Nakul Sharma /Data Analytics and Artificial Intelligence, 5(1), 2025, 44-46



Data Analytics and Artificial Intelligence Vol: 5(1), 2025 REST Publisher; ISBN: 978-81-948459-4-2 Website: http://restpublisher.com/book-series/daai/ DOI: https://doi.org/10.46632/daai/5/1/8



DOI: https://doi.org/10.46632/daai/5/1/8

Stress Detection Datasets: A Survey Nakul Sharma

Vishwakarma Institute of Information Technology, Kondhwa, Pune, India Corresponding author Email: nakul777@gmail.com

Abstract. The stress detection and its mitigation form an essential part within everyday life. Stress and its known and unknown causes can be detected by creating a model to predict its occurrence. This chapter introduces different datasets that have been used in detection of stress in the current literature. A review of these datasets has been provided. These datasets have been used for creating various Machine Learning or Deep learning models.

1. INTRODUCTION

Stress detection and its mitigation is a challenging task as symptoms are not visibly present. Stress when remaining undetected, can cause deep depression and can have several other side effects. One of the central tasks of Machine Learning Model is prediction. The ML model can help in providing predictions based on the symptoms of patient. There are different types of datasets used by these ML/DL models for detect stress. Different data-sets have different format to save data.

Data-sets form integral part of AI/ML/DL pipeline in order to predict more accurately and precisely. The standard methodology involving any ML/DL model is pre-processing of data, conducting Feature Engineering on the data-set, splitting the data-set into training and testing of data, conducting training and testing on the data-set. The results of accuracy, precision, recall are further enhanced by introducing hyper-parameter tuning.

In this chapter, PhysioNet, are discussed. PhysioNet is research resource supported by National Institute of Health (NIH). NIH primarily provides funds for conducting research and creating research artifacts which are useful to medical researchers [1]. The details of how PhysioNet resources help in stress detection is given in section-2 of this chapter. Another essential resource which helps in dress detection is Wearable Stress and Affect Detection dataset (WESAD). This dataset consists of multi-model information related to stress detection and also includes multiple affective states. It contains information about physiological and behavioral data. The author's contribution includes WESD dataset, different benchmarks for evaluation and bridges the gap between previous lab studies. The details about WESAD dataset is provided in section-3 of this chapter [2].

Psychometric tool for detecting stress is Perceived Stress Scale (PSS). The PSS has been used to assess the amount of stress perceived by different categories of people. This can be applied to both female and male patients across different ages as well. The stress perceptions were rated on a scale to determine the amount of stress being perceived by the patients. PSS is discussed in detail in section-3 of this chapter [3].

SayoPillow is a tool developed by National Sleep Research Resource (NSRR) in order to acquire from patients some parameters while sleeping. The tool was designed for SayoPillow (Smart Yoga Pillow). SyoPillow is discussed in detail in section-4 of this chapter [4].

Stress detection and its prevention are very much necessary for healthy life. Chronic cases of stress when left untreated can lead to deep depression and mental imbalance. The side effects of chronic stress are obesity, insomnia, diabetes, depression and cancer [4]. The medical fraternity has provided several mediums of stress detection and monitoring. These can be used in conjunction for better diagnostic and clinical evaluation.

2. PHYSIONET

PhysioNet provides a repository of medical research data and was also formerly managed by by Massachusetts Institute of Technologies laboratory for computational physiology. PhysioNet provides stress detection datasets in form of signals. PhysioNet also provides data-set of automobile drivers. This data-set consists of two files drive17a and drive17b which are part of one experiment. These files are of 29 and 25 minutes respectively. There are other 16 records consisting of 65 to 93

minutes [1]. The main resource bank of the website physionet.org is PhysioBank which consists of an archive of digital recordings related to cardiological signals. The different signals collected are from patients suffering from coronary sleep syndromes and cardiac anhythmias. There are several related databases provided in PyshioNet. These include ambulatory ECG recordings, non-EEG datasets related to neurological signals, wearable device based dataset. The website also includes PhysioToolkit which is contains different physiologic signals for analysis. There is an online forum as well for the purpose of exchanging and communicating recorded biomedical signals.

The specialized dataset of drivers who were driving on prescribed route is provided in the dataset titled Stress Recognition in Automobile Drivers. These drivers moved in and around city streets and highways of Boston and Massachusetts. The recorded signals include ECG, EMG, GSR measured on the hand and foot and respiration. These signals can help in investigation of stress during driving. Those patients who regularly undertake driving and also feel stressed can see spike in any one of these signals [1].

3. WSED DATASET

WSED stands for Wearable Stress and Affect Detection. This is a publicly available dataset for affective computing. It is a multi-modal dataset which contains physiological and motion data. It records wrist and chest worn device information of 15 patients daily. More specifically, it contains information several patients' parameters which provide an indication of stress. The datasets also provides affective states of neutral, stress, amusement. The overall accuracy of 93% was achieved using the existing datasets. The dataset provided an enhancement to the existing studies by providing three different affective states which categorized different states of having stress. The dataset research also provided a benchmark on various standard ML algorithms for improvement and possible deployment. The dataset can accessed online using the url https://ubi29.informatik.uni-siegen.de/usi/data_wesad.html. It also provides an interactive and fast in-browser version of the dataset using the url https://kristofvl.github.io/. The scope of study was using wearable devices to measure various stress related parameters [2].

The dataset included some general questions as well related to patients height, weight etc. It also included some stress specific questions about intake of alcohol, caffeine intake and smoking history [2]. The study protocol included preparing the participants for data collection. The next phase was preparing the base line condition, providing the amusement condition to the participants, exposing the participants to TSST test, meditation for normalizing the participants and final recovery [2].

Perceived Stress Scale (PSS) is a medium of checking for stress preceptors. The original PSS consisted of 14 items of selfreport to indicate specific conditions which have been stressful in life. This measure perceived the level of stressful events which have occurred in life. The original intention of the dataset was to predict stress for adults who had cleared high school. The question asked to the respondents consisted for basic information, along with specific information related to stressful conditions in respondent's life. The specific questions also delved into past month's reason for being upset in life. There were also general questions about inability to control essential events in life. PSS stress detection is more manual based method of detecting stress. There were various variants to tis model of checking stress. These included changing the questionnaire according to the specific situation which the respondents have gone through. The exact question also reflected on the quantum of stress perceived by the respondents. Another variation included creating different segments of the questionnaire to reflect the type of respondents filling up the questionnaire. These included different questionnaire for the chronically traumatized respondents and different from lightly traumatized respondents. The questionnaire when filled properly did help in perceiving stress of the respondents [3].

4. SRR SLEEP STUDY DATASET

SaYoPillow is developed to recognize the importance of good quality sleep which helps to reduce stress.

The system developed makes use of National Sleep Research Resource (NSRR) dataset. It contains information about heart disease and sleeping disorders. The dataset has 6,441 individuals information between the years 1995 to 1998. The dataset also contains information of Polysomnograms conducted on 5,793 people. The parameters specific to stress are used in study and assessment only and rest are ignored. These inputs are used for SaYoPillow software [4].



FIGURE 1. Stress detection datasets

5. OPERATIONS ON DATASETS

There are various operations which can be performed on the datasets. These include conduction of Feature Engineering and Exploratory Data Analysis (EDA). Before the dataset is fed into ML/DL algorithms, these tasks must be completed.

Feature Engineering includes tasks which are requisite steps for generating a more accurate and precise model. These tasks includes preprocessing to check for null values, missing values, data normalization and extraction of features. Each and every step mut be done properly and perfectly. Any step not conducted properly or having some deficiencies is likely to reuce accuracy and preceision of the dataset included. In additon to Feature Engineering, a more exhaustive EDA can also be performed to know better different patterns present within the dataset.

6. ML/DL ALGORITHMS

There are different ML/DL algorithms that can be applied to a preprocessed dataset. ML algorithms that can be applied depending on the type of prediction which will be done by created system. A combination of algorithms can also be applied. This combination can enable increase in the accuracy and precision of the created system. A good choice is to use Random Forest or SVM in combination of other ML algorithm. In case of Deep Learning (DL), it is essential to create a large size dataset are the pre-requisite. DL algorithms will work well with large size dataset. DL is computationally intensive task and utilizing the services of GPU/TPU with small dataset is not advisable.

7. CONCLUSION

In this chapter, different types of datasets which have been proposed or are being used for stress detection are discussed. There are several other stress detection which can are derived from these datasets or are newly proposed based on the existing datasets. These included PhysioNEt, WSED, PSS, SRR Sleep datasets.

REFERENCES

- [1] Costa, Madalena, George B. Moody, Isaac Henry, and Ary L. Goldberger. "PhysioNet: an NIH research resource for complex signals." Journal of electrocardiology 36 (2003): 139-144, https://doi.org/10.1016/j.jelectrocard.2003.09.038
- [2] Schmidt, Philip, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. "Introducing wesad, a multimodal dataset for wearable stress and affect detection." In Proceedings of the 20th ACM international conference on multimodal interaction, pp. 400-408. 2018. https://doi.org/10.1145/3242969.3242985

[4] Rachakonda, Laavanya, Anand K. Bapatla, Saraju P. Mohanty, and Elias Kougianos. "Sayopillow: a blockchain-enabled, privacyassured framework for stress detection, prediction and control considering sleeping habits in the iomt." arXiv preprint arXiv:2007.07377 (2020). https://doi.org/10.48550/arXiv.2007.07377

Chan, S.F., La Greca, A.M. (2020). Perceived Stress Scale (PSS). In: Gellman, M.D. (eds) Encyclopedia of Behavioral Medicine. Springer, Cham. https://doi.org/10.1007/978-3-030-39903-0_773.