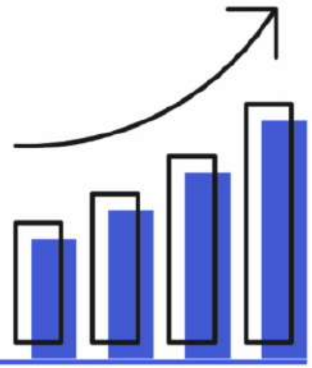
Data Science Project

Predicting Stress Level with PPG features



Learning Outcomes:

In this project we will:

- Learn and understand about classification using ANNs
- Learn about Photoplethysmography and Heart Rate Variability
- Learn how to clean and process dataset (outlier, missing values, etc.)
- Learn about Correlation and feature selection.
- Learn about Artificial Neural Networks.
- Deploy our model as a Flask webapp

Part I (i): Classification and ANNs

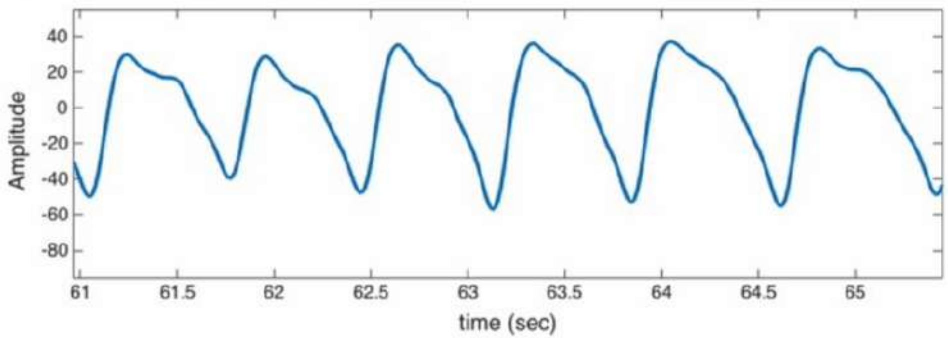
Classification is a common machine learning algorithm which is used to identify and categorize data into 2 or more different classes.

Few examples of classification problems can be:
- Object classification: is it a cat or a dog?
- Sentiment classification: Emotion from facial features
- Classification of different kinds of flowers
- Classification of tumour being benign or malignant.

In this project, we are going to look at such a classification problem, predicting the stress level from PPG data.

Part I (ii): Photoplethysmography

- Photoplethysmography, known mostly as PPG, utilizes an infrared light to measure the volumetric variations of blood circulation. This measurement provides valuable information about the cardiovascular system.
- The heart rate variability (HRV) is a measure of variation in time duration between consecutive heart beats. HRV has used widely as an indicator for stress, health, and various disease conditions.
- In a PPG signal, the location of a peak represents the instant of time at which a heartbeat occurs. Thus, the computation of HRV requires accurate identification of the location of peaks in the PPG signal, which consequently leads to precise computation of time intervals between consecutive heartbeats.
- Many relevant features can be derived from the HRV measurements such as the number of interval differences of successive NN intervals greater than 50ms (pNN50), and the square root of the mean squared differences of successive NN intervals (RMSSD). We can also get frequency domain features like LF, HF and LF-HF. *(It is recommended to go through https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4916478/)*



Part I (iii): Photoplethysmography

As mentioned previously, we get Heart Rate and Heart Rate Variability from the PPG directly. This can be stored in any cloud-based server which then can be easily downloaded as a .csv/.xlsx file.
MATLAB algorithms can be used to extract other features such as RMSSD, pNN25, pNN50, LF and HF.

- **RMSSD**: Root Mean Square of Successive Differences
- **pNN25/pNN50:** The proportion of number of pairs of successive NN (R-R) intervals that differ by more than 25/50 ms by the total number of RR intervals.
- **High Frequency power (HF):** frequency activity in the 0.15 - 0.40Hz range.
- **Low Frequency power (LF):** frequency activity in the 0.04 - 0.15Hz range.

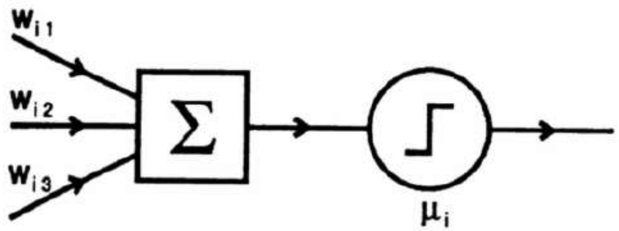
With regards to Heart Rate Variability, research has identified certain frequency bands that tend to correlate with certain physiological phenomenon, such as Parasympathetic nervous system activity.

Part I (iv): The Dataset

- Link: https://www.kaggle.com/datasets/vinayakshanawad/heart-rate-prediction-to-monitor-stress-level?select=Train+Data
- Make sure to read both the time domain features and frequency domain features and make a Dataframe with the features mentioned here in the PPT.
- The task at hand: Classifying the Stress Level.
- Let us explore the different features in the dataset and try to understand how we can proceed with model building.

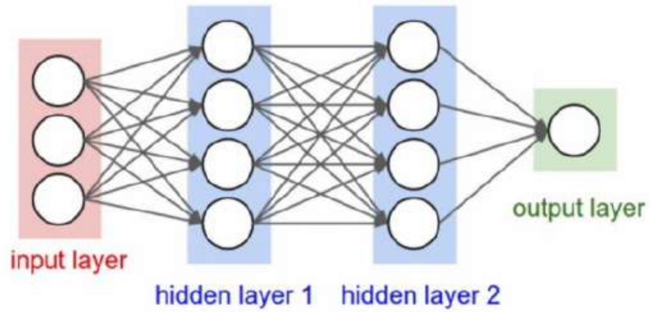
Part II (i): Artificial Neural Networks

- Artificial Neural Networks are essentially an adaptation of the human brain in machines.
- Just like there are neurons firing in the brain to do functions, ANNs also have "neurons" which have some input and output with activation functions. (Fig 1)
- We will be making a Feed-Forward Neural Network.
- Training is done by inputting the data and changing the weights iteration by iteration to get the desired output.
- Small increments are made to the weights to minimize the overall loss.
- Here, by loss, we mean the categorical loss between predicted class and actual class.
- Fig 2 shows the equation of the given neuron.
- Fig 3 shows a Feed-Forward Neural Network with 3 inputs, 2 hidden layers and one output.



$$v = \sum_{j=0}^{m} w_j x_j$$

$$w_0 = b$$



Part II (ii): Implementing FF-ANN

- We will implementing the Feed-forward Neural network using TensorFlow.

- TensorFlow provides tools to create complex NN architectures and gives the freedom to set and tune hyperparameters.

- Using the feature variables and FF-ANN, we will implement a Classification model.
- The model will be then saved and used to predict with test data and then be used to deploy a Flask webapp.