**1)**

**Dataset Information**

**Rechtschaffen and Kales standard** - The criteria established by Rechtschaffen and Kales provide a systematic way to visually analyse and score sleep stages based on specific patterns in the polysomnographic data. This standardized scoring system allows for consistency in sleep research and clinical assessments.

The sleep-edf. database contains 197 whole-night Polysomnographic sleep recordings, containing EEG, EOG, chin EMG, and event markers. Some records also contain respiration and body temperature. Corresponding hypnograms (sleep patterns) were manually scored by well-trained technicians according to the Rottschafer and Kales manual and are also available.

Fpz-Cz and Pz-Oz electrode locations

These patterns (hypnograms) consist of sleep stages W, R, 1, 2, 3, 4, M (Movement time)

The PSG files are formatted in EDF while the hypnograms are in EDF+. The specifications of EDF and EDF+

The 153 **SC\*** files (SC = Sleep Cassette) were obtained in a 1987-1991 study of age effects on sleep in healthy Caucasians aged 25-101, without any sleep-related medication [2]. Two PSGs of about 20 hours each were recorded during two subsequent day-night periods at the subjects’ homes. Subjects continued their normal activities

The 44 **ST\*** files (ST = Sleep Telemetry) were obtained in a 1994 study of temazepam effects on sleep in 22 Caucasian males and females without other medication. Subjects had mild difficulty falling asleep but were otherwise healthy. The PSGs of about 9 hours were recorded in the hospital for two nights

Frequency - 100 Hz

**2)**

**Electrode locations**

Fpz-Cz: This represents the electrode placement from the frontal pole (Fpz) to the vertex or central midline (Cz). The frontal pole is located at the most anterior part of the head, and Cz is at the centre of the head's surface.

Pz-Oz: This indicates the electrode placement from the vertex or central midline (Pz) to the occipital pole (Oz). Pz is again at the centre of the head's surface, and Oz is at the most posterior part of the head.

Fpz-Cz:

F: Frontal - Refers to the frontal region of the head.

p: Pole - Indicates the most anterior part of the head.

z: Zero - Represents the midline reference point.

So, "Fpz-Cz" signifies an electrode placement from the Frontal Pole (most anterior part of the head) to the midline reference point (Zero).

Pz-Oz:

P: Parietal - Refers to the parietal region of the head.

z: Zero - Represents the midline reference point.

O: Occipital - Indicates the most posterior part of the head.

Therefore, "Pz-Oz" represents an electrode placement from the midline reference point (Zero) at the Parietal region to the most Occipital (posterior) part of the head.

**3)**

**Why used CNN- LSTM over SVM and RF?**

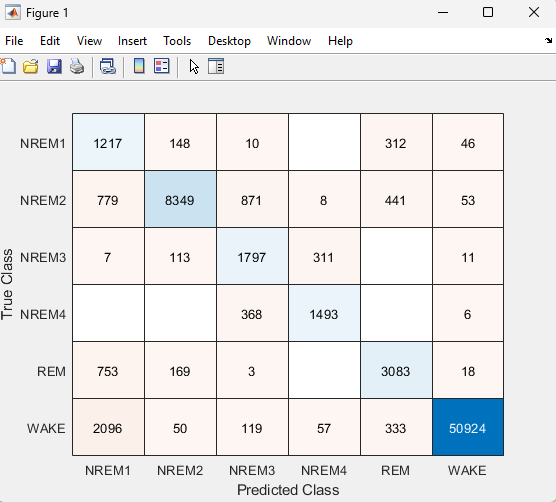
Conventional machine learning algorithms such as support vector machines (SVM) , random forests and neural networks to train the model for sleep stage classification based on the extracted features. Although these methods have achieved a reasonable performance, they carry several limitations including the need for a prior knowledge of sleep analysis and are not able to generalize to larger datasets from various patients with different sleep patterns. The second category includes the automated feature extraction-based methods such as deep learning algorithms, in which the machine extracts the pertaining features automatically.

One key reason we have used CNN-LSTM (deep learning model) is the availability of large amounts of data to learn the underlying complex pattern in the data sets. Due to availability of many sleep EEG recordings, deep learning algorithms have also been applied for sleep stage classification

**4)SVM Output**

A screenshot of a computer

Description automatically generated



**5)**

**RF Output**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**6)**

**Features**

**Mean:**

Description: The average value of the signal.

Significance: Provides a measure of the central tendency of the signal.

**Standard Deviation (std):**

Description: A measure of the dispersion or spread of the signal values.

Significance: Indicates the variability or consistency of the signal.

**Skewness:**

Description: A measure of the asymmetry of the signal's probability distribution.

Significance: Helps identify the shape of the distribution; positive skewness indicates a longer right tail.

**Kurtosis:**

Description: A measure of the "tailedness" of the signal's probability distribution.

Significance: Indicates whether the signal has heavy tails compared to a normal distribution.

**Maximum:**

Description: The maximum value in the signal.

Significance: Represents the highest amplitude reached by the signal.

**Minimum:**

Description: The minimum value in the signal.

Significance: Represents the lowest amplitude reached by the signal.

**Interquartile Range (IQR):**

Description: The range between the first quartile (25th percentile) and the third quartile (75th percentile) of the signal.

Significance: Provides a measure of the spread of the central part of the distribution, less sensitive to outliers than the full range.

**Shannon's Entropy (spentropy):**

Description: A measure of the signal's disorder or uncertainty.

Significance: Quantifies the amount of information or randomness in the signal.

**Fractal Dimensions (fd)**:

Description: The fractal dimension of the signal typically estimated using methods like the Whittle estimator for fractional Brownian motion.

Significance: Describes the complexity and self-similarity of the signal.

**Activity:**

Description: The variance of the signal, indicating its overall energy.

Significance: Measures the overall magnitude of the signal.

**Mobility:**

Description: The ratio of the standard deviation of the first-order difference to the standard deviation of the signal.

Significance: Indicates the rate of change or mobility of the signal.

**Complexity:**

Description: A measure of signal complexity, calculated as the ratio of the standard deviation of the second-order difference to the product of the standard deviation of the first-order difference and the mobility.

Significance: Combines measures of variability and mobility to quantify complexity.

**Root Sum of Squares (rssq):**

Description: The square root of the sum of the squares of the signal values.

Significance: Provides an overall measure of signal magnitude.

**Mean Frequency (meanfreq):**

Description: The mean frequency of the signal.

Significance: Describes the average frequency content of the signal.

**Median Frequency (medfreq):**

Description: The median frequency of the signal.

Significance: Represents the middle point of the frequency distribution.

**Signal-to-Noise Ratio (snr):**

Description: The ratio of the signal power to the noise power.

Significance: Indicates the quality of the signal relative to noise.

**Range:**

Description: The difference between the maximum and minimum values in the signal.

Significance: Represents the full amplitude range of the signal.

**7)**

**Confusion Matrix**

Where TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative)indicate the number of sleep stages correctly labelled, the number of sleep stages correctly identified as not correspond to the sleep stages, the number of sleep stages that incorrectly labelled, and the number of sleep stages which were not identified as the sleep stages that they should have been, respectively.

A table with numbers and letters

Description automatically generated