

# Sleep detection using Machine Learning - SAATHI

## Team 17 - Silicon Labs Entrepreneurs Challenge

Inter IIT Tech Meet 10.0

### Description

This machine learning algorithm detects the sleep cycles on the basis of the accelerometer data recorded by our IoT-based healthcare solution.

We have used the XGBoost algorithm for fall detection, which is renowned for its performance and efficiency gains. Principal Component Analysis (PCA) was additionally used to transform the data to a lower dimension. The dataset used is **ICHI 14** [1], which is available in open domain. We envision our model to be dynamically trained on the data collected by our device, which ensures that our algorithm adapts to the different variations in external conditions of use.

Our machine learning algorithm provides predictions regarding the sleep stages under three heads, namely:

- **Mild Sleep** (Label 0): corresponds to NoREM sleep stages 1-3.
- **Deep Sleep** (Label 1): corresponds to REM sleep stage.
- **Awake** (Label 2): corresponds to awake/movement.

We removed the data points which were labelled as *unknown* as they did not contribute to the model training and evaluation in any way whatsoever.

### Running the code

- Install the following packages

```
numpy==1.16.1
pandas==0.25.1
```

- Download the dataset zip files from the following link:

```
https://www.researchgate.net/publication/305212784\_ICHI14-Borazio
```

- Unzip the dataset and place the individual .npz files in the `data` subfolder as `./ICHI14 dataset/*`
- Ensure to remove the irrelevant file `pat_inf.npz` file from the above directory.
- Run `Data_preparation.py` to prepare the data in the required format for training.
- Run `Train.py` to train the XGBoost model on the generated data and save the trained model for future inference.
- Run `Eval.py` to get the metrics on the test split of the ICHI 14 dataset.

### Metrics on Test Split

Classification Accuracy = 74.12%

F1 Score (Weighted) = 73.43%

Precision (Weighted) = 72.88%

Recall (Weighted) = 74.12%

We found that our metrics are consistent and are at par with the existing research studies on the same dataset [\[2\]](#). The pretrained model may be trained dynamically on the recorded data from the use of our device, which may further increase its predictive capabilities over the long term.