Age Regression from Brain MRI

Group: 30

Alin Fulga, Pierre El Mqirmi, Panagiotis Fytas
{af1719, pte19, pf2418}@imperial.ac.uk

1 Summary

In this report, we are exploring three different approaches to predicting the brain age using MRI data, and measure their effectiveness in minimising the mean average error. From our conducted experiments, we observe how using volumes from trained segmentations for regression gives good test results of 7.87 years MAE, using processed grey-matter maps with PCA gives better test results of 5.27 years MAE, while training the regression model on processed grey-matter maps directly gives the best test results of 4.50 years MAE.

2 Part A

For segmentation pre-processing, we have tried to modify the image sizes, which has not, however, changed the overall performance of the model. In exchange, when experimenting with different spacings, we have realised that reducing the black background helps with performance, and we have changed the image spacing to 2.7, as seen in Figure 2.

The network implemented is an adapted version of the UNet [1], having a depth of 3 down and 3 up blocks, with an initial number of filters of 8. While using a batch size of 4 images and leaving the other parameters as default, we have managed to reach a Dice Similarity Coefficient of 0.85 on the test set. The similar decrease between the training and validation loss, as observed in Figure 3, indicates that the network is not overfitting to the training set. In contrast, the continuous decrease in loss shows that the network is also not underfitting.

As it can be seen from the Box and Whisker plots in Figure 4, our model has best performed on the WM tissue, seconded by the GM tissue, while the bottleneck to the Dice Similarity Coefficient consisted mostly in the CSF tissue scores, which generally scored lowest.

With the predicted segmentations, we are extracting the relative tissue volumes, and we are performing brain age regression using multiple models, such as Linear Regression, Bayesian Ridge Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Regression Tree and Support Vector Regression. In this case, as part of our 2 fold cross-validation, we have found that the Support Vector Regression model with a level 3 polynomial kernel yields the best results, as seen in Figure 5.

In terms of the features used, we have researched combining features such as GM and WM, or eliminating WM or GM, however the results were not on par with the performance obtained when utilizing all of the features.

3 Part B

For pre-processing, we have used a 3-level Gaussian Image Pyramid followed by smoothing using Gradient Anisotropic Diffusion, as seen in Figure 2. Our choice for the Discrete Gaussian downsampling was based on the need to reduce aliasing while downsampling, which would have reduced the image quality and therefore model performance. The level of downsampling was found by experimenting with both higher and lower factors, which were however not as well-performing. We are then denoising the

image with the anisotropic infusion method to preserve the edges in the image, while also smoothing out flat regions of the image.

After experimenting with different variance preservation rates for the PCA, we have found that the default 95% yields the best results. Similarly to part A, we are using the multiple regression methods aforementioned on the processed, and then dimensionality reduced data for the 2-fold cross-validation, and the results can be visualised in Figure 5. In this case, we have found that Ridge Regression with a coefficient of 0.9 gives the best result.

4 Part C

In this part, we are using the grey-matter maps as in Part B, with an image spacing of 2.7, and train our model for 150 epochs, using a learning rate of 0.0008 and a batch size of 32. Our model uses an architecture based on the SFCN architecture [2], using 16 initial filters and a drop-out rate of 0.4. The network uses a feature extractor comprised of 6 convolutional blocks formed from sequences of `Conv3d`, `BatchNorm3d`, `MaxPool3d` (except last block) and `ReLU`. Then, the probability estimator uses a pre-defined number of age bins to determine the size of the output layer.

After analysing the age distribution of our subjects, we have created 71 age bins, corresponding to the ages between 18 and 88. Our network computes a probability for each of the age bins, and it makes the final regression prediction by multiplying the probabilities with their corresponding ages. We train the network using the `MSE` loss between the regression predictions and the real ages of the subjects.

To make the comparison with the other two approaches equivalent, we have split our training data into 2 folds, and we have alternatively trained the network on one fold, while using the other fold for validation. We have then averaged the final validation scores to compute the cross-validation MAE. When evaluating the test set, we have made sure that we have used the whole training dataset for training.

5 Age Regression Results

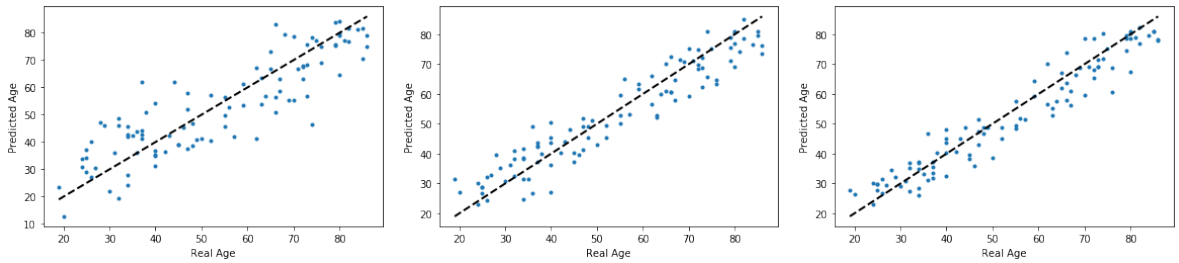

Figure 1: Scatter plots for age regression for three different approaches.

As it can be seen from Figure 1, we have obtained a continuous decrease in the mean average error from the first to the third approach. Firstly, for Part A we were able to achieve a MAE of 8.19 years for cross-validation and a MAE of 7.87 years for testing. Secondly, for Part B we were able to achieve a MAE of 5.20 years for cross-validation and a MAE of 5.27 years for testing. Finally, for Part C we were able to achieve a MAE of 5.89 years for cross-validation and a MAE of 4.5 years for testing.

From interpreting these results, we can observe that for Parts A and B, we have a small margin between the best cross-validation score and the test score, which proves the robustness of the two methods. The 2-fold cross-validation performance in Part C has a larger difference compared to the test performance than in the other two parts due to the decrease in the training data used to train the fully convolutional network on the regression task, which is more dependent on the amount of data provided.

Overall, we generally observe a better test performance than in cross-validation, which shows that our regression models have sound generalisation properties and do not overfit the data. As expected, with the increase in complexity of the model, reaching the fully-automated feature extraction in Part C, the models tend to perform better, which is a natural consequence of the capacity of more complex models to learn to generalise on input data of increased difficulty and greater detail, as in our case.

6 Figures

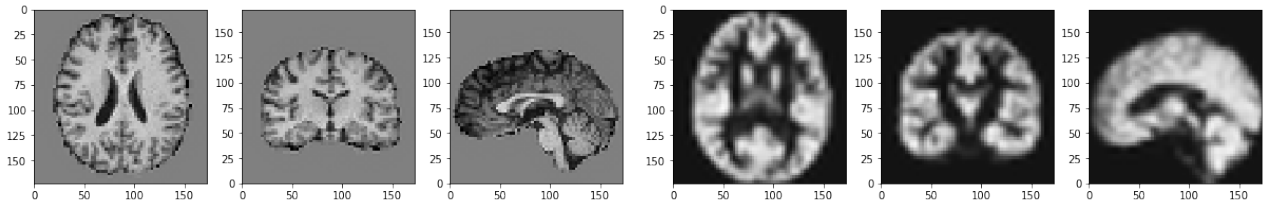

Figure 2: Pre-processing applied to MR images (part A) and grey-matter maps (parts B and C)

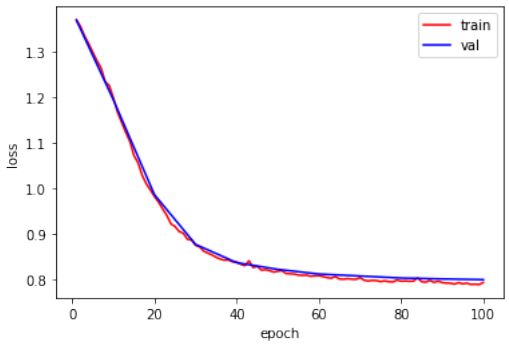

Figure 3: Loss curves during training of the segmentation model

Figure 4: Box and Whisker plots representing the Dice Similarity Coefficient per each tissue class

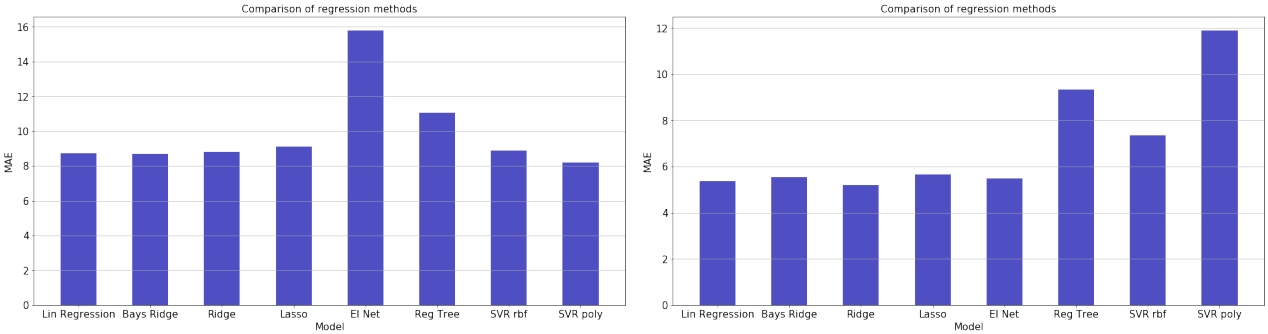

Figure 5: MAE with different regression methods for part A (left) and part B (right)

References

[1] Konstantinos Kamnitsas et al. "Ensembles of multiple models and architectures for robust brain tumour segmentation". In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 450–462.

[2] Han Peng et al. "Accurate brain age prediction with lightweight deep neural networks". In: *bioRxiv* (2019).