**Sagnik Mukherjee**
Applied Electronics & Instrumentation Engineering
Heritage Institute of Technology,Kolkata

# GSoC 2021: Project Proposal for INCF

## Personal Details

- **Name:** Sagnik Mukherjee
- **Email:** sagnik2207@gmail.com **or** sagnik.mukherjee.aeie22@heritageit.edu.in
- **Location(Country and City):**Country-India , City- Kolkata
- **College:** Heritage Institute of Technology,Kolkata
- **Major:** Applied Electronics & Instrumentation Engineering
- **Degree Program:** BTech (4 years)
- **GitHub:** Sagnik220
- **Linkedin:** Sagnik Mukherjee
- **Personal Website:** https://sagnik220.github.io/cashmoney/
- **Medium Blog:** https://medium.com/@sagnik2207
- **Phone number:** +91-8017262239 (also on WhatsApp)
- **Skype ID: live:.cid.7b240e9e3dff5924** (Sagnik Mukherjee)
- **Time zone:** UTC+05:30

## Background Information

- I am a third-year undergraduate student majoring in Applied Electronics & Instrumentation from Heritage Institute of Technology.
- CGPA: 8.21/10

Project Proposal

INCF

# Neurodata Without Borders projects

## Project Name:Interactive visualization of neurophysiology data in NWB Widgets

**Lead mentor:** Ben Dichter

**Backup mentor:** Ryan Ly, Oliver Ruebel

**Tags:** NWB, Python, Jupyter, ipywidgets, matplotlib, plotly

## Project Synopsis:

A neurophysiologist is a scientist, who records brain activity from humans as well as other animals and uses experiments to try to understand how the brain works (in diseased and in normal states). The goal here will be to assist neurophysiologists in using their data to more deeply understand how brains function.We will be using the results of DeepLabCut Algorithms and will be reading the results and providing interactive visualization

components that will allow scientists to explore those results across many datasets.

**Why is this project important?**

The goal of NWB is to develop a unified, extensible, open-source data format for cellular-based neurophysiology data, one of the most common and important data types in neuroscience.

The goal of this specific project are as follows:

**1.**Interactive visualization,helps researchers to gain scientific insights.

**2.**Play around with the data to draw various scientific conclusions.

**3.**Navigate data to visualize specific data elements.

**Primary Targets:**

**1.** Visualizations for videos of neural activity acquired through a microscope.

**2.** Develop interactive tools that would allow a user to navigate easily through the stages of data
processing.

**Secondary Targets**(These will be completed if time permits):

**1.** Enhance Usability for existing visualizations

**2.** Scalability of existing visualizations

## Specifications

### 1. Collecting Data :

Firstly we are going to need a video dataset to feed into DeepLabCut's markerless 3D pose estimation during behavior across species. [*DeepLabCut™ is an efficient method for 3D markerless pose estimation based on transfer learning with deep neural networks that achieves excellent results* .]

### 2. Training and Analyzing the Data:

Using the Notebooks provided by DeepLabCut we are going to feed our dataset and train and analyze the input video.

Steps involved :

- Load demo data
- Create a training set
- Train a network
- Evaluate a network
- Analyze a novel video

Analyzing the new video using the deepcutlab function analyze_videos.

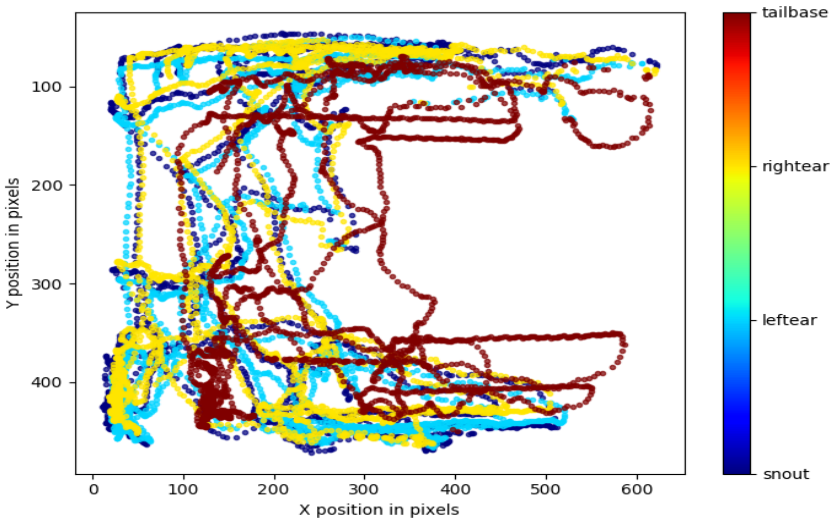
```
deeplabcut.analyze_videos(path_config_file,videofile_path, videotype='.mp4')
```

“The user can choose the best model from the evaluation results and specify the correct snapshot index for the variable **snapshot index** in the **config.yaml** file. Otherwise, by default the most recent snapshot is used to analyse the video.”

Then use functions in Deeplab Cut library, to plot the trajectories of the analyzed videos.

```
deeplabcut.plot_trajectories(path_config_file,videofile_path)
```

This function plots the trajectories of all the body parts across the entire video. Each body part is identified by a unique color.The plot belows shows the trajectory of different body parts of the mouse.



There is some correlation between the movement of body parts & neural activity. Deep learning networks are used to understand which neurons in the brain encode fine motor movements in mice.

Datasets entailing calcium imaging data of active neurons and high-resolution videos when mice perform tasks.

We want to use recent advances in deep learning to

1. Estimate the poses of mouse body parts
2. Extract behaviorally-relevant information &
3. Align them with neural activity data.

Behavioral video analysis is made possible by transfer learning, the ability to take a network that was trained on a task with a large supervised dataset and utilize it on a small supervised dataset. *[This has been **used e.g. in a human pose–estimation algorithm called DeeperCut**. Recently, such algorithms were tailored for use in the laboratory in a **Python-based toolbox known as DeepLabCut, providing a tool for high-throughput behavioral video analysis.**]*

DeepLabCut utilizes the feature detectors (ResNets + readout layers) of one of the state-of-the-art algorithms for human pose estimation called DeeperCut.

Faster variants with MobileNetV2 backbones.

3. **Implementation of the widget for videos of neural activity acquired through a microscope:**

    1. Apply network to large datasets using DeepLabCut and extract behaviorally-relevant information.

    2. Align behavioral data with neural activity data and perform cross-correlations.This will allow to understand the relationship between motor function and neural activity in specific cell-types in the brain.

    3.We can use **CaImAn** an open source tool for scalable calcium imaging data analysis.As we need to correlate the neural activity of the brain & behavior video analysis from DeepLabCuts algorithm.

```
#Class for Neural Activity Widget
class NeuralActivityPlotlyWidget(widgets.VBox):
```

This class will contains different functions/methods that I want to implement:

1. **Visualizing the movements/dynamics:**

    Considering a mouse dataset example, if the movement is fast the neuronal activity of the mouse is more in that state. Similarly if it is resting then the neuronal activity will be less. So I was thinking about making a visualization that will incorporate these 2 things plot dynamics vs neuronal activity as a widget.

First we take data from movement to classify them as fast or slow movements and then to correlate them with confusion matrices if possible, or simply plot the movements with the spikes occurred during stimulus.

2. **Visualizing the Spontaneous movement of the animal as spontaneous movement dominate cortical activity :**

Using the output data from csv of DeepLabCut , I am thinking of possibly using that data from each frame to add them over the video dataset to calculate the total distance.

As the mouse is moving, we know that spontaneous movement dominates the cortical activity so I was thinking of making a visualization that will correlate the spikes in neural activity during the movement with the distance travelled.As more the body moves more is the neural activity.

3. **Visualization for total distance,velocity,acceleration jerk:**

i) Using the output from DeepLabCut if we calculate the Euclidean distance of each feature between two frames and execute it for all the frames of a video, we can add it up for total distance.

ii) Velocity vector stores the difference between two given distances of one feature.

iii) Jerk (derivative of acceleration) measures change in acceleration through video, frame to frame.

### 4. Secondary Targets

**i) Enhance Usability:** Usability is important because it can make the users complete the task accurately. So after ending primary goals we will look into the existing codebase to enhance the existing widgets.

## Proposed Milestones

| | Dates | Task |
|---|---|---|
| **Pre-GSoC Period** | Present - May 17th | Interacting with the mentors, about the project. Setting up the nwb-jupyter-widgets repository in local and trying to understand the codebase. |
| **Community bonding** | May 17th - Jun 7 | Discuss implementation details related to the project. Create blueprints for my first task. Thus with the help of my mentor, I will become absolutely clear about my future goals.I would like to start contributing,in the community bonding period only if possible |
| **Coding Period-I** | Jun 7th-<br><br>July 12th | **Task 1:** Implementation of widget for neural activity.(4-5 weeks approximately) |

| | | |
|---|---|---|
| **Code review-I** | July 12th<br>July 16th | Review of the code submitted till date |
| **Publishing Medium Article** | July 16th-18th | After successful completion of First Evaluation,will publish an medium article citing all the experience & project work through the first half of GSoC'21. |
| **Coding Period-II** | July 16th-Aug16th | **Task 2:** Implementation of interactive tool using ipywidgets for data processing stages<br>(3-4 weeks approximately) |
| **Final Review** | Aug 16th | Submit entire code, project summaries, and final evaluations of their mentors. |
| **Publishing Medium Article** | Aug 23-<br>Aug 30th | After the final evaluation from mentor's I would like to publish a medium article showing insights of the project I did during GSoC'21 |

**Plan for communication with mentors** *How will you and the mentors keep in contact? (Via weekly Hangouts/Skype calls, via email, via chat…?)*

I am available at all the platforms and I intended to work closely with my mentor during this period by staying in touch via Hangouts/Skype/Google Meet or any other means of communication.

**Plan after GSoC'21 Period:**

Even after this GSoC is over, I plan to

- Achieve extended goals for this project.
- Remain in touch with my mentors
- Contribute to the nwb-jupyter-widgets
- I have some ideas that I would like to share & implement in future

    i) Interactive Widgets using Deep Learning and Computer Vision, for detection and segmentation of anatomical structures and the detection of lesions, such as hemorrhage, stroke, lacunes, microbleeds, metastases, aneurysms, primary brain tumors ect.

    ii) Interactive Widgets using Deep Learning to understand emotions by using the neural activity data to draw some information about neural signature.

    iii)GUI app using tkinter for neuroimaging.

**Statement of Purpose:**

I have been related to the open source world for around a year. I have always loved the concept of open source and wanted to contribute for the same.

One of my projects is related to sending resumes and mail using Speech Recognition API,where the user voice is used to write the text of the mail as the API is very inaccurate I was finding some alternatives and then I came across an Open Source Voice Recognition API named DeepSpeech I am

currently learning DeepSpeech to enhance my project as according to researchers voice data trained over DeepSpeech module can decrease the error to as less as 7%. That is when I loved the idea of Open Source & now as GSoC provides a good platform to dive right into the middle of open source development with the opportunity to work in big open source projects with the core developers. This makes me excited and eager to participate in it and give back to the open source community.

I chose this particular project because it is very well aligned with my interests and matches with the skillset. Other than that, this project provides a good opportunity to apply my learnings on a practical scale. This is my field of interest and therefore the natural inclination to this project.

**Candidate details:**

**Motivation - Why do you want to do this project?**

This project aligns with my skill set, and brings a new opportunity of learning.I am not well versed with neuroscience so I thought as my technical skill was aligning to the project I just might as well apply as GSoC is not a sprint but a marathon. During these 3 month period I hope to engage with the mentors to learn about neuroscience and their technological implementations.Most of the INCF projects align with Biomedical Engineering,which is a new domain I wanted to explore.The projects in NeuroData without Borders were mostly aligned to my current skill set and also providing me a chance to deep dive into a new domain.As I previously mentioned, I want to be a part of the community even after the GSoC ends, I have previously mentioned some of my own ideas that I want to implement

with the organization if possible.Thus I believe the opportunity of learning from mentors during the GSoC will help me contribute my own ideas to the organization.

**Coding Skills**

1.Languages: Python,C/C++

2.Framework:Tensorflow,Pytorch

3.Libraries: scikit-learn,numpy,pandas,matplotlib,plotly

**Match/Related Past Work Experience:**

I have worked with all the libraries used in this project,such as scikit-learn,numpy,pandas,matplotlib,ipywidgets etc.

So I believe that my skill set aligns with the project I believe to be a good match for this.

I have also shared my previous work related with these libraries/packages below:

**1.Digital Signal Processing:** Implemented different types of filter such as Chebyshev-I,II,Butterworth,Elliptical etc using scikit-learn,numpy,pandas & matplotlib libraries. My articles are also published @GeeksforGeeks.

**Is this the only project that you will apply for?**

Yes this is the only project I am applying for GSoC'21.

**How much time do you plan to spend on your GSoC?**

I plan to work full time 40-45 hours a week for the project, however due to uncertain conditions in India my college has not yet scheduled the exam dates, but most probably it will occur in the month of June, so during my exams (2weeks) I plan to work 30-35 hours.

**Do you have any other plans for the summer (school work, another job, planned vacation)?**

No, I have no other plans for summer, other than 2 week exams in the month of June. As I mentioned before during that period I plan to work 30-35hrs instead of regular 40-45hours.

**Preferred mode of Communication?**

I am perfectly fine with Gitter, Slack, IRC, Email, etc. I also can join any pre-planned video meetings (no timezone constraints :-) ).
My preferred language for communication is English.

**My CV:**

https://drive.google.com/file/d/12gjHgc4-bB1_IbxUQ9okwlVDWPFJHSfP/view?usp=sharing

## Resources Used:

1. CaImAn an open source tool for scalable calcium imaging data analysis
2. Combining in vivo calcium imaging datasets and deep learning networks for video analysis (DeepLabCut) to identify novel brain regulators of fine motor learning in mice
3. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning
4. Neurodata Without Borders Youtube Channel
5. DeepLabCut Youtube Channel

## Thank You!