A CONVOLUTIONAL NEURAL NETWORK DISTINGUISHING BETWEEN PARKINSON'S DISEASE AND HEALTHY CELLS USING MICROSCOPY IMAGES

Eddie Cai, Wolfgang Reintsch, Chanshaui Han, Sneha Shinde, Roxanne Larivière, Andrea Krahn, Carol X.Q. Chen, Emmanualle Nguyen-Renou, Anna Franco-Flores, Thomas M. Durcan, Edward A. Fon, Rhalena A. Thomas

Institut-Hôpital neurologique de Montréal
Montreal Neurological Institute-Hospital

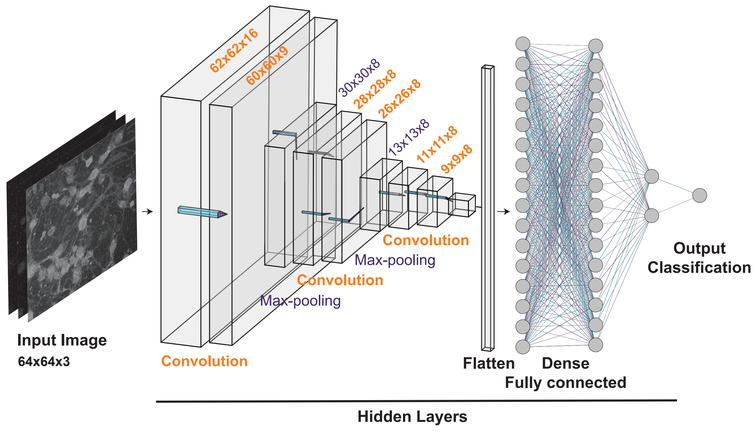
The Neuro Early Drug Discovery Unit (NeuroEDDU), Department of Neurology and Neurosurgery, Montreal Neurological Institute-Hospital, McGill University, Montreal, Quebec H3A 2B4, Canada.

McGill
Centre universitaire de santé McGill
McGill University Health Centre

BACKGROUND AND RATIONALE

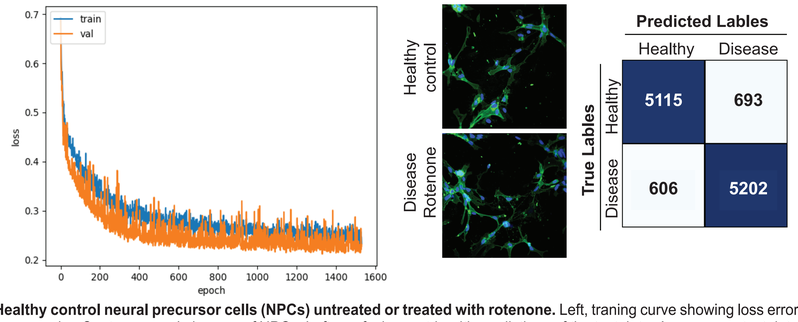
- Parkinson's disease (PD) is a neurodengerative disorder marked by loss of dopaminergic neurons in the midbrain, affecting millions of people worldwide, without disease modify treatements
- The molecular pathways leading to dopaminergic neural cell death in PD remain unclear
- Induced pluripotent stem cells (iPSCs) reprogrammed from blood or skin and differentated into dopaminergic neurons
- Measurements extracting defined features from high content microscopy images are time consuming and not reproduced well
- Feature selection is bias, doesn't account for feature combinations and misses things difficult to measure or define
- We aimed to create a classifier that can accurately classify healthy and unhealthy cells from high content immunofluorescence microscopy images using organoid dyes
- We tested the hypothesis that biological differences between PD models and healthy control cells are represented by morphological changes detectible by deep learning

CNN MODEL ARCHITECTURE



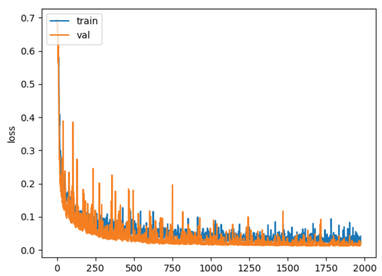
Schematic representation of the starting convolutional neural network. A simplified verson of the VGG-16 modle with 8 layers and 6042 trainable parameters. The dimensional outputs are indicated.

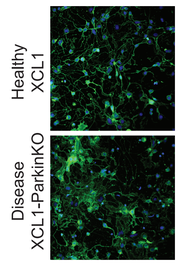
TRAINING CNN MODELS TO PREDICT DISEASE STATE

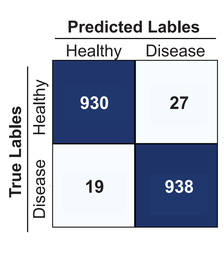


Healthy control neural precursor cells (NPCs) untreated or treated with rotenone. Left, traning curve showing loss error over epochs. Center, example images of NPCs. Left, confusion matrix with predictions of the test data. Accuracy on test data 88.8%.

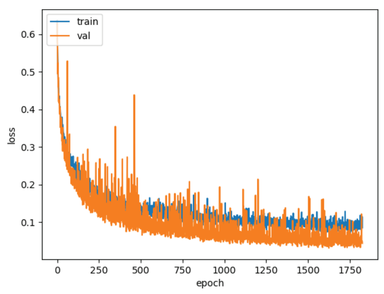
TRAINING CNN MODELS TO PREDICT DISEASE STATE

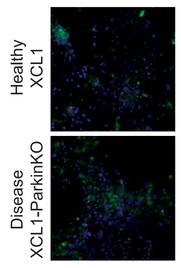


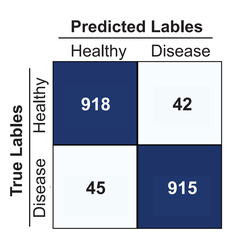




NPCs differentiated from heathly control iPSCs and PD model iPSCs. Left, traning curve showing loss error over epochs. Center, example images of NPCs. Left, confusion matrix with predictions of the test data. Accuracy of test data is 97.6%.

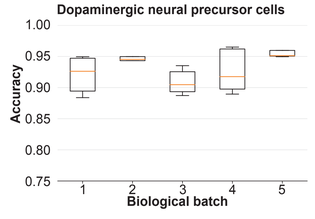


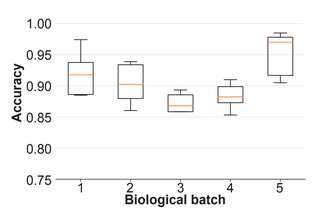


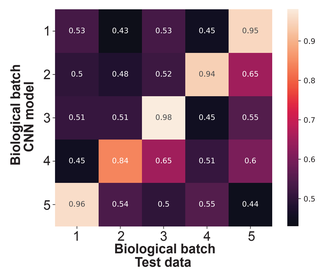


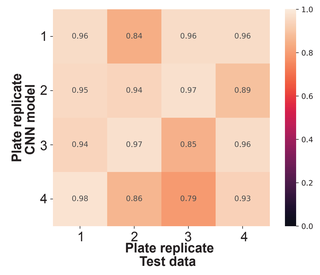
Dopaminergic neurons differentiated 4 weeks from heathly control iPSCs and PD model iPSCs. Left, traning curve showing loss error over epochs. Center, example images of NPCs. Left, confusion matrix with predictions of the test data. Accuracy of test data is 95.5%.

CNN MODELS ARE ROBUST AND REPRODUCIBLE



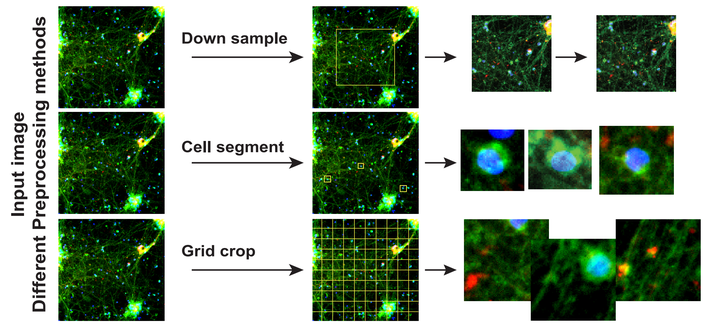


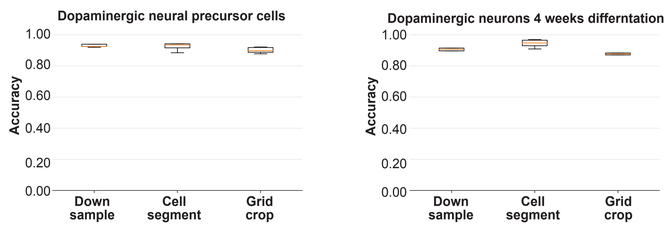




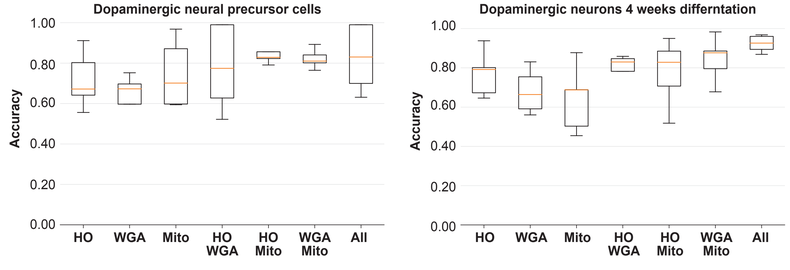
Boxplots show accuracy of models on test data. Each model was trained five times with separate spliting of the training and validation data. Five separate cultures were imaged as biological replicates. Left heatmap shows the accuracy of predicting healthy control or Parkin-KO NPCs in the indicated test batches using the CNN models trained separaetly for each batch of images. Heatmap right, predictions of test images on different plates from within one biological batch.

EXPLORING IMAGE INPUT CONDITIONS





Different image preprocessing methods all yeild accurate predictive models. Top, a schematic represenation of three different preprocessing methods tested. An example image with example of the end processed images input into the CNNs is shown. Box plots show five replications of training CNN models and testing the hidden test images with the different preprocessing methods.



Importance of channels to model accuracy. Models were trained with individual channel inputs and combinations of channels. HO (Hoechst) nuclear stain, WGA cell membrane stain, Mito (mitotracker-red), mitochondrial stain. Using all three channels produces the most accurate channels.

CONCLUSIONS

- We created a CNN architechture that accurately predicted the classification of healthy control or PD cellswhen trained with rotenone treated vs untreated NPCs, control vs parkin-ko NPCs and dopminergic neurons
- The models are highly reproducible and have low variability across replicates
- CNN models trained from one biological batch do not generalize to other biological batches
- CNN models generalizes across plates within the same biological batch indicating these models would be useful in phenotypic drug screening
- Image preprocessing methods have minimal impact on predictive accuracy
- Image input channels combinations effect the model accuracy. All channels are the most accurate indicating each channel contributes information to the predictive model.

NeuroEDDU

Fon lab