

CAGI slide pack guidelines

I agree to my slides being used according to the following terms:

Laksshman Sundaram

- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 2013. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk
- Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed





Herbert Wertheim
College of Engineering
UNIVERSITY of FLORIDA

DEEP BIPOLAR

A Deep Neural network approach to predict Bipolar Disorder

Laksshman Sundaram, Rajendra Rana Bhat, Dr.Xiaolin Li

Outline

- Introduction
- Bipolar disorder & Challenge objectives
- Data pre-processing
- Intro to Deep learning
- Deep Bipolar-1
- Results
- Deep Bipolar-2
- Enhancements
- Questions

About the team

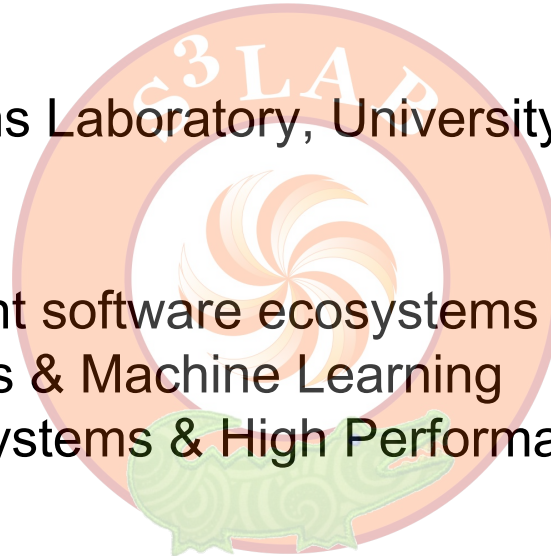
Scalable Software Systems Laboratory, University of Florida

Areas of Interest :

- Scalable and intelligent software ecosystems
- Deep Neural Networks & Machine Learning
- Cloud , Big data ecosystems & High Performance Computing

Applications :

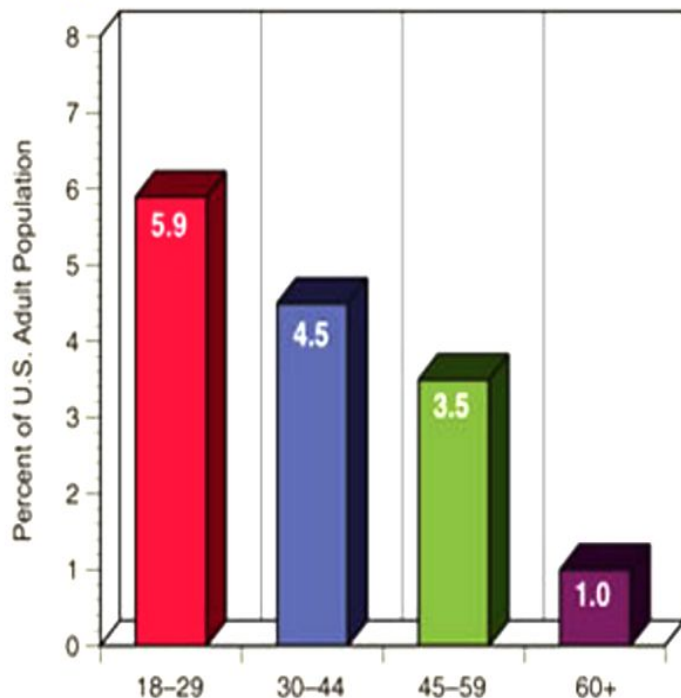
- Deep Health, Deep Sky, Deep Cyber, etc



Bipolar disorder & Challenge objectives

Demographics (for lifetime prevalence)⁵

• Age:



- Bipolar disorder is a brain disease that causes unusual mood shifts
- Estimated 51% of affected population go untreated in a given year
- Detection not straightforward - symptoms and test metrics not too dissimilar from other brain disease
- Recent studies indicate heritability and genetic factors as causes opening new area of detection using genome data.
- CAGI challenge given to predict the bipolar disorder using exomes .
- Exome sequencing data of 1000 samples with 500 for training and 500 for prediction challenge

Image source <http://www.nimh.nih.gov/health/statistics/prevalence/bipolar-disorder-among-adults.shtml>

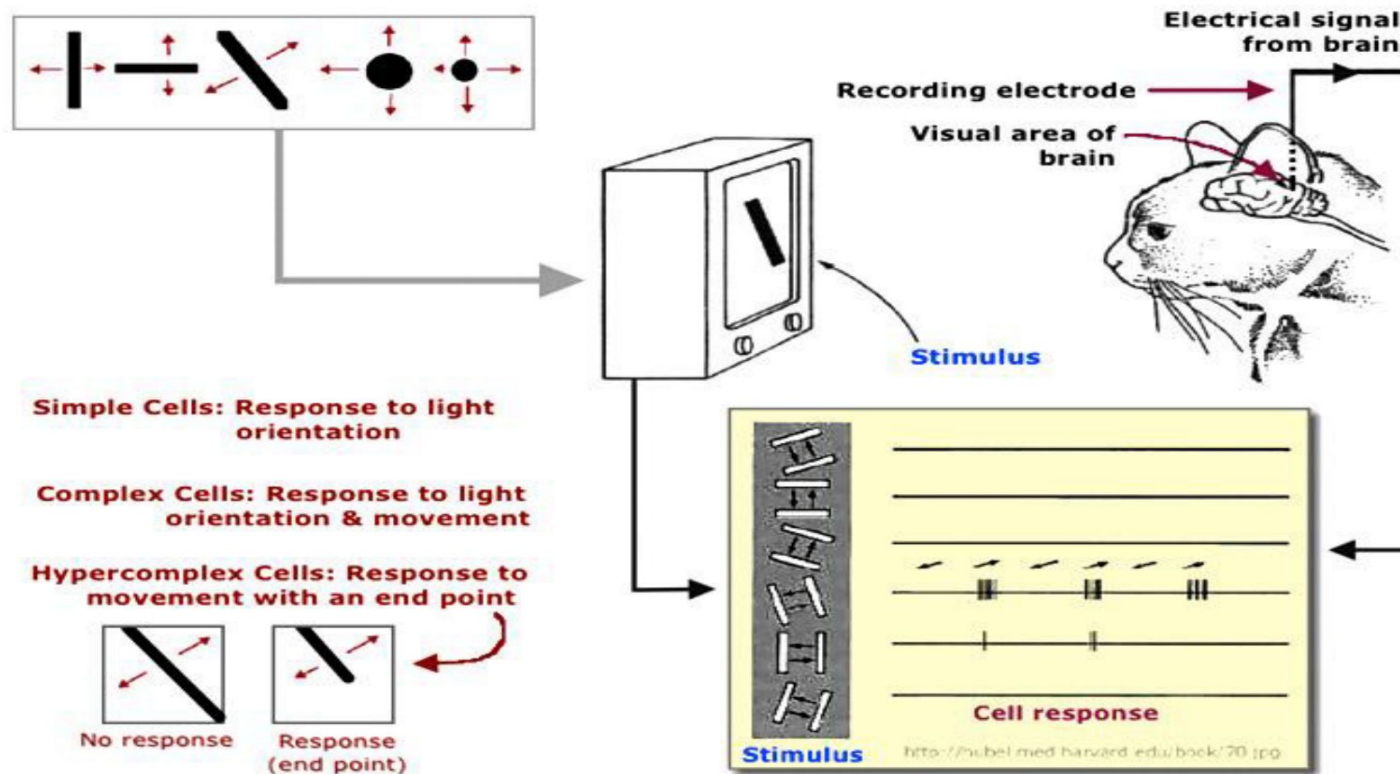
Data Pre-Processing

- Extracted genotype information from the exomes
- The genotypes were 0/0,0/1,1/1 and ./.
- One-hot-encoding transformation on the genotypes i.e 0/0 encoded as 0100,
- 0/1 encoded as 0010,etc.
- One hot encoding treats all categorical variables equidistant
- Literature references with similar pre-processing steps

[1] - Deep SEA, <http://www.nature.com/nmeth/journal/v12/n10/full/nmeth.3547.html>

[2] - DANQ, <http://biorxiv.org/content/early/2015/12/20/032821>

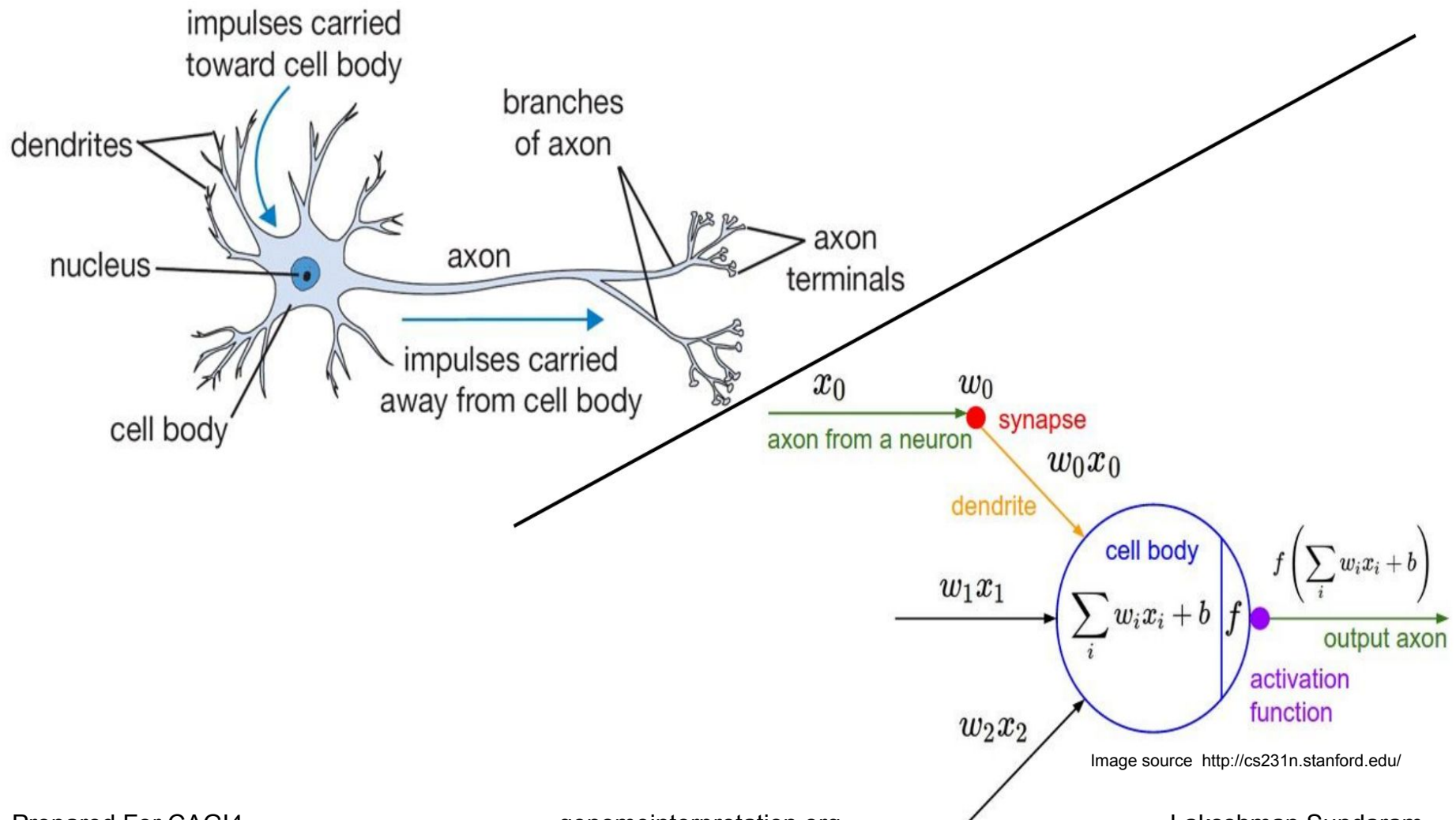
The Hubel & Wiesel Experiment



Hubel & Wiesel, 1959

Image source <http://cs231n.stanford.edu/>

The human model



Summary about Deep Neural Networks

- Deep neural networks is a hierarchical feature learning model
- Stacked layers of neurons learning particular aspects of the input
- Modelled after the structure of the cerebral cortex in brain
- Each filter produces a feature map of the the given input

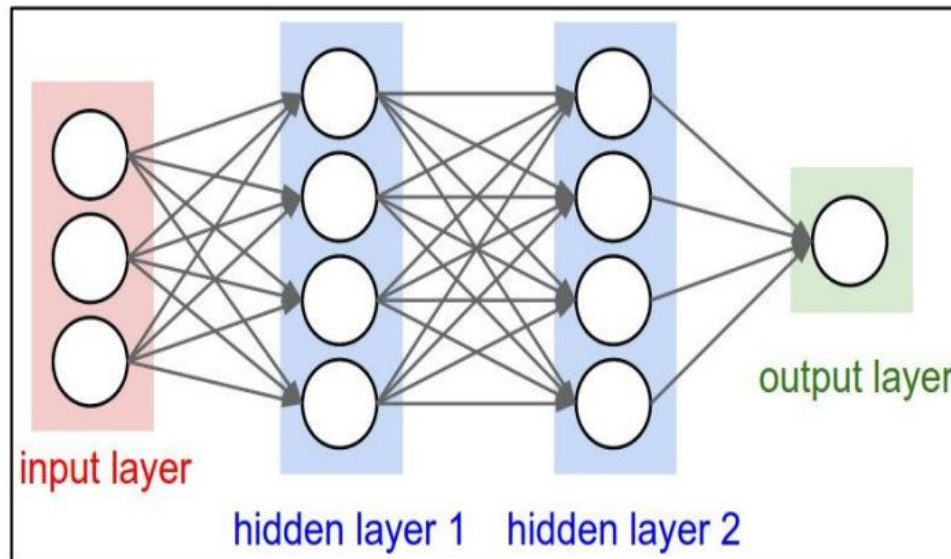
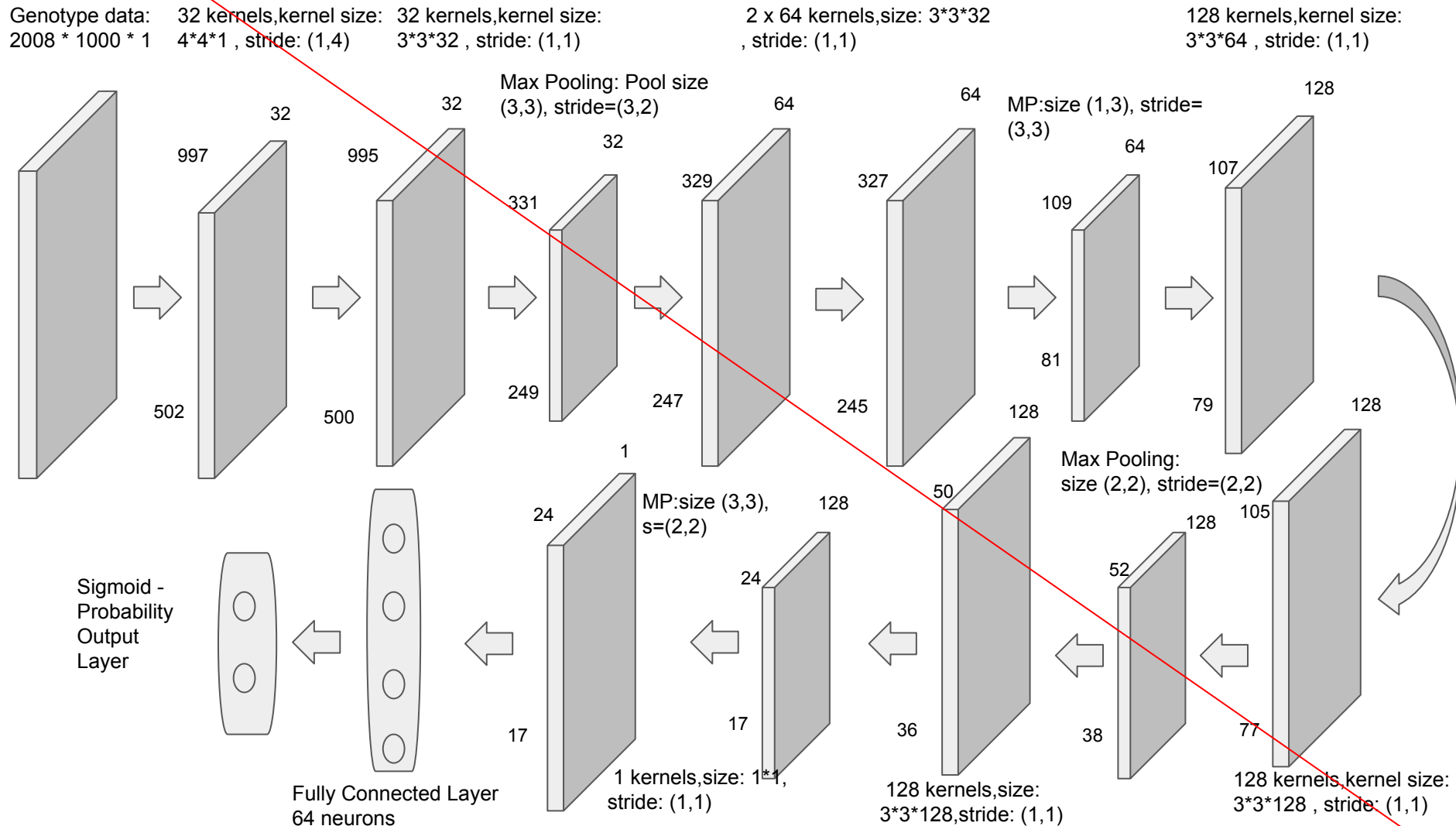


Image source <http://cs231n.stanford.edu/>

Deep Bipolar 1



Deep Learning Architecture Details

- Deep Bipolar 1 - 8 Convolutional Layers, 1 Fully Connected layer.
- Weight Regularization : L1 (Lasso)
- Activation Function : ReLU
- Parameter & Overfitting controllers : Dropouts, Batch Normalization

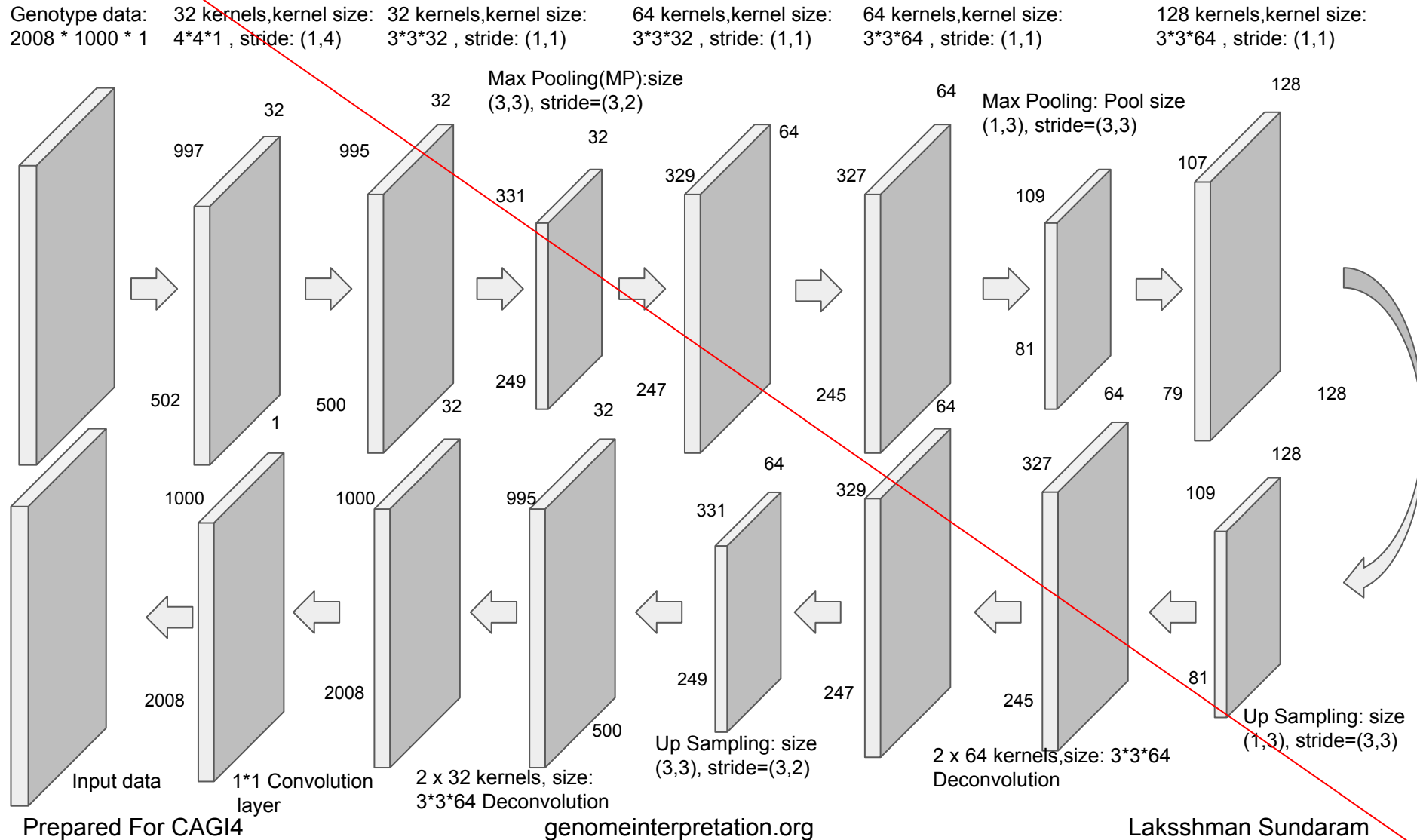
Results

- Accuracy for prediction challenge ~65%
- We got 321/500 correct predictions for the challenge. Below is the class wise split up.

Class	Original diagnosis	Our Prediction
Unaffected - 0	249	169
Bipolar Disorder -1	251	152

Our Model has similar results for both the classes

Deep Bipolar 2-Convolutional AutoEncoder



Enhancements

- Deep Bipolar 1 was the model used for the competition ~65% accuracy
- Deep Bipolar 2 - Convolutional Neural Networks with Autoencoder approach has better predictive power
- Target further refinement of the second model to achieve over 85% accuracy
- Future approaches include transfer learning
- Eradicates the problem of small sample size

Acknowledgements

- CAGI and data providers for the challenge & the data
- Dr. Xiaolin Li & Rajendra Rana Bhat for their assistance in the design of the neural network model

Thank you