by using lists, for loops, and packages available in python. After this, it was found that the number of observations for chest-worn device data was extremely higher than that of the wrist-worn device data. This was due to the difference in sampling frequencies for the various devices. In order to tackle this problem upsampling or downsampling technique has to be put in use. More about dealing with challenges has been explained in the data preprocessing and feature extraction section.

One of the other challenges faced was the imbalance in the new dataset that was created after dealing with the difference in sampling frequencies of chest-worn and wrist-worn device. In the process of exploratory data analysis, it was found that the dominant label was 0. This has been visually presented in the count chart created using the Seaborn package in Python (Figure 3.1).

From Figure 3.1, it is evident that 0 is the dominant category followed by 1, 4, 2, and 3. This is often found in many real-world datasets. If the imbalance is not checked for in the beginning it could affect the interpretation of accuracy obtained. In such cases, the baseline classification accuracy would be higher than the baseline classification accuracy of a balanced dataset. Baseline classification accuracy refers to the accuracy obtained when all the predictions are the same as the most frequent class label. It is calculated as follows :

$$\frac{\text{Count of the most frequent class}}{\text{Total number of entries}}$$

Suppose one obtains an accuracy of 85% but the baseline accuracy is 85% as well then it cannot be said that the model is working as the minimum accuracy for the model

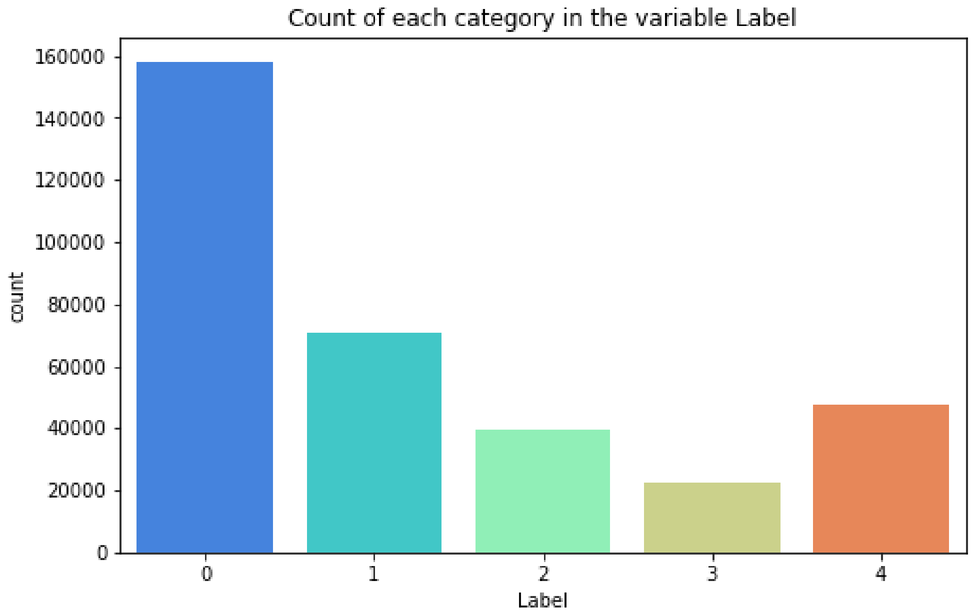

Figure 3.1: Count Plot of Label Variable

will be 85%. Usually, imbalance in categories does not greatly affect the model but if the imbalance is quite high then certain techniques have to be employed in order to cope with the high imbalance. One of such methods is to use weights. One could apply more weight to classes that are not dominant and less weight to the dominant class so that the representation becomes balanced. However, in this project, as the imbalance is not quite problematic weights have not been applied to all the classifiers.

It is given in the dataset readme file that there were 12 males and 3 females who participated in the study but the information about which data is related to males and which is related to females could not be found anywhere. Hence, the possibility of implementing gender-based machine learning algorithms was ruled out. Gender could have had an impact on the dataset as it is well-known that adult men and women have different normal ranges for various parameters. For example, the average heart rate of adult men is in the range of 70 and 72 beats per minute whereas the average heart rate for adult women is between 78 and 82 beats. The difference is caused by the size of the heart as the size of the heart of a woman is typically smaller than that of a man (Prabhavathi et al. 2014) . This could induce an inconsistency in the results obtained if new data on a large number of women are introduced.

Finally, the age group of the participants is very narrow and this could pose a problem while predicting the emotional state of people who belong to an enitrely different age group (say 70-75) because age also has an effect on stress (Folkman et al. 1987).

3.3 Preprocesing and Feature Extraction

The dataset in total is about 17 GB and the features have been extracted from the synchronised data. As the variables have been sampled at different frequencies, the number of observations vary for each sensor. For example for subject S1, there are 4,255,300 measurements for chest-worn device variables like ECG, EMG, EDA and so on, 194,528 wrist-worn ACC samples across 3 axes, 389,056 samples for BVP measurement and 24,316 measurements for wrist-worn EDA and temperature. In addition to this the number of measurements for chest-worn data (or wrist-worn data) is not the same for every subject, i.e., subject S2 has 4,545,100 chest-worn device measurements whereas subject S3 has 4,496,100 measurements for chest-worn device data. Same applies for wrist-worn device measurements. For example subject S2 has 207,776 samples for wrist ACC across 3 axes, 415,552 for wrist BVP and 25,972 for wrist EDA and temperature whereas subject S3 has 205,536 measurements for ACC, 411,072 measurements for wrist BVP and 25,972 measurements for wrist EDA and temperature. In addition to this, there are other files which have not been used for the purpose of this research paper.

In such a case, the data cannot be combined to a single data frame directly. Either the measurements of the slow sensors have to be replicated to match the number of records for the fastest sensor or the measurements of the sensors that are faster than the slowest sensor have to be downsampled to match the slowest sensor. In this case, the chest-worn device is the fastest sensor with a frequency of 700 Hz and the slowest sensors are the sensors that are used to measure the wrist EDA and wrist temperature having a frequency of 4 Hz. The way in which the slowest signals are replicated in order to match the frequency of the fastest sensor is upsampling and the way in which all sensors are shrunk in order to match the frequency of the slowest sensor is downsampling (*Downsampling and Upsampling of Images â Demystifying the Theory | by Aashish Chaubey | Analytics Vidhya | Medium* n.d.).

For the purpose of this study, the technique that has been used is downsampling. A window size of 0.25 seconds has been applied to all the data from all sensors other than the sensors having a frequency of 4 Hz. For the chest-related data, for each subject, the measurements have been grouped with each group having 175 samples in it and statistical measurements have been applied to this group to combine it into a single number. This is done in order to have a matching number of records for chest-worn device data and EDA and temperature. In a similar way, wrist ACC measurements are grouped in such a way that each group has 16 samples in it. Wrist BVP measurements have been grouped so that each group has 32 samples in it. After all this preprocessing of the raw data, a combined dataset comprising of data from both chest-worn devices and wrist-worn devices is obtained for each individual.

In regards to the label, the same technique that was used to downsample chest-worn device data has been used. In order to get one single value, the mode has been used. Statistically, the mode is that value that occurs most frequently (*Mode Definition* n.d.). So, if in a categorical variable 1 occurs 12 times and 2 occurs 13 times then 2 would be the mode. In this case, both the category as essentially dominant but due to the property of mode, 2 is chosen as the mode. This could potentially be a problem while summarising the labels. This was checked and it was found that in each group the majority of labels belong to a single category (for example, a group might have 90% of labels as 0 and 5 percent as 1 and the rest divided among 2,3 and 4). Therefore, the mode is used to summarise the data.

In order to create a data frame consisting of all the records of all subjects, the dataset of each individual is combined to create a new dataset. From this new dataset, label 5, 6, and 7 have been removed as they have been asked to be ignored in the readme.pdf file of the WESAD dataset. This significantly reduced the size of the dataset but not to an extent where there is not enough data for training and testing. This new data set is used for all data analysis, model training, and other purposes.

The column Label, which is the target variable was of type int64 and was converted to categorical type.

The dataset obtained after combing all preprocessed raw data of individual records has 337,860 rows and 52 columns of which one is Label (target column).

In addition to this dataset, another dataset was also created using a window size of 1 second with everything else remaining the same. This dataset has 84465 rows and 57 columns (including the target). The reason for more columns is because Wrist Temperature and Wrist EDA had to be compressed as the window size is large and mean, standard deviation, minimum and maximum were used to compress the 2 columns.

The statistical measures that have been used are:

- Mean
- Standard Deviation
- Minimum
- Maximum

Below is a table giving details about the basic statistical measures like mean and standard deivation of the extracted features using window size of 0.25 seconds.

Variable	Mean	Std Dev
Mean Chest ACC1	0.8098	0.1309
Mean Chest ACC2	−0.0441	0.1036
Mean Chest ACC3	−0.2649	0.3313
Mean Chest EDA	4.8792	3.5197
Mean Chest EMG	−0.0030	0.0015
Mean Chest Temp	33.9014	1.1942
Mean Chest Resp	0.0523	4.0590
Mean Chest ECG	0.0010	0.067
Mean Wrist ACC1	11.7089	43.9294
Mean Wrist ACC2	−2.1229	28.1881
Mean Wrist ACC3	17.5164	29.5609
Mean Wrist BVP	0.001509	55.706287
SD Chest ACC1	0.008032	0.017411
SD Chest ACC2	0.007367	0.012667
SD Chest ACC3	0.011705	0.017596
SD Chest EDA	0.010020	0.009226
SD Chest EMG	0.014154	0.011113
SD Chest Temp	0.025100	0.264302
SD Chest Resp	0.339756	0.324870
SD Chest ECG	0.189996	0.178646
SD Wrist ACC1	1.555208	3.581251
SD Wrist ACC2	0.982580	2.388348
SD Wrist ACC3	1.445972	3.420960
SD Wrist BVP	18.984678	30.145167
Min Chest ACC1	0.793654	0.130685
Min Chest ACC2	-0.058995	0.104980
Min Chest ACC3	-0.291882	0.328287
Min Chest EDA	4.853411	3.519839
Min Chest EMG	-0.046199	0.038140
Min Chest Temp	33.832615	1.267618
Min Chest Resp	-0.744402	3.920364
Min Chest ECG	-0.260904	0.222661
Min Wrist ACC1	9.547031	44.412052
Min Wrist ACC2	-3.447961	28.398072
Min Wrist ACC3	15.561647	30.168274
Min Wrist BVP	-28.733465	71.254006

Variable	Mean	Std Dev
Max Chest ACC1	0.826814	0.138306
Max Chest ACC2	-0.029001	0.106707
Max Chest ACC3	-0.237225	0.340390
Max Chest EDA	4.923049	3.509973
Max Chest EMG	0.037246	0.038468
Max Chest Temp	33.986307	1.159236
Max Chest Resp	0.634512	4.139568
Max Chest ECG	0.556699	0.616213
Max Wrist ACC1	13.892867	44.158896
Max Wrist ACC2	-0.801243	28.422393
Max Wrist ACC3	19.555695	29.782934
Max Wrist BVP	27.399611	63.883372
Wrist EDA	1.804678	2.348282
Wrist Temp	32.594447	1.501841

Table 3.1: Mean and Standard Deviation of Features Extracted

Chapter 4

Models and Training

4.1 Models

4.1.1 Decision Tree

According to IBM, 'A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.' (*What is a Decision Tree | IBM* n.d.)

Just as the name says, the decision tree is basically a tree with branches, leaves and nodes. The start of a decision tree is called the root node. This root node has only outgoing branches and does not contain any incoming branches. The nodes following the outgoing branches are called decision nodes. The end nodes are called as terminal nodes or lead nodes. Leaf nodes only contain incoming branches and does not contain any outgoing branches.

The decision regarding splitting is mainly influenced by the information gain and entropy (*Entropy and Information Gain. Yet another tool used to make Decisionâ¦ | by Steven Loaiza | Towards Data Science* n.d.). Information gain, as the name suggests gives the information that a feature in the dataset provides about the labels/class and the aim is to try to maximise the information gain. The decision tree algorithm chooses the parameter with highest information gain for splitting. Entropy, on the other hand is the measure of impurity and it is used to calculate how well a node can split the data correctly.

In Python, one of the ways to implement a decision tree is by using the sklearn library. It can be used in program using the command 'from sklearn.tree import DecisionTreeClassifier' (*sklearn.tree.DecisionTreeClassifier â scikit-learn 1.1.2 documentation* n.d.).

This is accompanied by several parameters which can be used along with the classifier in order to improve the performance. Some of these parameters include the criterion to measure the quality of a split such as Gini impurity, the strategy to choose split, maximum depth of the decision tree, the minimum samples to split a node, minimum number of samples for a leaf node, random state if the results need to be replicated and many others.

4.1.2 Random Forest

Random forest is another supervised machine learning algorithm that can be used for classification as well as regression (*Random Forest | Introduction to Random Forest Algorithm* n.d.). Random forest is an ensemble ML algorithm, i.e., it combines several base models (decision trees) to create a final optimal model. It creates decision trees from various samples, relying on their majority for categorization and average for regression. Therefore, in general random forests give slightly better accuracy when compared to decision trees.

In Python, one of the ways to employ random forests is to use sklearn library (*sklearn.ensemble.RandomForestClassifier â scikit-learn 1.1.2 documentation* n.d.). The classifier can be put to use by importing it using the command from 'sklearn.ensemble import RandomForestClassifier'. It has various parameters that can be specified if the default parameters have to be modified. These include (but not limited to) maximum tree depth, criterion (just like decision tree), minimum sample split, maximum leaf nodes and so on.

4.1.3 K Nearest Neighbour

This is also a supervised distance based ML algorithm that can be used for both classification as well as regression tasks. KNN algorithm finds the K number of nearest neighbours to the data point and assigns the value to the new data point accordingly. It is greatly influenced by value of K (*What is the k-nearest neighbors algorithm? | IBM* n.d.). As a rule of thumb, K is chosen to be an odd number (for 2 class problem) because if K is even it could cause problems while assigning the label to the new data point (*KNN Classification Tutorial using Sklearn Python | DataCamp* n.d.).

In Python, one of the ways to implement a K Nearest Neighbour algorithm is by using the sklearn library (*sklearn.neighbors.KNeighborsClassifier â scikit-learn 1.1.2 documentation* n.d.). It can be used in program using the command 'from sklearn.neighbors import KNeighborsClassifer'. This comes with several parameters such as the number of neighbours, weights, algorithm, leaf size, power parameter, metric to use for distance computation and so on. Weights parameter is extremely useful when dealing with imbalanced datasets. If weight is set to 'distance', the algorithm will work in such a way that the nearby neighbours will have larger impact when compared to the neighbours that are not very close.

4.2 Implementation

The models that have been used in this research study are random forests, decision trees and K Nearest Neighbours. Other techniques like Principal Component Analysis, scaling, adding of weights have also been used to observe the impact on the classifiers. An important fact that is to be noted is that scaling has not been used for decision trees and random forest because scaling has less/no impact on the results from tree based models as they are not affected by the variance present in the dataset (*Do Decision Trees need Feature Scaling? | by Praveen Thenraj | Towards Data Science* n.d.). However, scaling does have a huge impact on a distance based model like KNN. For all the models, 70 percent of the data has been used for training and the rest has been used for testing. It is important to note the value of F1 score as it is imbalance dataset, thus F1 score is to be noted (*The F1 score | Towards Data Science* n.d.).

4.2.1 Model 1 : Decision Tree

Window Size 0.25 Seconds

Firstly, the model was trained on the dataset that was preprocessed using a window size of 0.25 seconds. For the 5 class classification problem an accuracy of 97.9% was obtained and an F1 score of 96.1% was obtained. The most important features were Wrist Temperature having feature importance of 12%, Wrist EDA with feature importance of 11%, Mean Chest EDA with an importance of 10% followed by Mean Chest Temperature and Max Chest ACC (measured across the third axis) having feature importance of 9% and 8% respectively.

After this, principal component analysis was applied to the decision tree classifier so that it can handle any correlations and overcome the overfitting problem of decision tree classifier. Using 10 components an accuracy of 90.2% was obtained along with an F1 score of 90.65%.

Window Size 1 Second

For the 5 class classification problem for the data with window size of 1 second, an accuracy of 96.29% was obtained and F1 score of 90.65% was obtained as well. The most important features were Standard Deviation of Chest EMG (11.6%), Mean Chest EDA (8%), Maximum Chest ACC (7%), Maximum Wrist Temperature (6%) and Minimum Wrist Temperature (6%).

For the 3 class classification problem (amusement, stress and baseline) of dataset preprocessed with 1 second window size an accuracy of 97.27% and F1 score of 97.24% was obtained. Max Chest ACC, Min Chest ACC, Mean Chest EDA, Mean Wrist Temperature, Mean Chest Temperature were the features with highest values of features importances of 9.7%, 8.5%, 7.7%, 7.4% and 6.3% respectively.

4.2.2 Model 2 : Random Forest

Window Size 0.25 Seconds

The next classifier that is performing slightly better than the decision tree classifier is the random forest classier. For the dataset that was preprocessed using 0.25 seconds window size an accuracy of about 99.2% for the 5 class classification problem was obtained and took approximately 145 seconds to train. The F1 score was approximately 99.25% . As in the decision tree classifier, no scaling was used here either as it has no effect on the performance of the random forest classifier as well. The most important features used in this classifier were Wrist Temperature, Wrist EDA, Mean Chest EDA, Mean Chest Temperature and Minimum Chest EDA with importance of 7.52%, 6.08%, 5.12%, 4.56%, and 4.49% repsectively. After this, PCA (Principal component Analysis) was applied to this classifier as well. The number of components used was 10. The accuracy was about 93.27% and F1 score was about 92.263%.

Window Size 1 Second

For 5 class classification problem for the dataset with window size of 1 second the accuracy was 98.7 percent and F1 score was 98.8%. The most important features were Standard Deviation Chest EMG, Mean Wrist Temperature, Minimum Wrist Temperature, Maximum Chest ACC, Maximum Wrist Temperature and for the 3 class classification problem the accuracy was 99.2% and F1-Score was 98.99%. The most important features were the Standard Deviation of Chest ACC, Mean Wrist Temperature, Minimum Wrist EDA, Mean Chest Temperature and Maximum Chest ACC with features importances of 18.9%, 9.4%,9.06%,9% and 4% respectively.

4.2.3 Model 3 : K Nearest Neighbour

Window Size 0.25 Seconds

Without scaling, the KNN algorithm did not perform as well as the other classifiers and had an accuracy of about 82% and F1 score of 79.2%. However, after scaling, its accuracy improved to a great extent and went up to 96.12% and F1-Score of 96.10% for the 5 class classification problem with dataset having window size of 0.25 seconds. When KNN was trained with weights as well as scaling the accuracy improved again and was approximately 96.9% with F1 score of 96.1%.

Window Size 1 Second

For the dataset with window size of 1 second and 5 class classification, the accuracy was 82.73% and F1 score was 82.65%. KNN performed well with scaling giving an accuracy of 95.3% and F1 score of 95.5%. After applying weights and scaling, the accuracy went upto 95.86% and F1 score was about 95.97%.

For the 3 class classification problem and dataset having window size of 1 second without scaling KNN had an accuracy of 91.303% and F1 score of 91.29% and with scaling the accuracy improved to 97.9% with F1 score of 97.2% After applying scaling and weights the accuracy significantly improved and went to 98.5% with F1 score of 98.2%.

4.3 Comparing Models

A summary of the results obtained are presented below in the tables.

Algorithm	Accuracy	F1 Score
Decision Tree	97.9%	96.1%
Decision Tree and PCA	90.2%	90.65%
Random Forest	99.2%	99.25%
Random Forest and PCA	93.27%	92.26%
KNN	82%	79.2%
KNN with Scaling	96.12%	96.10%
KNN with Scaling and Weights	96.9%	96.1%

Figure 4.1: Accuracy and F1 Score for Dataset with Window Size of 0.25 seconds

Machine Learning Model	5 Class Classification		3 Class Classification	
	Accuracy	F1 Score	Accuracy	F1 Score
Decision Tree	96.29%	90.65%	97.27%	97.24%
Random Forest	98.7%	98.8%	99.2%	98.99%
KNN	82.73%	82.65%	91.303%	91.29%
KNN with Scaling	95.3%	95.5%	97.9%	97.2%
KNN with Scaling and Weights	95.86%	95.97%	98.5%	98.2%

Figure 4.2: Accuracy and F1 Score for Dataset with Window Size of 1 second

All the algorithms have better performance for the dataset that has been preprocessed using a window size of 0.25 seconds. This is because when the window size becomes smaller, the dataset becomes larger making more data available for training and testing.

KNN is the worst performing model among the three but the performance of is significantly improved with the help of scaling. Weighted KNN performs better than KNN in terms of classification and applying scaling to the dataset has a significant influence on the results. This is so because KNN uses feature values in its distance calculation.

When one of the feature values is significantly higher than those of other features, it will dominate the distance and affect the results. Applying weights is beneficial because the dataset is imbalanced and thus applying weights ensure that enough importance is given to points closest to the test point. Since KNN is a distance based algorithm, it doesn't take much time to fit the model. However, it takes a long time, approximately 20 minutes in order to predict the labels of the test data. The time taken would increase if the size of the test set is increased. (*K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint* n.d.)

Decision trees were giving good results and the time taken for computation was also around 5 to 7 minutes. The best results were obtained from random forests. An important fact to note here is that although random forests were giving highest accuracy and F1 score, it not significantly better than decision trees and the time taken for building the model was comparatively more than the time taken to build the decision tree model (12-14 minutes). Thus, using decision tree is as good as random forest if the slight improvement in accuracy is not highly significant.

Chapter 5

Conclusion

5.1 Summary and Results

After training and testing all the models it can be concluded that for both the datasets, random forest is performing the best. The results obtained are better than the results from the paper titled 'Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data' (Bobade, and Vani 2020). One of the reasons for the better results can be attributed to using the complete Chest data as it is has the highest feature importance of 18.9% for the 3 class classification problem (using random forests) and this feature has not been used in 'Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data' (Bobade, and Vani 2020) paper. The results also have significant improvement from the original WESAD dataset research paper titled 'Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection' by Schmidt et al. 2018.

5.2 Future Work

One of the things that could have been explored if there was more time would be personalisation of the results. These results were obtained by combing the data of all the fifteen subjects and then training models using it. Another interesting analysis would be to use some of the subjects for training the models and some for testing instead of combining the data of all the subjects.

The preprocessing that should be done in order to train models subjectwise is as given. Firstly, the data for each subject from the synchronised .pkl file should be extracted and stored in lists. After choosing a desired window size, each of the dataframe made for individual subjects should be preprocessed using this window size and the statistical measures desired should be applied in order to form a new data frame. This new dataframe would be for each subject, so a total of 15 dataframes would be present.

After this, using a suitable language like Python or R, a program should be written so that for each model some of the subjects will be chosen (randomly) for training and the remaining would be for testing. The results could be surprising as it could be affected by the gender of the subject as well. Machine learning techniques like K fold cross validation, leave one out cross validation could be used if the results aren't up to the mark. In leave one out cross validation, instead of leaving a row out, here the program has to be written in such a way that it leaves one data frame out.

Although the results obtained were very good, the preprocessing was a major step for the WESAD dataset. Another perspective that could be put into use is to use deep learning techniques as they are known to be extremely effective with huge datasets. Instead of using statistical measures like mean, standard deviation, minimum and maximum, the features could be extracted from the synchronised pickle file using a suitable window size to make a dataframe and this dataframe could be used to train deep learning models. Some of the models that could be trained include the ANN (Artifical Neural Network) model, single layer perceptron model, multilayer perceptron model and so on. One of the major advantages of using deep learning for this classification problem is that the preprocessing stage would be simplified to a great extent. At the same time, one of the disadvantages would be the expense of training a deep model. It would take more time to train deep learning model than the time it would take to train a machine learning model. In addition to this, one would require a good system and software in order to successfully train the deep learning model. However, if everything is available then deep learning models might provide results that are as good as the machine learning models or even better without having the hassle of preprocessing and using statistical measures to get the dataset used to train ML models.

Bibliography

Bobade, Pramod, and M. Vani [2020]. "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data". In: *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 51–57. DOI: `10.1109/ICIRCA48905.2020.9183244`.

Do Decision Trees need Feature Scaling? | by Praveen Thenraj | Towards Data Science [n.d.] `https://towardsdatascience.com/do-decision-trees-need-feature-scaling-97809eaa60c6`. (Accessed on 09/01/2022).

Downsampling and Upsampling of Images â Demystifying the Theory | by Aashish Chaubey | Analytics Vidhya | Medium [n.d.] `https://medium.com/analytics-vidhya/downsampling-and-upsampling-of-images-demystifying-the-theory-4ca7e21db24a`. (Accessed on 09/01/2022).

Entropy and Information Gain. Yet another tool used to make Decisionâ¦ | by Steven Loaiza | Towards Data Science [n.d.] `https://towardsdatascience.com/entropy-and-information-gain-b738ca8abd2a`. (Accessed on 08/31/2022).

Everything to Know About Stress: Causes, Prevention, and More [n.d.] `https://www.healthline.com/health/stress`. (Accessed on 08/31/2022).

Fight, Flight, or Freeze: How We Respond to Threats [n.d.] `https://www.healthline.com/health/mental-health/fight-flight-freeze`. (Accessed on 09/01/2022).

Folkman, Susan et al. [1987]. "Age differences in stress and coping processes." In: *Psychology and Aging* 2.2, 171â184. DOI: `10.1037/0882-7974.2.2.171`.

Healey, J.A., and R.W. Picard [2005]. "Detecting stress during real-world driving tasks using physiological sensors". In: *IEEE Transactions on Intelligent Transportation Systems* 6.2, pp. 156–166. DOI: `10.1109/TITS.2005.848368`.

Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial | SpringerLink [n.d.] `https://link.springer.com/article/10.1007/s12671-018-0905-4`. (Accessed on 08/31/2022).

Joblib: running Python functions as pipeline jobs â joblib 1.2.0.dev0 documentation [n.d.] https://joblib.readthedocs.io/en/latest/. (Accessed on 09/01/2022).

K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint [n.d.] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning. (Accessed on 09/02/2022).

Keshan, N., P. V. Parimi, and I. Bichindaritz [2015]. "Machine learning for stress detection from ECG signals in automobile drivers". In: *2015 IEEE International Conference on Big Data (Big Data),* pp. 2661–2669. DOI: 10.1109/BigData.2015.7364066.

KNN Classification Tutorial using Sklearn Python | DataCamp [n.d.] https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn. (Accessed on 09/01/2022).

Koldijk, Saskia et al. [2014]. "The SWELL Knowledge Work Dataset for Stress and User Modeling Research". In: *Proceedings of the 16th International Conference on Multimodal Interaction.* ICMI '14. Istanbul, Turkey: Association for Computing Machinery, 291â298. ISBN: 9781450328852. DOI: 10.1145/2663204.2663257. URL: https://doi.org/10.1145/2663204.2663257.

Koolhaas, J.M. et al. [2011]. "Stress revisited: A critical evaluation of the stress concept". In: *Neuroscience Biobehavioral Reviews* 35.5, pp. 1291–1301. ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2011.02.003. URL: https://www.sciencedirect.com/science/article/pii/S0149763411000224.

McCorry, Laurie Kelly [Aug. 2007]. "Physiology of the autonomic nervous system". en. In: *Am. J. Pharm. Educ.* 71.4, p. 78.

Mode Definition [n.d.] https://www.investopedia.com/terms/m/mode.asp. (Accessed on 08/31/2022).

Monique Tello, MD [Oct. 2016]. *Regular meditation more beneficial than vacation.* URL: https://www.health.harvard.edu/blog/relaxation-benefits-meditation-stronger-relaxation-benefits-taking-vacation-2016102710532.

Open Sourcing Mental Illness Wants to Break the Mental Health Stigma [n.d.] https://www.businessinsider.com/open-sourcing-mental-illness-stigma-tech-2019-12?r=US&IR=T. (Accessed on 09/01/2022).

PhysioBank Databases [n.d.] https://archive.physionet.org/physiobank/database/. (Accessed on 08/31/2022).

Prabhavathi, K et al. [Aug. 2014]. "Role of biological sex in normal cardiac function and in its disease outcome - a review". en. In: *J. Clin. Diagn. Res.* 8.8, BE01–4.

Python pickling: What it is and how to use it securely | Synopsys [n.d.] https://www.synopsys.com/blogs/software-security/python-pickling/#:

~:text=Pickle\%20in\%20Python\%20is\%20primarily,transport\%20data\%20over\%20the\%20network.. (Accessed on 09/01/2022).

Random Forest | Introduction to Random Forest Algorithm [n.d.] https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/. (Accessed on 08/31/2022).

Reddy, U Srinivasulu, Aditya Vivek Thota, and A Dharun [2018]. "Machine Learning Techniques for Stress Prediction in Working Employees". In: *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–4. DOI: 10.1109/ICCIC.2018.8782395.

Research :: Open Sourcing Mental Health - Changing how we talk about mental health in the tech community - Stronger Than Fear [n.d.] https://osmhhelp.org/research. (Accessed on 08/31/2022).

Schmidt, Philip et al. [2018]. "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection". In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction.* ICMI '18. Boulder, CO, USA: Association for Computing Machinery, 400â408. ISBN: 9781450356923. DOI: 10.1145/3242969.3242985. URL: https://doi.org/10.1145/3242969.3242985.

sklearn.ensemble.RandomForestClassifier â scikit-learn 1.1.2 documentation [n.d.] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. (Accessed on 08/31/2022).

sklearn.neighbors.KNeighborsClassifier â scikit-learn 1.1.2 documentation [n.d.] https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. (Accessed on 08/31/2022).

sklearn.tree.DecisionTreeClassifier â scikit-learn 1.1.2 documentation [n.d.] https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html. (Accessed on 08/31/2022).

Stress effects on the body [n.d.] https://www.apa.org/topics/stress/body. (Accessed on 08/31/2022).

Stress effects on the body [n.d.] https://www.apa.org/topics/stress/body. (Accessed on 09/01/2022).

The F1 score | Towards Data Science [n.d.] https://towardsdatascience.com/the-f1-score-bec2bbc38aa6#:~:text=The\%20F1\%20score\%20becomes\%20especially,grid\%20search\%20or\%20automated\%20optimization.. (Accessed on 08/31/2022).

UCI Machine Learning Repository: WESAD (Wearable Stress and Affect Detection) Data Set [n.d.] https://archive.ics.uci.edu/ml/datasets/WESAD+\%28Wearable+Stress+and+Affect+Detction\%29. (Accessed on 08/31/2022).

What is a Decision Tree | IBM [n.d.] https://www.ibm.com/topics/decision-trees. (Accessed on 08/31/2022).

What Is the Difference Between Chronic and Acute Stress? - AFC Urgent Care [n.d.] https://www.afcurgentcarehixsontn.com/what-is-the-difference-between-chronic-and-acute-stress/. (Accessed on 08/31/2022).

What is the k-nearest neighbors algorithm? | IBM [n.d.] https://www.ibm.com/topics/knn#:~:text=The\%20k\%2Dnearest\%20neighbors\%20algorithm\%2C\%20also\%20known\%20as\%20KNN\%20or,of\%20an\%20individual\%20data\%20point.. (Accessed on 08/31/2022).

Workplace Stress Responsible For Up To $190B In Annual U.S. Healthcare Costs [n.d.] https://www.forbes.com/sites/hbsworkingknowledge/2015/01/26/workplace-stress-responsible-for-up-to-190-billion-in-annual-u-s-heathcare-costs/?sh=1a544a83235a. (Accessed on 08/31/2022).

Appendix

All the codes used in this paper can be accessed via the following link :

`https://github.com/Aparna-K28/Research-Project`