Brain disease classification using multi-channel 3D convolutional neural networks

Oral defense seminar

Andreas Christopoulos Charitos
Supervisor : Anders Eklund

LINKÖPING UNIVERSITY

LINKÖPING UNIVERSITY

LINKÖPING UNIVERSITY

Welcome to the final presentation

What will be presented?

1. Overview of the thesis problem
2. Current work status
3. Pending work
4. Conclusion
5. Final words

Thesis Overview

- The topic of this thesis was to create a classifier to distinguish between subjects with Autism Spectrum Disorder (ASD) and Control (CON)
- the data consist of 1035 fMRI 3D brain images, with 505 being subjects with ASD and 530 typical CON
- each image has been pre-processed to remove time component and unwanted noise with different methods resulting different images (derivatives) while preserving the 3D dimensions
- in the current project 9 derivatives are being considered :
  (i) alff (ii) degree binarize (iii) degree weighted
  (iv) dual regression (v) eigenvector binarize
  (vi) eigenvector weighted (vii) falff (viii) reho (ix) vmhc

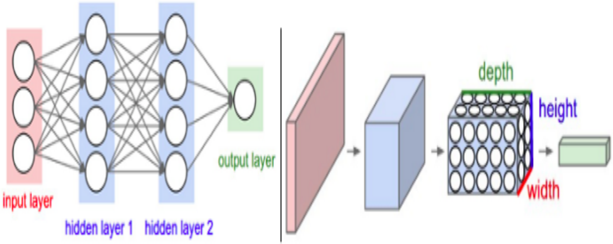
Thesis Overview

- resulting images are 3D volumes of size (61×73×61) except dual regression derivative which has (61×73×61×10)
- in total for every subject we have 19 3D image volumes those will be the inputs to training algorithm
- since our data consists of 3D images the proper algorithm for training is a Convolutional Neural Network (CNN)
- CNNs is a class of neural networks that designed to process data in the form of multidimensional arrays, like for example colored images

Thesis Overview

- the layers are organised in 3 dimensions: width, height and depth
- the main advantage over the traditional Artificial Neural Networks (ANNs) is their sparse connectivity that allows fewer connections thus less training parameters
- the neurons in one layer do not connect to all the neurons in the next layer but only to a small region of it
- that is achieved by convolving a multidimensional array (e.g. image) with smaller kernel matrices known as filters

Illustration of ANN vs CNN



Left: A regular 3-layer Neural Network. Right: A ConvNet arranges its neurons in three dimensions (width, height, depth), as visualized in one of the layers. Every layer of a ConvNet transforms the 3D input volume to a 3D output volume of neuron activations. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be 3 (Red, Green, Blue channels).

Thesis Overview

- The data are split to train/validation/test with ratios 70/15/15 and normalized to [0,1]
- the performance evaluation of the models will be conducted with the Monte Carlo (MC) dropout *(Gal and Ghahramani 2015)* with accuracy as performance metric
- MC dropout is a method that offers a more probabilistic evaluation of the model
- in simple words, MC dropout is performed by evaluating the model on the test data for a defined number of times. That gives a vector of values for the defined evaluation metric

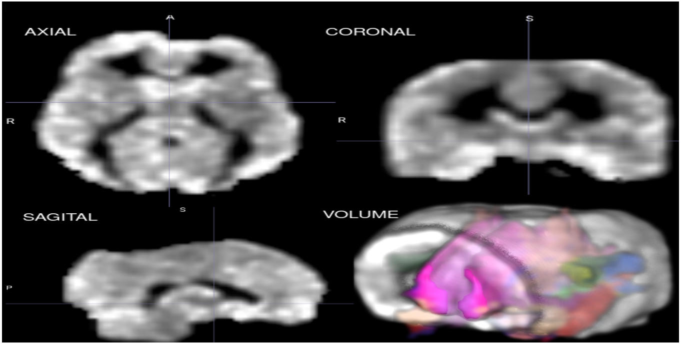


Figure: I

llustration of the three views (axial, sagital, coronal) and the volume for an fMRI image of ABIDE dataset that has processed with **eigenvector weighted** derivative for a subject with ASD.

Illustration of axial view with different derivatives for a single subject

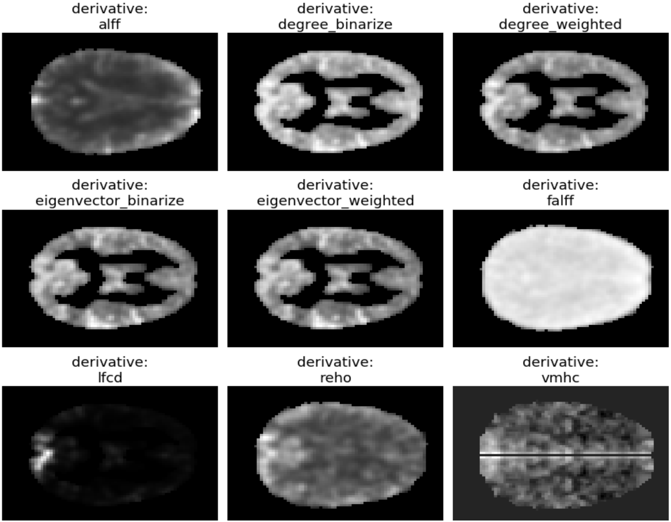
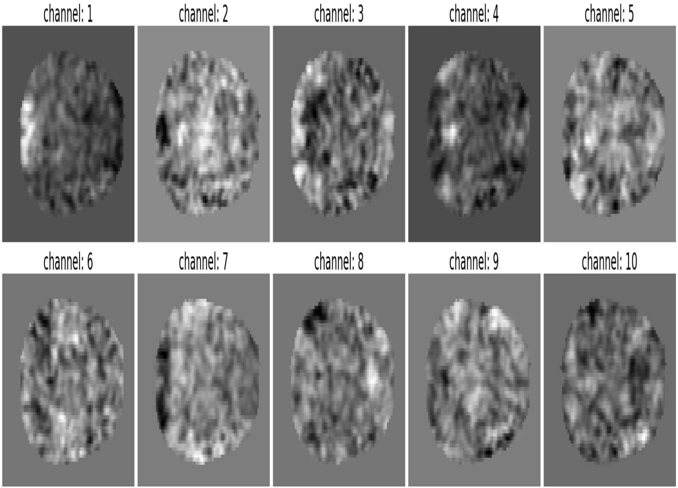


Illustration of axial view for dual regression derivative for a single subject



LINKÖPING UNIVERSITY

Questions of interest

During the project we interested in answering the following questions:

RQ1

What is the relation of the derivatives in the classification? We are interested to know if a certain derivative or derivative combination gives better classification

RQ2

What kind of 3D augmentation (e.g. rotations, scaling of each volume) are most important in the 3D CNN training?

Questions of interest

RQ3

What is the best architecture of the 3D CNN? More specifically we wish to know how many layers and filters to use. Also,what is the best regularization method for the network?

RQ4

Is the classification uncertainty reduced when training the 3D CNN with augmentation, compared to when no augmentation is used?

Questions of interest

- for answering the RQ1 we test 3 combinations of interest
- one combination with the dual regression derivative (10 channels), one with all the derivatives except dual regression (9 channels) and finally one combination with all the derivatives (19 channels)
- In addition, to answer RQ4 we test the previously mentioned combinations with and without augmentations applied
- five different augmentation methods are implemented :
  - rotation
  - flip
  - center crop
  - blur
  - elastic deformation

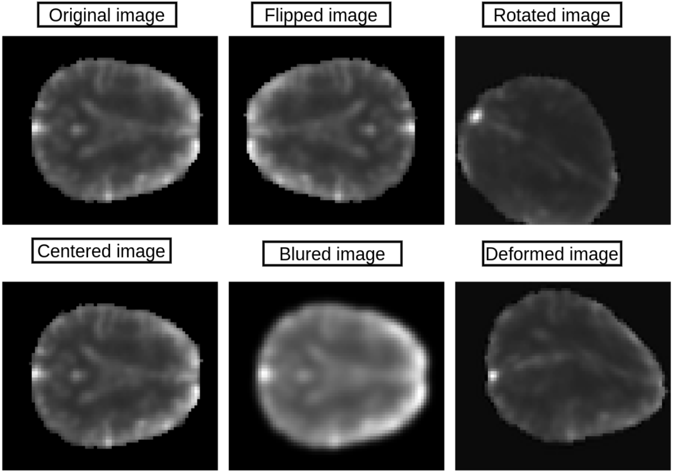
Questions of interest

- augmentations are applied only to train and validation data while test is used as hold-out for evaluation
- Also, the search for the best architecture (RQ3) is performed in the content of the grid search
- we define a search space (matrix) with discrete values to test some hyper-parameters of interest
- The hyper-parameters of interest are :
  - number of convolutional layers - nConv
  - number of filters in the convolutional layers - nFilt
  - dropout rate - dropRate
  - dense node - denseNum

Hyperparameters	**Grid search space**			
Number of filters in the first convolutional layer	32	64	128	
Dropout rate	0.4	0.5	0.6	
Nodes in the dense layer	128	256	512	1024
Number of convolutional blocks (after first convolutional layer)	2	3	4	

Table: Grid search space. The table provides the different values that will be tested for the hyper-parameters of interest.

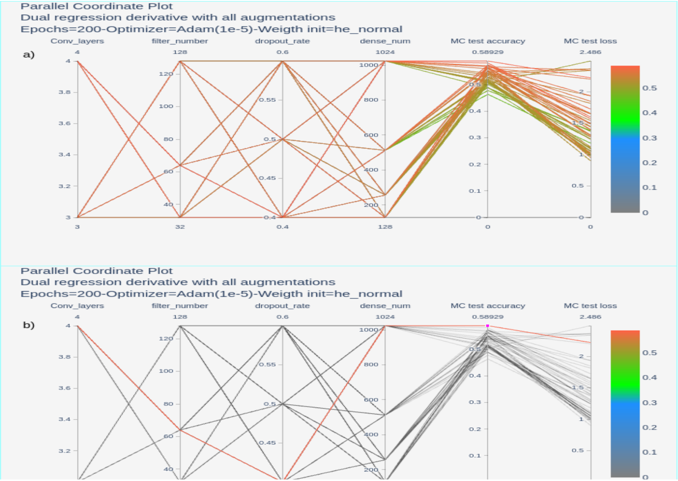
Different augmentation methods applied to a ALFF derivative fMRI image for a subject with ASD



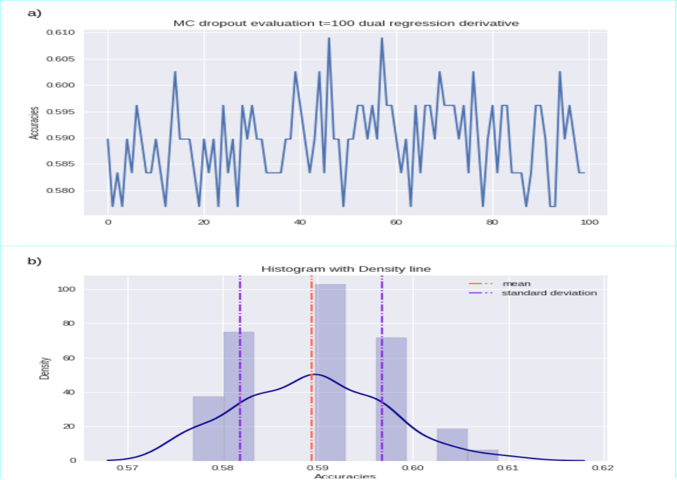
LINKÖPING UNIVERSITY

Training with augmentations

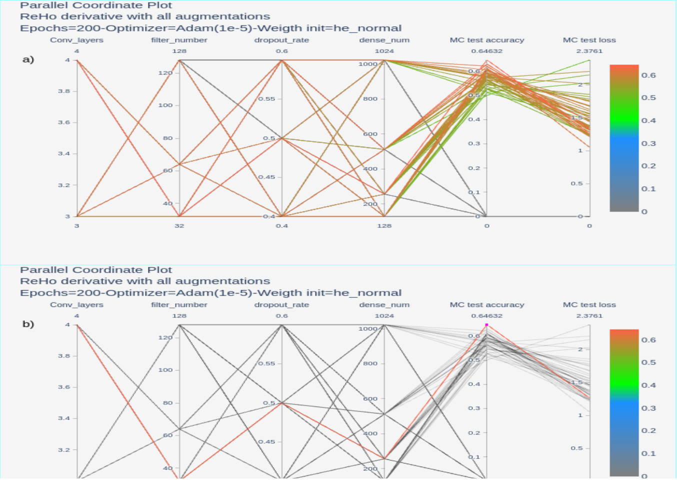
Dual regression derivative with all the augmentations



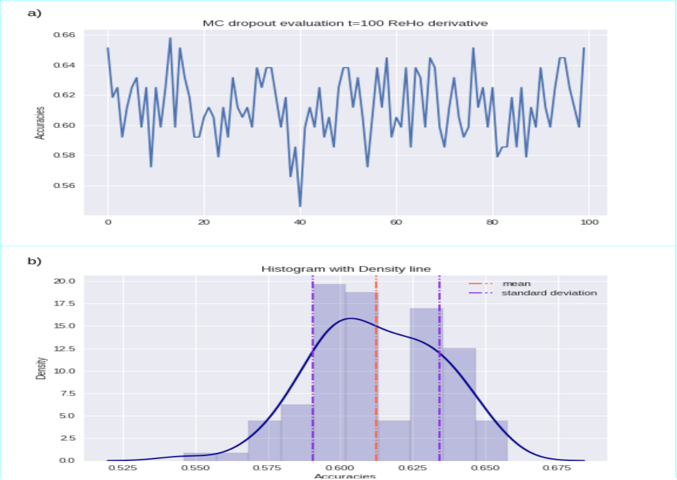
MC evaluation for dual regression derivative with all augmentations



ReHo derivative with all the augmentations

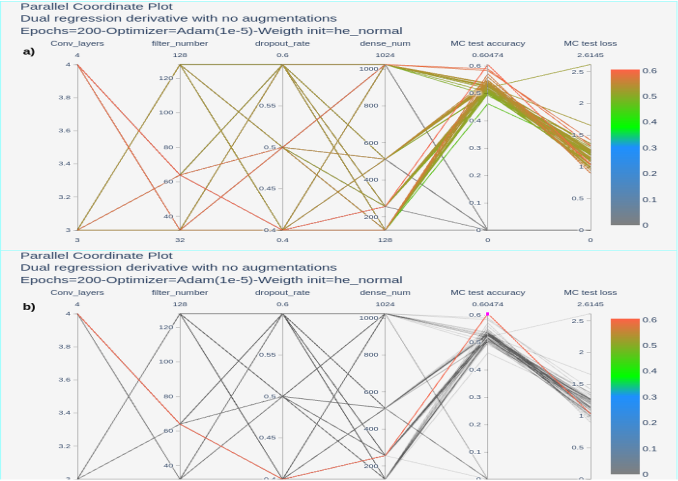


MC evaluation for ReHo derivative with all augmentations

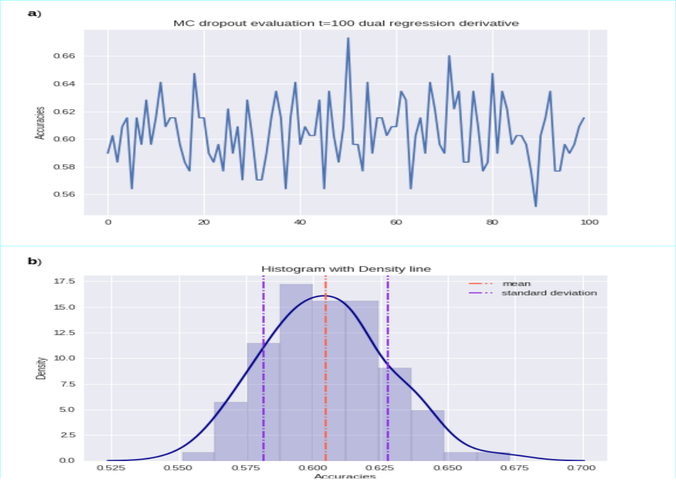


Training without augmentations

Dual regression derivative with no augmentations



MC evaluation for dual regression derivative with no augmentations

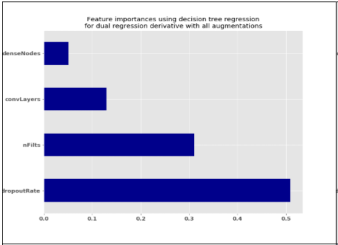


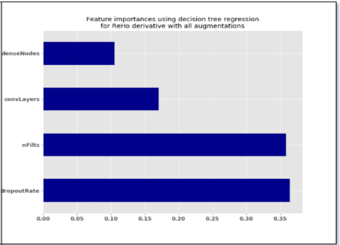
Derivatives	best accuracy	convolutional layers	filter number	dropout rate	dense nodes
Dual regression with no augmentations	0.6	4	64	0.4	256
Dual regression with all augmentations	0.58	4	64	0.4	1024
ReHo with all augmentations	0.65	4	32	0.5	256

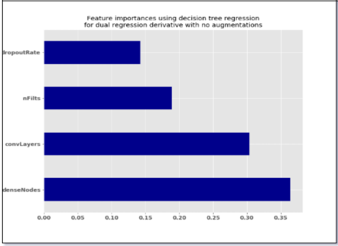
Table: Obtained accuracies for best architecture. Illustration of the accuracy achieved for each derivative combination alongside the hyper-parameters for the best architecture.

Statistical analysis-feature importance

- In order to further investigate RQ3 an analysis in the content of feature importance is performed.
- more specifically, linear regression is used with accuracy as dependent variable ($y \equiv accuracy$) and the hyper-parameters of interest as independent $(x_1, x_2, x_3, x_4) \equiv (denseNodes, convLayers, dropoutRate, nFilts)$.
- the scope of this analysis is to find which feature is more important for the accuracy.


Feature importances using decision tree regression
for dual regression derivative with all augmentations
denseNodes
convLayers
nFilts
ropoutRate
0.0
0.1
0.2
0.3
0.4
0.5


Feature importances using decision tree regression
for ReHo derivative with all augmentations
denseNodes
convLayers
nFilts
ropoutRate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35


Feature importances using decision tree regression
for dual regression derivative with no augmentations
ropoutRate
nFilts
convLayers
denseNodes
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35

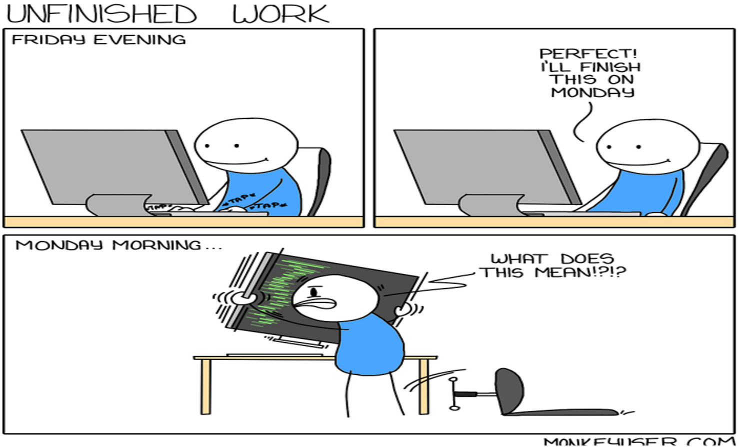
Results

	MC Evaluation 100 steps		
Derivatives	mean accuracy	standard deviation	$mean \pm sd$
Dual regression with no augmentations	0.6	0.023	0.60 ± 0.023
Dual regression with all augmentations	0.59	0.0075	0.59 ± 0.0075
ReHo with all augmentations	0.61	0.022	0.61 ± 0.022

Table: Summary results table. In this table the mean accuracy and standard deviation is reported as evaluated with the MC dropout method.

LINKÖPING
UNIVERSITY

Were are the rest of the results?



Training error

Error

During the training some errors were occurring possibly due to the fact of changing the GPU driver.

How to solve them?

Solutions

- Test another version of the software (Tensorflow)
- Perform the training on a computer with high RAM and GPU memory
- Change the model structure, the training parameters might be too many to perform training.

What is pending?

- train the rest of combinations:
  (i) all 19 derivatives with/without augmentations
  (ii) 9 derivatives (all except dual regression) with/without augmentations
  (iii) ReHo without augmentations
- perform MC evaluation with 1000 steps
- further feature impotence's analysis

LINKÖPING
UNIVERSITY

Analysis of the present results

- dropout rate is the most important hyper-parameter for training with augmentations
- densenodes is the most important hyper-parameter when training with no augmentations
- the accuracy has no significant improvement with the augmentations
- the model uncertainty seems to decrease with augmentations applied

LINKÖPING UNIVERSITY

Final words

- Perform grid search is a challenging task where many parameters involved
- augmentations have proven to increase model performance but require tuning and more data intuition
- Markov chain MC (MCMC) is another method that can be used to approximate the posterior predictive distribution and used for evaluation of the models

Acknowledgements

I would like to thank

- My supervisor Anders Eklund for his quick responses to my queries and his continuous support during the project
- my examiner Johan Alenlöv for his helpful comments
- all of my teachers at the university and especially the program coordinator Oleg Sysoev for his patience and understanding


That's all Folks!

References

Gal, Yarin and Zoubin Ghahramani (June 2015). "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of The 33rd International Conference on Machine Learning*.

Tack för idag!

www.liu.se

LINKÖPING UNIVERSITY