**Stress-Predict Dataset**

This pilot study is to develop a stress prediction model that classify individuals to stress or not stressed groups based on signal data obtained from Empatica watches.

This study comprises of 35 volunteer participants who performed three different stress inducing tasks (i.e., Stroop color word test, interview session and hyperventilation session) with baseline/relax period in-between each task, for 60 minutes. Blood volume pulse (BVP), Inter-beat-intervals, and heart rate were being continuously recorded using Empatica watches while respiratory rate was estimated using PPG-based respiratory rate estimation algorithm.

**Dataset Files**

**|-- *Processed\_data***

|----- ***heartrate\_resprate\_timestamps\_labels folder***

|----- ***Improved\_Combined\_hr\_rsp\_binary\_PX.csv*** (contain information of heart rates and respiratory rates along with timestamps and labels (for nonstress/baseline and 1 for stress task duration). Here *X* is participant number)

|----- ***Time\_logs.xlsx*** (contain date and start/end time of each task for each participant, Irish standard time)

|----- ***heartrate\_timestamps\_labels folder***

|----- ***PX\_comb\_binary.csv*** (contain information of heart rates along with timestamps and labels (for nonstress/baseline and 1 for stress task duration). Here *X* is participant number)

|----- ***resprrate\_timestamps\_labels folder***

|----- ***Improved\_PX\_comb\_10sec\_binary.csv*** (contain information of respiratory rates along with timestamps and labels (for nonstress/baseline and 1 for stress task duration). Here *X* is participant number)

|----- ***Improved\_All\_Combined\_hr\_rsp\_binary.csv*** (contain information of heart rates and respiratory rates of all the participants along with timestamps and labels (for nonstress/baseline and 1 for stress task duration))

|----- ***Questionnaires\_scores.xlsx*** (contains information about the PSS and STAI questionnaire scores of each participant)

|----- ***Time\_logs.xlsx*** (contain date and start/end time of each task for each participant, Irish standard time)

**|-- *Raw\_data***

|----- **S*X* *folder*** (folders with raw files from Empatica E4. Where *X* is participant number)

|----- ***ACC.csv*** (contains accelerometer data (x, y, z axis))

|----- ***BVP.csv*** (contains raw BVP data)

|----- ***EDA.csv*** (contains EDA data (skin conductance))

|----- ***HR.csv*** (contains heart rate data)

|----- ***IBI.csv*** (contains inter-beat-interval data)

|----- ***info.txt*** (contains information about all the csv file and sampling rate)

|----- ***tags\_SX.csv*** (contains timestamp tags. start-end time of each task)

|----- ***TEMP.csv*** (contains skin temperature data)

**Libraries**

Following libraries were used for analysis:

* Descriptive Analysis (python):
  + pandas
  + numpy
  + seaborn
  + matplotlib
* Classification Analysis (python):
* Numpy
* Pandas
* Scikitlearn
* Scipy
* Matplotlib

**References**

These materials are required for citation when using the dataset:

1. Talha Iqbal, Andrew Simpkin, Nicola Glynn, John Killilea, Jane Walsh, Gerard Molloy, Adnan Elahi, Sandra Ganly, Eileen Coen, William Wijns, and Atif Shahzad. *“Stress Levels Monitoring Using Sensor-Derived Signals from Non-Invasive Wearable Device: A Pilot Study and Stress-Predict Dataset.”,* Nature Scientific Data [Under Review]
2. Talha Iqbal, Adnan Elahi, Sandra Ganly, William Wijns, and Atif Shahzad. *"Photoplethysmography-Based Respiratory Rate Estimation Algorithm for Health Monitoring Applications."* Journal of medical and biological engineering 42, no. 2 (2022): 242-252.