Phase1 – feature extraction

- Time Domain Features
  - Peek to peek (PTP)
    Calculates the difference between the maximum and minimum value in input.

$$PTP = X_{Max} - X_{Min}$$

  - AASS
    Calculates total vertical length of a signal, with value $\tau = 1$

$$AASS = \frac{1}{N} \sum_{i=0}^{N-\tau} \frac{1}{\tau} |x_{i+\tau} - x_i|$$

  - Singular Spectrum Analysis (SSA)
    Shows how well different components can be separated from each other.

$$SSA = \sum_{i=\tau}^{N-\tau} \frac{1}{2} | \frac{x_{i-\tau} - x_i}{|x_{i-\tau}| - x_i} + \frac{x_{i+\tau} - x_i}{|x_{i+\tau}| - x_i} |$$

  - Log Detect
    Logarithmic detectors are particularly used in pulse detection.

$$\log Dectect = e^{\frac{1}{N}\sum_{k=1}^{N} \log(|x_k|)}$$

    Source : Research Study of stability of time-domain features for electromyographic pattern recognition - Dennis Tkach, He Huang, and Todd A Kuiken

  - Zero Crossings (ZC)
    It is the number of times signal x crosses zero within an analysis window; To avoid signal crossing counts due to low-level noise, a threshold ε was included (ε = 0.01) The ZC count increased by one if

$$\{ x_k > 0 \, and \, x_{k+1} < 0 \} \, or \, \{ x_k < 0 \, and \, x_{k+1} > 0 \})$$
$$and \, | x_k - x_{k+1} | \geq \varepsilon$$

    Source : Research Study of stability of time-domain features for electromyographic pattern recognition - Dennis Tkach, He Huang, and Todd A Kuiken

- Statistical Features
  - Mean

    Simply divides the sum of all values in a data set by the number of values

$$\text{mean} = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

  - Median

    It first Orders dataset from lowest to highest value, the median is the data point that separates the upper half of the data values from the lower half.
    If the size of the dataset n is odd, the median is the value at position p where

$$p = \frac{n+1}{2}$$

$$\tilde{x} = x_p$$

    If the size of the dataset n is even, the median is the average of the values at positions p and p + 1 where

$$p = \frac{n}{2}$$

$$\tilde{x} = \frac{x_p + x_{p+1}}{2}$$

  - Percentile

    A percentile is defined as a score at or below which a given percentage falls.

$$Percentile = \frac{Number\ of\ Values\ Below\ some\ value\ \text{“}x\text{”}}{Total\ Number\ of\ Values} \times 100$$

  - Standard Derivation (STD)

    The standard deviation is a measure of the amount of variation or dispersion of a set of values.
    A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

  - Histogram

    It represents the frequency of data distribution, With 10 equal-width bins , At the end, the sum of the obtained values is calculated.

- Entropy
  - Sample Entropy

    It quantifies a signals complexity irrespective of the signal length, used for assessing the complexity of physiological time-series signals.
    It is based on the conditional probability that two sequences of length 'n+1' randomly selected from a signal will match, given that they match for the first 'n' elements of the sequences. Here 'match' means that the distance between the sequences is less than some criterion 'k' which is usually 20% of the standard deviation of the data sequence taken into account.

$$\text{Sample Entropy} = -\log\left(\frac{A^n(k)}{B^n(k)}\right)$$

    where $B^n(k)$ is the estimated probability that two sequences match for n points, and $A^n(k)$ is the estimated probability that the sequences match for n+1 points.

    Source : EEG-based human emotion recognition using entropy as a feature extraction measure - Pragati Patel, Raghunandan R and Ramesh Naidu Annavarapu

  - Approximate Entropy

    It is a technique used to quantify the amount of regularity and the unpredictability of fluctuations over time-series data, which measures the randomness of the fluctuation in given data set.
    Smaller values indicates that the data is more regular and predictable.

$$\text{ApEn}\,(n,k,N) = \ln\left(\frac{C_n(k)}{C_{n+1}(k)}\right)$$

    where n is subseries length , k is similarity tolerance(coefficient), N is data length, $C_n(k)$ and $C_{n+1}(k)$ are pattern mean of length (n) and (n + 1).

    Source : EEG-based human emotion recognition using entropy as a feature extraction measure - Pragati Patel, Raghunandan R and Ramesh Naidu Annavarapu

  - Spectral Entropy

    It is defined to be the Shannon entropy of the power spectral density (PSD) of the data. It is a measure of the random process uncertainty from the frequency distribution. A low SE value means the frequency distribution is intense in some frequency bands.

$$H(x, sf) = -\sum_{f=0}^{f_s/2} P(f)\log_2[P(f)]$$

    Where P is the normalized PSD, and f_s is the sampling frequency. Using Welch periodogram as Spectral estimation method and mean frequency in dataset as sample frequency.
    Source : A review of feature extraction and performance evaluation in epileptic seizure detection using EEG - Poomipat Boonyakitanont , Apiwat Lek-uthai , Krisnachai Chomtho , and Jitkomut Songsiri

- Permutation Entropy

The highest value of PE is 1, signifying that the data series is purely unpredictable; whereas the lowest value of PE is 0, signifying that the data series is entirely predictable.

$$\mathrm{PE} = -\sum_{i=1}^{n} P_i \log_2 P_i$$

where p_i represents the relative frequency of possible sequence patterns, n implies permutation order of $n \geq 2$

Source : EEG-based human emotion recognition using entropy as a feature extraction measure - Pragati Patel, Raghunandan R and Ramesh Naidu Annavarapu

- Singular Value Decomposition Entropy (SVD)

It measures the dimensionality of the data.

$$H = -\sum_{i=1}^{M} \overline{\sigma}_i log_2(\overline{\sigma}_i)$$

Where M is the number of singular values of the embedded matrix Y and σ1,σ2,...,σM are the normalized singular values of Y.

The embedded matrix Y is created by

$$y(i) = [x_i, x_{i+\text{delay}}, \ldots, x_{i+(\text{order}-1)*\text{delay}}]$$

$$Y = [y(1), y(2), \ldots, y(N - (\text{order} - 1)) * \text{delay})]^T$$

Source : A review of feature extraction and performance evaluation in epileptic seizure detection using EEG - Poomipat Boonyakitanont , Apiwat Lek-uthai , Krisnachai Chomtho , and Jitkomut Songsiri