**Notes / bullet points reading assignment 8DM50**

*By Myrthe Boone*

* Gives a summary of the application domain of the paper
* 11% of all births are preterm, main global cause of death and disability in children < 5 years
* Preterm birth represents one of the most pervasive perinatal risk factors for atypical neurodevelopment such as anxiety, autism spectrum disorders, depression and more
* Therefore important to study / understand how preterm birth affects structural brain development and could lead to these neurodevelopmental defects
* Perinatal brain connectivity maps may give insight into the (erroneous) development of motor, cognitive and behavioral abilities in later life
* Diffusion MRI is able to depict the characterization of microstructural changes in orientation and organisation of the major white matter tracts: the connectome.
* Individual differences in the (development of) structural brain connectivity can be detected with Diffusion MRI
* However, these images are difficult to interpret and complex in nature
* Machine learning techniques as deep neural networks and random forests might be able to analyse these complex scans and uncover patterns that are not possible to detect with the human eye
* With this approach different metrics are predicted from the brain connectivity maps /neonatal connectome such as brain maturation index and gestational age as indication of demographic and neurodevelopmental characteristics.
* Finally a sensitivity analysis to obtain the most relevant features is executed. With this, the connections in brain regions contributing most to the final prediciton can be discovered.
* Gives a summary of the used (Machine Learning) methodology and evaluation metrics
* Two methodologies / approaches are used in this paper: random forests regression and deep neural networks regression.

Sensitivity analysis good to get insight to reduce black box factor of the machine learning approaches (especially DNN)

*Random Forest Regressor*

* For the random forests regression Scikit Learn RandomForestRegressor was used with MSE as loss function
* Bagging good
* Bootstrapping?
* Overfitting tried to reduce
* Cross validation less biased because of also including test set
* Class imbalance is addressed
* Hyperparameters number of trees in forest (estimators) and max depth of each tree.
* Tuned separately for PMA a scan and GA at birth predictions 🡪 grid search on independent dataset (n=73)

*Deep NN regression*

* Keras
* Basic architecture built from previous work and common DNN knowledge (hyperparameter tuning separately too computationally expensive)
  + What is common DNN knowledge?
  + Network too basic? Too basic architecture?
  + No residual blocks / skip connections
* Optimized via manual refinement architecture search
* Optimization was done on same set of 73 structural connectomes from independent sample (same as for RF) to avoid overfitting separately for GA at birth and PMA at scan prediction
* Models were trained using MSE and Adam with different learning rates

*Evaluation*

* Evaluation metric calculated on test data excluded from training and hyperparameter tuning
* 5 groups / folds
  + 20% of data used for testing each fold
  + remaining 80% : 65% training and 15% validation to tune hyperparameters
* Min max normalisation fitted on training / validation set and applied to test set
* Bias correction to correct age dependency of training residuals
* Final performance: averaging test set performance over 5 folds.
  + Mean Absolute Error (MAE) as evaluation metric
* Pearson correlation and p-value between actual Y and predicted output Y for each test set
* Heteroscedasticity in predictions 🡪 comparing variance of absolute error
* Discusses the strong and weak points of the methodology and evaluation metrics

Strong points RandomForestRegressor

* Intuitive
* Feature extraction with sensitivity analysis easy and interpretable
* No nonsense approach, baseline machine learning approach. Reproducible. Simple but doordacht. Eerste onderzoek in dit gebied dus niet meteen te specifieke architectuur.

Weak points RandomForestRegressor

* For this particular problem with high degree of complexity and only tuning number of trees and depth of each tree maybe the random forest regressor is too simple / not able to capture the complexity
* It will not be able to predict any value outside the available values since averaging is a big part of random forest models.

Strong points DNN

* Able to interpret higher dimensionality / complex data by means of adjusting each neuron’s weights during back propagation

Weak points

* Difficult to interpret / black box
* Lot of parameters to tune, not evident which architecture is the best
* Architecture itself relatively simple
* Suggests alternative methodology, evaluation metrics and ideas for improvement

Convolutional Neural Networks? No they deliberately chose not to use CNN because spatial distribution of adjacency matrices is not necessarily reflective of brain region locality and connectivity characteristics 🡪 in this study, when predicting age directly from connectome data they chose to use RF and DNN which do not require data known to have local correlations

Prediction bias still present after correlation 🡪 bias correction values fitted on training set and applied blindly to respective test sets to prevent overfitting 🡪 bias cannot be perfectly corrected. Especially for random forests

Batch normalization

Learning curves

Weights initialization

What is the goal? Interpretability? Why not just a linear regression model?

Results are good.

Common knowledge? What do they mean? Should have specifically mentioned for reproducibility e.g. weight initialization, batch normalization.

Geometric deep learning? Explain why it would be good.

GANs?

Pretrained neural network

Adversarial network

**Summary**

* Lars

**Summary machine learning**

* Lars

**Strong & Weak points + recommendations**

* Willem
* Myrthe

**Notebook week 5 afmaken (+ week 6 misschien)** : Lieke en Noortje

Myrthe: notebook week 5 online zetten

**Maandag 17 oktober 13:30 meeting:** think about question for transformers discussion and send it to Mitko