AWS Calculator
https://calculator.s3.amazonaws.com/index.html#r=IAD&key=files/calc-5a5813e79fcb01745b627c45ca7134f21f940f48&v=ver20190926fD
Monthly cost = $830.75
Total cost over 12 months = $9,969

1. Brief description of problem to be solved.
Understanding how the brain is wired will allow medicine to address diseases that are untreatable or difficult to diagnose. A starting step is to use computational tools to build out a map of neuron connections in simple animal models, such as mice. Recent advances in neuroimaging have allowed scientists at Janelia to map out 1,000 individual neurons in the mouse brain. However, a complete characterization of connectivity will require mapping out at least 100,000 neurons. Such a task will require an automated approach to classifying and detecting neurons from high-resolution brain images. Our group (https://neurodata.io/) is working with Janelia on this data, and we hope to improve the performance of neuron detection. We will accomplish this by building an open source Python package (SaVANNA: Sparse Vectors Applied to Neural Net Architecture), which will allow other groups to use the methods we develop for their own research.

Convolutional neural networks (CNNs) are the gold standard for detecting objects within images. However, training CNNs is a very expensive operation; large amounts of training data and computations are necessary to train the network. Unlike natural images, images of the brain are not very abundant and are resource-intensive to acquire. Thus neuroscientists wishing to leverage CNNs for neuron detection tasks are often limited by small sample sizes in training sets.

Recently, improvements in alternative, cheaper, methods for image classification such as Random Forest make them more viable as a possible contender for deep models. We have already seen these improved methods outperform CNN's on smaller data sets. We believe that combining deep learning techniques with recently developed manifold oblique random forests (MORF: https://arxiv.org/abs/1909.11799) will improve classification and detection ability, creating a niche of lower-cost object detection methods which avoid significant computation.

Challenges we will address:
- Classifying and detecting neurons from high-resolution brain images (two-photon tomography)
- Building a high performance classifier that does not require large sample sizes to train

2. Proposed AWS solution (including specific AWS tools, timeline, key milestones).
AWS Tools:
Batch: for processing brain images in parallel
S3: for storing raw and processed data
EC2: for performing the computation

ECR: for storing a Docker container with our processing pipeline

Timeline and Key Milestones:

1 month: set up batch infrastructure for running our package on neuroimaging datasets.

3 months: 5+ neuroimaging datasets run on our pipeline using AWS.

6 months: open-source tool with which other neuroscientists can analyze their neuroimaging datasets.

3. Plan for sharing outcomes (tools, data, and/or resources) created during project.
All of our code is open source and will be shared on Github at github.com/NeuroDataDesign/savanna. It will be thoroughly documented and tested with Travis CI to support robust usage by the scientific community. Additionally, our tool will be made available on the PyPI and Anaconda repositories for public use and easy installation. We plan to publish a research paper on our methods/algorithm and its applications to neuroimaging data.

4. Any potential future use of AWS beyond grant duration by individual research group or broader community.

SaVANNA will continue to be made available on neurodata.io and openneuro.org to facilitate long-term, reproducible neuroscience research through the AWS cloud infrastructure. The tool will be continually supported by the Neurodata team for the scientific community.

5. Names of any AWS employees you have been in contact with (this is not a prerequisite for the application).
N/A

6. Any AWS Public Data Sets to be used in your research.
N/A

7. Keywords to facilitate proposal review.
Computational neuroscience, machine learning, computer vision, neuroimaging