Machine Learning for Imaging – Coursework Report
Age Regression from Brain MRI

Group Leader: **Horsler**

Adam Horsler
{adam.horsler19}@imperial.ac.uk

1 Part A

This part consisted of the following steps:

- Data Pre-Processing
- Segmentation Model Training and Validation
- Segmentation Model Testing
- Feature Calculation
- Age regression

1.1 Data Pre-Processing

The image size chosen is (90, 90, 90) and the image spacing is (1,1,1). The same values were chosen across the spatial dimensions due to practical aspects of using CNN models. These values are a balance between spatial resolution and memory usage, as going above these values leads to memory limitations.

1.2 Model Training and Validation

The model chosen is is based off the U-Net architecture.

U-Net is commonly used in image segmentation tasks due to some advantageous properties. For example, it is able to capture both low-level and high-level features of an image through an encoder and decoder architecture. The encoder and decoder are composed of a series of convolutional layers, pooling layers, and rectified linear units (ReLUs). This pathway captures context and semantic information from the image, allowing the network to learn the high-level features important for brain tissue segmentation. The network also uses skip connections to improve learning.

Early stopping is implemented with a patience of 3 using the validation loss to prevent overfitting.

The training and validation loss curves are shown in Figure 1. Both training and validation curves show an initial large decrease in cross-entropy loss. Both losses then begin to tail off as the number of epochs goes beyond 30. Figure 2 shows 3 sets of images that clearly shows the improved segmentation ability of the model.

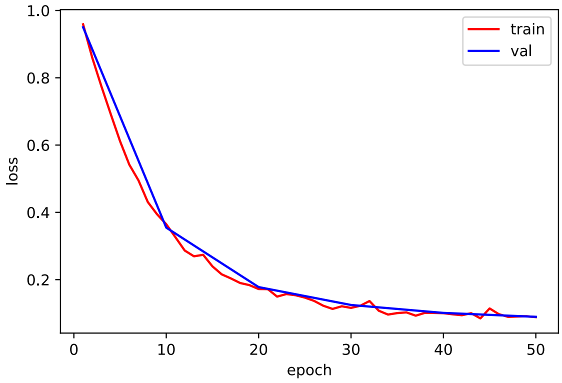

Figure 1: Segmentation Training and Validation Curves

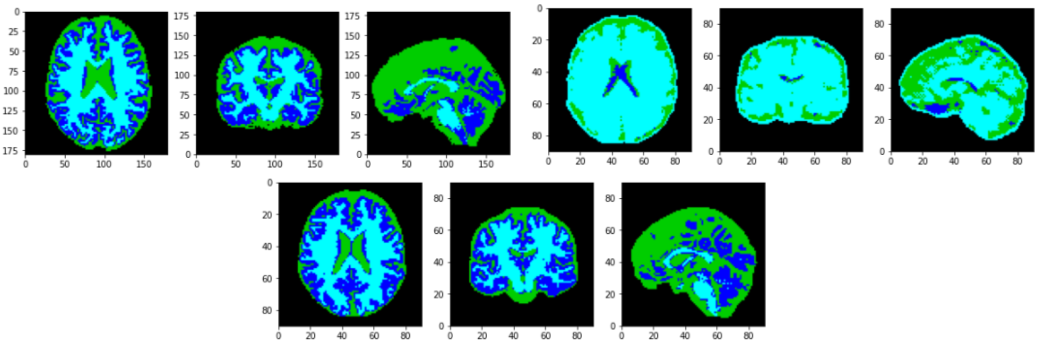

Figure 2: Segmentation Image Examples. Top left: Reference Segmentation, Top right: Segmentation Before Training, Bottom: Segmentation After Training.

1.3 Segmentation Model Testing

The test results using common segmentation and machine learning metrics is reported in Table 1. The total cross-entropy loss is 0.09.

Segmentation Error Metric	Value
Dice Score	[0.99, 0.83, 0.91, 0.94]
Mean Absolute Error	19.26
Cross Entropy Loss	0.10

Table 1: Segmentation Model Error metrics

1.4 Feature Calculation

Using the trained segmentation model, CSF, GM and WM normalised brain tissue volumes is estimated on the test set. The normalised plot is shown in Figure 3. As age increases, GM tissue volume decreases, while CSF increases. Therefore, they are useful predictors of age. The squared of these tissue volumes was also included as input features, however this did not significantly impact performance and so is omitted in this report.

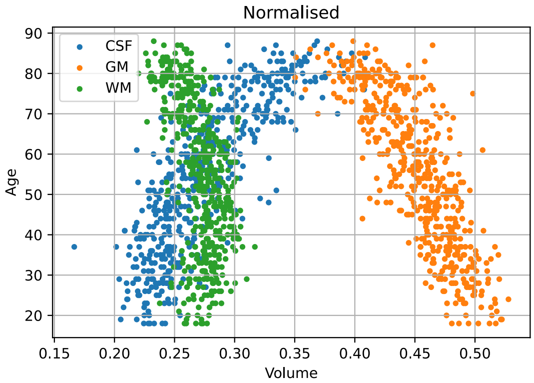

Figure 3: Brain Tissue Volumes (Normalised and Un-normalised)

1.5 Age Regression

Three regression methods are used: Decision Trees, Linear Regression and Support Vector Regression (SVR). The performance of each method is illustrated using a plot of target age against prediction age with a line of best fit. These plots are shown in Figure 4. The R_2 and MAPE metrics of each method is shown in Table 2.

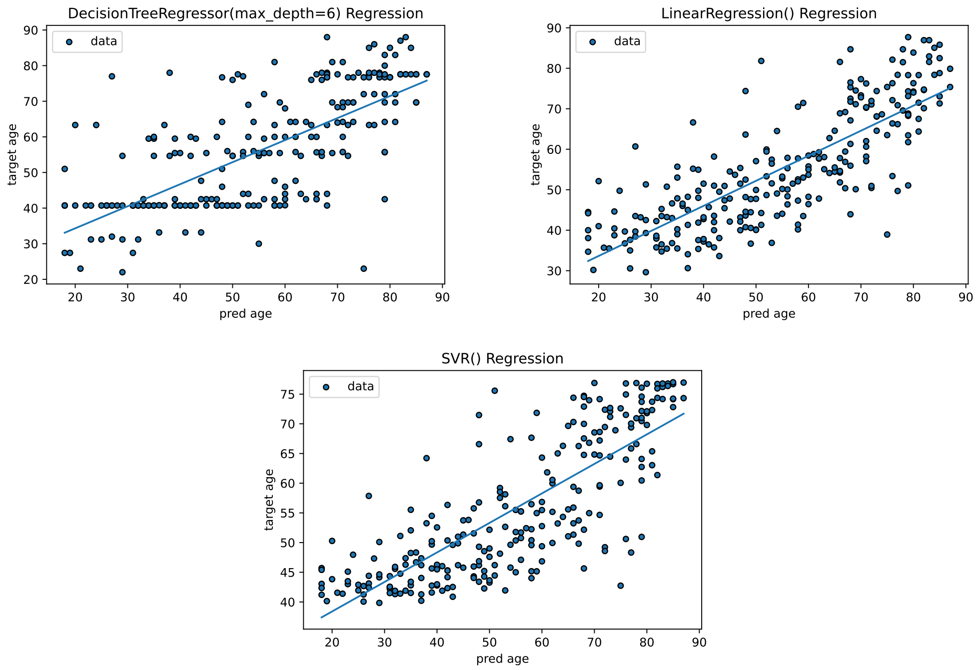

Figure 4: Age Regression Method Results

Regression Method	MAPE	R^2
Decision Tree	0.19	0.60
Linear Regression	0.18	0.68
SVR	0.24	0.57

Table 2: Regression Method Error Metrics

2 Part B

A Fully Convolutional CNN Regression model is trained to directly predict age from an MRI image. A CNN model with a final fully connected layer was also tested. The backbone architecture is the encoder architecture of the U-Net model, and the model output is a scalar. The loss functions tested were MAPE, MAE and MSE.

A similar training approach is used as in Part A, for example using early stopping and hyper-parameter selection as well as the data pre-processing from Section 1.1.

The decreasing training and validation loss curves are shown in Figure 5, illustrating that a CNN model is able to directly perform age regression from MRI brain scans.

The test set MSE and MAPE metrics are shown in Table 3. Figure 6 shows the CNN regression results on the test set as a plot similarly to the regression method plots from Part A.

The CNN regression results do not show good performance results. Due to this, several CNN models were trained, each with increasing complexity as the model may suffer from overfitting due to high number of parameters. It was found that model parameters did not greatly change the performance. Thus, it is likely due to insufficient training data (only 52 datapoints) and so the model cannot sufficiently learn the patterns for age regression in an MRI scan. Data augmentation techniques (e.g. rotation, scaling) were not used as this may affect the MRI images' medical meaning and thus lead to worse performance.

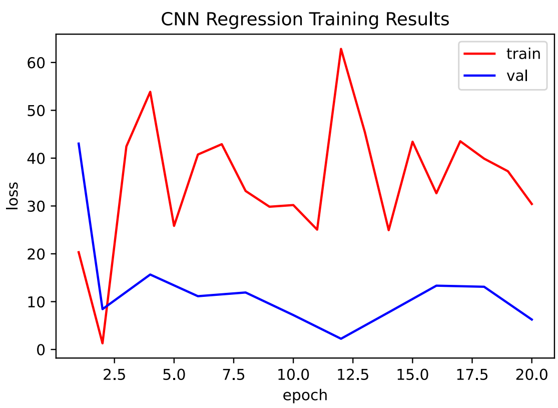

Figure 5: Regression CNN Training and Validation Curves

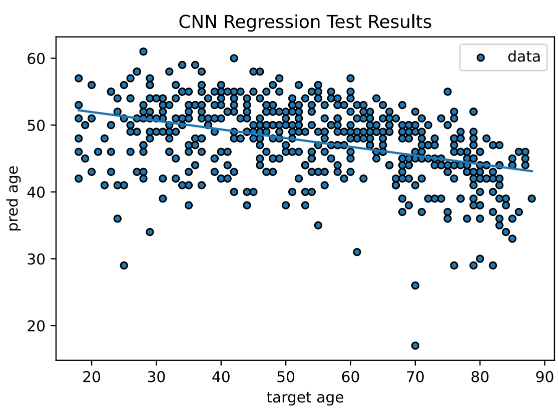

Figure 6: Regression CNN Results

3 Age Regression Results

Final testing was conducted on the hold-out set using the optimal pipelines from Part A and Part B respectively. For both parts, final training was done on the full training set (500 datapoints). The training and testing procedure for both parts also ensures there were no decisions or training/validation being done after receiving the final test set. This ensures proper evaluation of performance.

For Part A, the U-Net based architecture and linear regression model was chosen as it had optimal results compared to SVR and decision tree regression, with minimal hyperparameter tuning (reducing bias).

For Part B, the Fully Connected CNN model was chosen with MSE loss criterion as this was optimal. The final training curves are shown in Figure 8.

The regression results for Part A and B are shown in Figure 9 and Figure 10 respectively. In addition, the final test set segmentation results surpass the initial Part A results, as shown in Table 3. Despite this, the linear regression model was fitted on the full training set reference segmentation's as they represent the ground truth, ensuring optimal fitting.

The Part A performance is not very good on the final test set results as seen in Figure 9. When analysing the steps before regression prediction, the final test set CSF and GM scatter plots do not exhibit the same patterns as described in Section 1.4, thus the already fitted linear regressor predicts poorly.

Therefore, the Part B model performs better on the final test set. This is because it is able to automatically learn the important features for age prediction, instead of the manually calculated features as in Part A. However, Part B produces some unrealistic results, for example age predictions 120. This is likely due to outliers in the dataset.

The Part A method needs less training data for good performance (e.g. the 52 initial data points). However, the Part A method is more expensive as it requires ground truth segmentation's as well as ages.

Segmentation Error Metric	Value
Dice Score	[0.99, 0.83, 0.92, 0.95]
Mean Absolute Error	3.29
Cross Entropy Loss	0.077

Table 3: Segmentation Model Error metrics

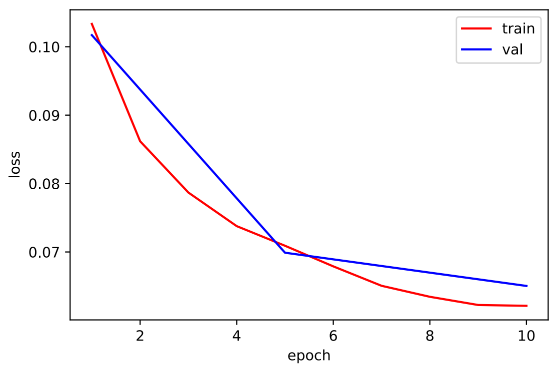

Figure 7: CNN Segmentation Training and Validation Graphs.

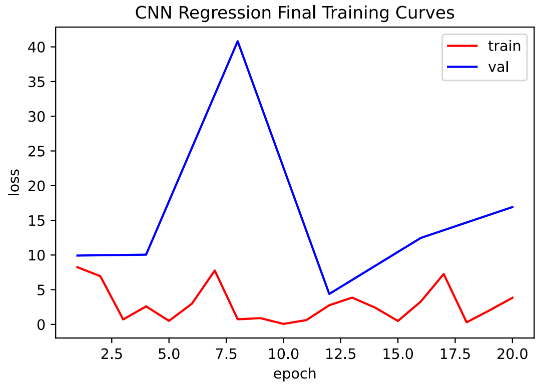

Figure 8: CNN Regression Final Training Curves

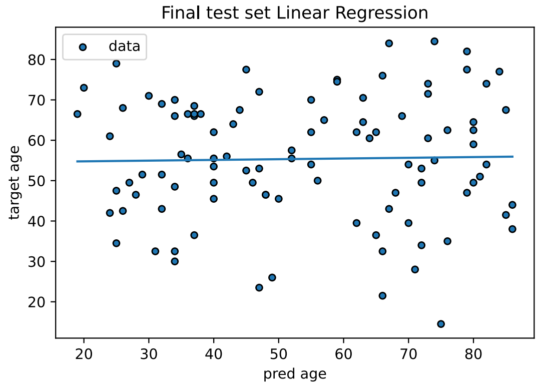

Figure 9: CNN Segmentation and Linear Regression Final Test Set Results.

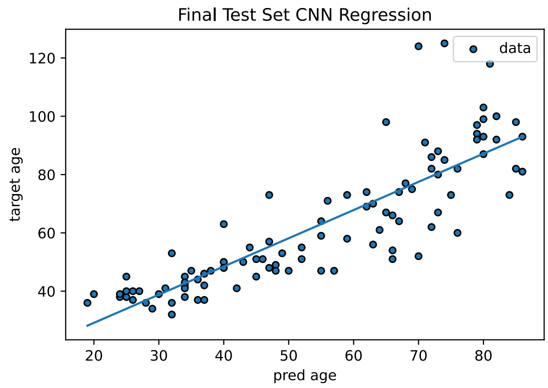

Figure 10: CNN Regression Final Test Set Results