

How are Physiological Time Series Data used for Health Research with AI?

By Benjamin M. Fox, Jolie T. Hoang, and Joy J. Jiang

The purpose of our educational module is to introduce a general audience to how complex physiological signals, like those used in polysomnography or sleep studies, are broken down and understood by artificial intelligence (AI) models. We intend this module towards a wide audience with an interest in AI/ machine learning (ML) methods for time series data, wearables, sleep, and health. There are two major aspects of this project: (1) poster for a general audience and (2) a corresponding code Colab notebook of tokenization and visualization of time series data. We also created a video walkthrough of the notebook, with linked time stamps included in the notebook, as a supplement for viewers who prefer to learn by audio and video. We expect the audience to choose the aspects of the module that best suit their interests and provide resources for them for further learning.

The poster is meant for learners of all backgrounds, which gives an overview of using physiological time series data with AI/ML models for health research. To concretize the lessons, we choose polysomnography, or sleep studies, by way of demonstrating how various types of data can be synthesized through this pipeline. The poster view time is 1-2 minutes.

Then, the Colab notebook is for those who lack a technical background but are interested in the code underpinning the principles outlined in the poster. There is a Table of Contents of the core sections. Carefully curated annotations throughout the notebook guide the learner along each section of code. For each key section of the notebook, time stamps and links are provided to follow along in the video tutorial. The runtime of the notebook is approximately 10 minutes, with a Colab GPU. Between poster and notebook, the module will take approximately 15 minutes to complete.

For those who might want more explanation, the supplementary video reviews the Colab notebook with a didactic and interpretative perspective, again for an audience without a technical background. The video walks through the Colab notebook and elaborates on the results and visualizations to explain how AI models take raw physiological time series data and transform it into something they can process and learn from, using a real world sleep study example and pretrained model. The full video watch time is approximately 25 minutes at 1x speed, or 17 minutes at 1.5x speed for those who would like to learn more. We also included chapters in the video corresponding to sections of the code notebook, so participants can skip to sections of interest.

Submitted is the poster for the general audience. To learn more, learners can review the Colab notebook and/ or visit all or parts of the supplementary video tutorial. We provide copies of the notebook and video in our submission. However, to best visualize the graphics in the notebook, follow the link mentioned in the poster, <https://tinyurl.com/5bzm8v2>, and to link notebook timestamps to the video, open the supplementary Loom video online, <https://tinyurl.com/au87rakx>.

Overall, students will learn the following key objectives:

- Define time series encoding and tokenizing
- Outline the process for simplifying big data into small, meaningful chunks that can help prepare time series data from multiple physiological measurements for ML
- Delineate the function of a self-supervised time series encoder and tokenizer
- Examine patterns and meaning in encoded time series data and describe how this is different from tokenized data
- Explain these steps work together in AI models to predict an example outcome in sleep (sleep staging), with a supervised classifier

These various resources were specifically designed to cater to a general audience who may have varying familiarity with ML concepts. The goal is to make cutting-edge AI concepts approachable, visual, and relevant to real-world health. We are happy to provide any supporting materials, such as discussion questions or follow-up activities. Thank you for considering this educational resource. We hope it inspires curiosity in the next generation of students.