MATH3004 Industrial Project

Semester 1, 2024

Final Project Report

Age estimation using DNA methylation

Michael Molloy

Supervised by

Professor Nicola Armstrong

Bachelor of Science (Data science)

Declaration

The work presented in this report is my own work and all references are duly acknowledged.

This work has not been submitted, in whole or in part, in respect of any academic award at Curtin University or elsewhere.

(Michael Molloy)

(21/10/24)

Contents

Introduction

DNA methylation is an epigenetic process used to influence gene expression and cellular function without changing the structure or sequence of the DNA. Methylation usually occurs when a methyl group attaches to cytosine in DNA (Moore, Le, and Fan 2012). Methylation is commonly measured using an Illumina 450k array which includes a methylation level between 0 and 1 for 450 000 methylation sites around the body.

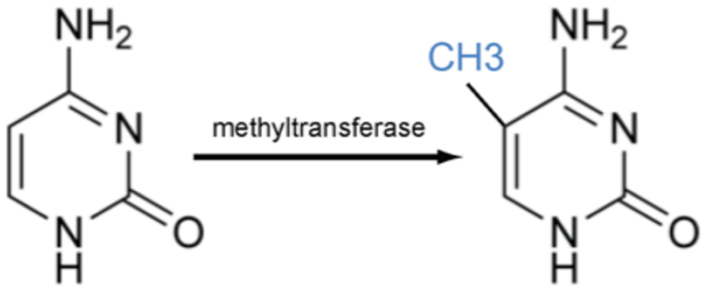


Fig 1: Methylation process (methyl group attached in blue) (Zou, n.a)

Changes in DNA methylation can occur due to several external factors including age, diet and lifestyle (Kader, Ghai, and Maharaj 2018). Since age has a significant effect on an individual's methylation level there has been increasing research into using methylation as a biological marker to predict age. Methylation has been used in recent years as a biomarker for 'epigenetic clocks' models. 'Epigenetic clocks' are models which are trained using a person's DNA methylation profile to predict their age. A person's predicted age is commonly referred to as their 'biological age' and their real age is referred to as their 'chronological age'.

Existing studies

Hannum et al. (2013) explored the link between chronological age and DNA methylation. Whole blood samples (656) were collected from individuals aged between 19 and 101. An Illumina 450k array was used to analyse the blood and give the methylation levels of the samples. The main purpose of the study was to create a model to accurately estimate age using methylation data. Elastic-net regression was used to predict age as it can be applied to high dimensional datasets with

multicollinearity due to its ability to group correlated variables and remove them as a group. The trained model used 71 methylation markers, gender and weight to predict age. The correlation between predicted age and actual age was very high with a 96% correlation and a standard error of 3.9 years. Through the training and application of the model it was observed that the methylation rate varies with gender. The model was then applied to paired healthy and cancerous tissue samples. The model predicted the age of cancerous tissue to be on average 40% older than the actual age of the person, indicating that cancerous tissue displays epigenetic features of a much older tissue.

Hyen et al. (2012) researched the epigenetic difference in methylomes between a newborn child and a 103-year-old man. This study found that there are high levels of methylation in young children which decreases over their lifetime, resulting in low levels of methylation in old age.

Zheng et al. (2024) analysed methylation data from a 450k Illumina array for two Chinese cohorts using whole blood samples. An Elastic-net regression was trained on a 250-sample dataset with an equal number of male and female samples. The DNA methylation epigenetic clock from this study "exhibited a close association with lifestyle and disease status" (Zheng et al. 2024 (para. 1)) indicating its future potential for measuring aging and different aging intervention strategies. The 'epigenetic clock' in this study achieved a correlation value of 0.97 and 0.77 on two Chinese cohorts. Three different epigenetic clock models were then tested in the model including Hovarth's clock (Horvath 2013), Hannum's clock (Hannum et al. 2013) and the 'PhenoAge' clock (Kader, Ghai, and Maharaj 2018). The 'epigenetic clock' created in the Zheng et al. (2024) study outperformed the other Elastic-net models, providing insight into the effects ethnicity may have on age predictions.

Limitations of existing studies

A limitation of the existing Hannum et al. (2012) study was their use of weight as a predictor for the age prediction model they created. The inclusion of weight as a parameter may be problematic and limiting as methylation data is not accompanied by weight data. The model created by Hannum et al. (2012) while accurate is not readily applicable. Furthermore, as weight is usually self-measured, bias can be introduced by the patient.

Objectives and significance of research

The objective of this research is to create an epigenetic clock model to predict chronological age from a person's DNA methylation profile. Elastic-net regression and neural networks will be used to model age, and the performance of both models compared.

This paper aims to produce an accurate epigenetic clock model which has significant application in research areas such as diet and lifestyle. The 'biological' age (predicted by epigenetic clock model) is important as it can be compared to a person's 'chronological' age to quantify the positive or negative effect a specific diet or lifestyle has on the overall health of a person.

This paper will provide insight into the relative prediction accuracy of elastic-net and neural networks age estimation when applied to a large methylation dataset. While there is research available that applies these methods singularly, there is no extensive research on their comparative predictive abilities when applied to the same data. The best neural network will be compared against the best elastic net regression to compare the two methods. The two neural networks will also be compared to check the dimensionality reduction capabilities of PCA vs UMAP. The ability of the two reduction methods to retain the global and local structure of the data will affect the ability of the neural network to learn and accurately predict age based on the reduced datasets.

Methodology

Data source

The data set chosen for this study includes Illumina 450k DNA methylation array data. The dataset is from the Powell et al. (2012) study which studied the epigenetics of families and the heritable similarity between family member's epigenetics. The data includes 257 samples of 450 000 variables from people between the ages of 10 and 75. The data is not independent as it includes the epigenetic profiles of families which may share heritable epigenetic traits. In this research the data is assumed to be independent as it is unlikely to introduce significant bias into age prediction.

The data will be used to train an elastic-net regression and a neural network to predict the age of a person. The data is provided in two different files with one file containing the ID of the person and all their 450 000 methylation levels. The second file contains the ID, age, gender and twin status.

Data preprocessing

The data set provided from the Powell et al. (2012) study is pre-processed to avoid bias and error from the array collection method. The Illumina 450k array was used as it is a cost-effective method to mapping epigenetics. Powell et al. (2012) described the normalisation process which was carried out with the Illumina software 'Genome Studio'. The pre-processing included the "removal of background chip effect, removal of outliers, computation of average bead signal and calculation of detection p-values using negative controls presented on the array" Powell et al. (2012). The negative values gained from the regularization process were changed to NA values as you cannot have a negative methylation percentage.

Elastic-net regression process

The next process is to apply the data to an elastic-net regression to predict age. Elastic net regression is a linear hybrid regression technique which uses the Lasso penalty and Ridge penalty to fit data to a linear model with the lowest Residual Sum of Squares (RSS). The Lasso and Ridge penalties are used to prevent overfitting with each penalty having a specific purpose.

Lasso regression fits the model by minimising a loss function which includes the Residual Sum of Squares (RSS) and the Lasso penalty as shown in equation 1. The loss function in equation 1 shows the RSS term on the left and the Lasso penalty term on the right. The Lasso penalty is a sum of parameter weights included in the model. The Lasso penalty therefore encourages a sparce model without large parameters. A model without large parameters is preferred as large parameters can cause a model to be unstable, have poor generalised predictive ability and highly variability as small changes such as noise in the data can significantly affect the model's prediction (Zou and Hastie 2005).

$$\text{Lasso Loss} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \qquad \text{Equation 1}$$

Ridge regression fits the model by minimising a loss function which includes the RSS and Ridge penalty as shown in equation 2. The loss function in equation 2 the RSS term on the left and the Ridge penalty term on the right. The Ridge penalty is the sum of the squared parameter weights included in the model. The Ridge penalty therefore produces a large model with most parameter weights shrunk close to zero. The resulting model is therefore stable and handles multicollinearity well as it can shrink correlated values to close to zero.

$$\text{Ridge Loss} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad \text{Equation 2}$$

The Elastic-net model is a hybrid regression which includes the RSS, Lasso penalty and Ridge penalty in its loss function. The loss function in equation 3 has the RSS on the left, the Lasso penalty term in the middle and the Ridge penalty term on the right. The inclusion of the two different penalties ensures the model is fit using characteristics from a combination of the Lasso and Ridge regressions. The Elastic-net regression can therefore either shrink or remove coefficients from the model making it advantageous for high dimensional datasets.

$$\text{Elastic Net Loss} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \qquad \text{Equation 3}$$

The elastic-net regression model will be trained first using the 'glmnet' package in R studio. The 'cv.glmnet' function uses 10 fold cross validation to train, test and optimise the model parameters while using only the training data. The 'cv.glmnet' function uses an alpha value to specify the ratio of ridge penalty and Lasso penalty used in the loss function to optimise the model. An alpha value of 0 produces a ridge regression and an alpha value of 1 gives a Lasso regression with any value between producing an elastic-net regression with a mix of both models.

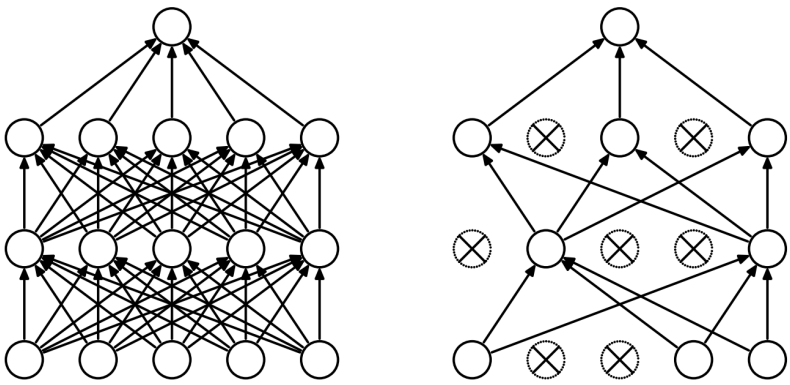
Elastic-net regression is the common choice for methylation data as it handles high dimensional datasets with multicollinearity well. As explained previously Hannum et al. (2013) and Zheng et al. (2024) achieved impressive accuracy using Elastic-net regressions. Furthermore Elastic-net regression is a common choice in genomic analysis due to the interpretability of the models it creates. As Elastic-net regression can only remove and shrink parameters to fit its model therefore the resulting model will include the methylation sites around the body which are most correlated to aging. The model is therefore significantly easier to interpret unlike other machine learning techniques such as neural networks which can contain hundreds of thousands of neurons.

Neural network process

A neural network contains multiple layers of full connected neurons. The neurons optimise their parameter weight as training data is forward and backward propagated through the network in the training of the network. The first layer called the input layer has a neuron for each feature in the dataset. The input layer is then connected to the first hidden layer; a neural network can have one or many hidden layers. The last hidden layer is then connected to the output layer which produces the model's final prediction. Each neuron is input with data, it then applies its parameter weight and bias before passing the result to an activation function which introduces non-linearity into the model (Sarita, 2023).

A neural network is trained by through forward and backward propagation. The training data is first entered into the network with random neuron parameter weights. The output of the randomly weighted model is then compared to the target outcome using a loss function (Roy, 2022). Stochastic Gradient Descent is then used to optimise the gradient function and therefore change the parameter weights and improve its predictive accuracy (Brownlee, 2021). The process described above including forward propagation, loss calculation and backward propagation is repeated for each datapoint in the training set is called an epoch.

The dropout regularization technique will be used in the training of the neural networks to prevent overfitting. As shown in figure 2 dropout technique randomly removes a percentage of nodes from each hidden layer before each epoch in the training of the network. The removed nodes parameter values are not updated for the specific epochs in which they are removed. The random removal of the nodes prevents the model from overfitting to the data and prevents the model relying mostly on specific nodes in the network for its predictions. The model trained with dropout regularization has better generalised predictive abilities (Gal et al. 2016)



(a) Standard Neural Net (b) After applying dropout.

Fig 2: Dropout regularization diagram (Srivastava et al. 2014)

Neural networks can be applied to methylation data due to their ability to handle large complex datasets and learn non-linear relationships between variables. While Neural networks have many beneficial features, the models they produce are not interpretable and may require a significant amount of data to train due to the large number of neurons in the model.

Dimensionality reduction

Dimensionality reduction is often used to reduce the number of input features therefore decreasing the amount of data needed to train the network as the larger a neural network the more samples needed to fully train all the parameters in the model. Two commonly used dimensionality techniques are Principal Component Analysis (PCA) and Uniform Manifold Approximation (UMAP). The two different techniques both aim to capture the variance of the data in a lower dimension while retaining local and global structures, however, they use different mathematical method to achieve this.

Principal component analysis

Prior to PCA data must be standardised (values between 0 and 1) to ensure all variables are of the same magnitude. A covariance matrix is then calculated including all parameters showing the relationships between parameters. The eigenvectors and eigenvalues of the covariance matrix are then found by solving the characteristic equation shown in equation 4.

$$\det(C - \lambda I) = 0 \qquad \text{Equation 4}$$

The eigenvalues indicate the amount of variance captured by the PCA component. The eigenvalues are ordered in descending order so that the first components account for more variance in the original data compared to later components. The PCA can be used as the entire dataset or a subset of the first components can be used (Abdi et al. 2010).

Uniform manifold approximation

UMAP dimensionality reduction first creates a weighted k-Nearest neighbours' graph with all points. Each point is connected to 'n' number of other points, with closer points having larger weights then further points. The weighted graph is then conversed into a 'fuzzy simplicial set' representing local connectivity of points in the original high dimensional space. Each connection is treated probabilistically as UMAP uses a 'softmax' activation function to define probability of connection for each connection. The fuzzy simplicial set is then optimised into a lower dimension using the Kullback-Leiber (KL) divergence measure shown below which measures the different of probability distributions between low and high dimensional representation of the data. The KL formula (Equation 5) calculates the loss of information when data is reduced from P original dimensions to Q reduced dimensions.

$$\text{KL Loss} = D_{KL}(P\|Q) = \sum_{x} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \qquad \text{Equation 5}$$

After the KL divergence measure has optimised the structure of different points to minimise the divergence measure while generating a low-dimensional imbedding of the original data points. The final UMAP reduction can be modified through different parameter values.

The 'n_neighbours' parameter changes the number of neighbours each point connects to when UMAP initially creates the initial high dimensional graph. A small 'n_neighbour' parameter will focus on the local structure of the data as each point will only consider the closest points to itself. A large n_neighbour parameter will preserve the global structure however there may be loss of some local features. The next parameter is min_dist which is the minimum distance between points in the low dimensional space. A large 'min_dist' will have loosely clustered points and a small 'min_dist' will have tightly clustered points.

Results and analysis

Data preparation

Initially the data was transformed from a 'data frame' in R studio to a 'numeric matrix'. Data frames in R Studio are useful as they allow for the simple storage of multiple datatypes with easy access and modification however it may not be the best option for large datasets. The methylation data from the 257 samples used 8Gb of RAM to store as a data frame which made any calculations or modification of the data time consuming. The same data when transformed to a numeric matrix only used 1Gb of RAM to store therefore making analysis of the dataset faster and more practical.

Data analysis

The missing values in the data were then analysed finding 0.24% of the matrix contained missing values. As Hannum et. al (2013) showed that the gender of a person effects their methylation profile missing values were imputed depending on gender.

The data was then analysed to check the distribution of ages in the data. As the data was from a Powell et al. (2012) study on heritable traits in methylation there are two obvious peaks in the age distributions which represents parents and their children. The lack of data for people between the ages of 20 and 35 might cause problems with model generalisation as the model may not learn to predict 20-35 year olds age accurately. As shown in figure 3 there is one person with an age of 75 which is significantly older than the other samples. After analysis the 75-year-old data point was removed as it may interfere with minmax scaling which is applied later in the analysis.

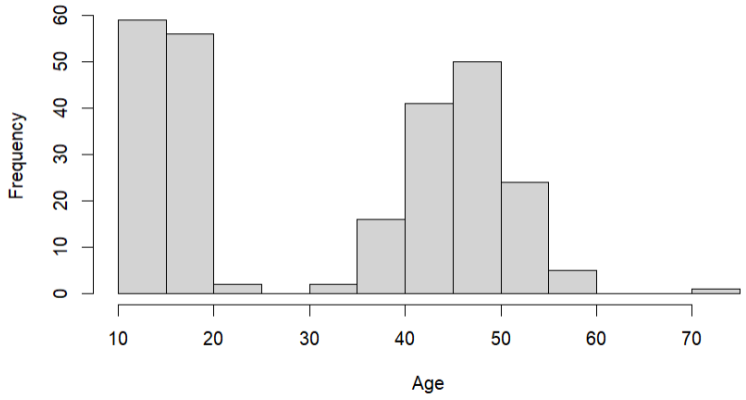


Fig 3: Age distribution in samples

The data was split into a test set and a train set to allow for unbiased testing of the trained models. The train set contained 80% of the data and the test set contained 20% of the data.

PCA dimensionality reduction

Principal component analysis was first applied and produced 255 principal components. The PCA plot of the first two components in figure 4 displayed a significant male outlier that had a PCA 2 component of over 30 which was significantly larger than the remaining samples.

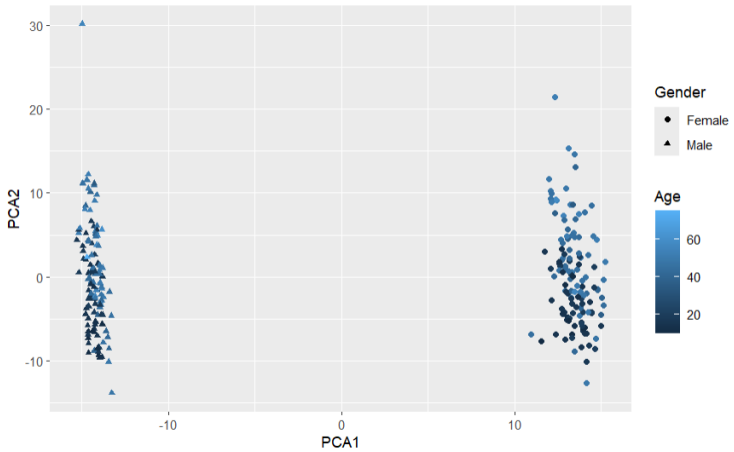


Fig 4: Plot of the first two PCA components

The outlier was removed from the dataset and the PCA analysis was completed again with no outliers present. The PCA data was then transformed using the 'MinMaxScaler' function from the 'sklearn' package in python. Scaling was applied to the data as neural networks trained with features of the same scale have improved gradient descent in the training of a neural network resulting in a faster convergence (Al-Faiz et al. 2018).

PCA neural network training

Initially a neural network with 4 hidden layers was trained will all 255 principal components. The neural network prediction accuracy was then tested with new data resulting in a root mean square error (RMSE) of 24.07 meaning that on average the predicted age was 24 years different then the real age. The inclusion of all PCA components therefore overfits the train data to the neural network resulting in poor generalised predictive ability.

Several neural networks, each with different numbers of principal components (ranging from 40 – 200), were trained. The accuracy of each trained model is presented in table 1.

Table 1: Accuracy of PCA trained neural networks with varying number PCA component inputs

PCA component number	RMSE	MSE	R squared correlation *
200	13.55	183.59	-0.592
150	10.52	110.66	0.058
125	9.53	96.43	0.103
100	6.62	43.81	0.808
90	6.59	43.45	0.763
80	8.47	71.73	0.639
60	5.378	32.38	0.880
50	5.85	34.23	0.873
40	7.60	57.81	0.701

*Adjusted R squared was not used as most models had more predictors than observations in the test set which resulted in invalid adjusted r squared values greater than 1.

The predicted accuracy of the best model (60 PCA components) was then plotted against real age below in figure 5. The best PCA trained neural network performed accurately on younger people, however it was less accurate when predicting the age of older people. The best PCA trained neural network relied on a neural network with a 60 neural input layer, 4 hidden layers decreasing from 40 to 5, then an output layer. Three dropout layers were implemented between the first four layers to ensure the neural network did not overtrain. The optimal model achieved a standard error of 5.378 years.

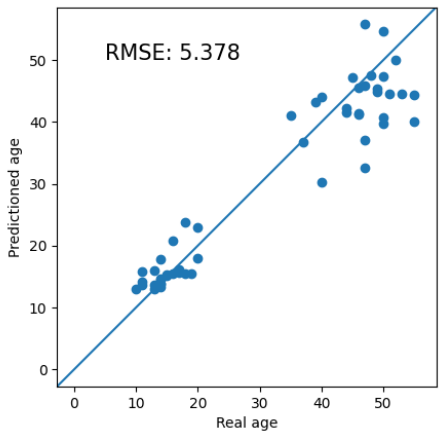


Fig 5: PCA neural network predicted age and real age plot

UMAP dimensionality reduction

UMAP dimensionality reduction was applied to the data. The UMAP transformation was computationally expensive therefore 16 features were the maximum number of features it could produce. Through experimentation and testing a value of 20 neighbours was found to give the best results allowing both global and local features to be preserved in the reduced data. The optimal minimum distance was found to be 0.1. The resulting lower-dimensional dataset showed significant separation between gender and some separation of different ages as shown in figure 6.

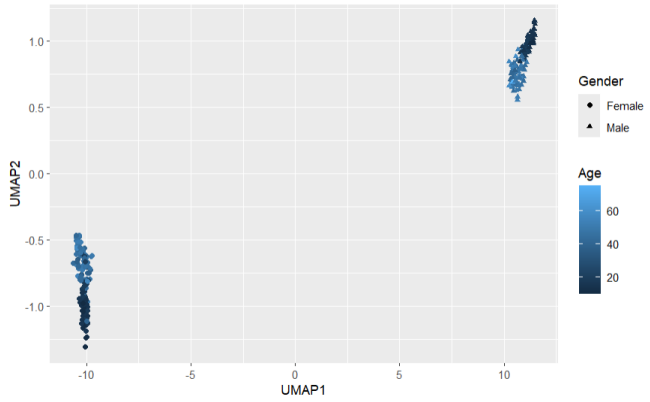


Fig 6: Plot of the first two UMAP components

UMAP neural network

Several neural networks were trained using between 6 and 16 UMAP components as input variables into the neural network as shown below in table 2.

Table 2: Accuracy of UMAP trained neural networks with varying number of UMAP component inputs

UMAP component number	RMSE	MSE	R squared correlation
16	9.08	82.43	0.695
14	10.7	116.23	0.569
12	10.44	108.94	0.596
10	9.40	97.04	0.700
8	10.88	118.30	0.593
6	10.25	105.13	0.611

The best UMAP trained neural network used all 16 components and achieved a standard error of 9.08 years and an R squared value of 0.695 and presented in table 2. The best UMAP trained neural network relied on a neural network with a 16 neural input layer, 4 hidden layers decreasing from 14 to 5, then an output layer. Three dropout layers were implemented between the first four layers. The UMAP neural network did not fit the data well with the model predicting younger people older than their age and older people younger than their age (figure 7).

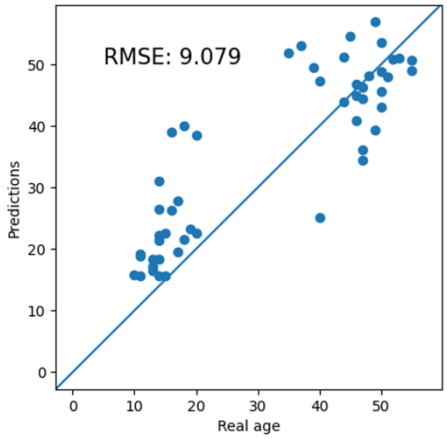


Fig 7: UMAP neural network predicted age and real age plot

Elastic-net model

Eleven different elastic-net models were trained using a range of alpha values between and 0 and 1. The models were tested using the test dataset to check their predictive ability. As expected the models with a lasso penalty component performed better due to the ability to remove variables from the large input number (figure 8). The ridge regression (alpha = 0) as predicted performed significantly less accurately than all other models as it could not remove any variables. The low predictive accuracy of the model indicates the model overfit to the training data due to the large number of input parameters.

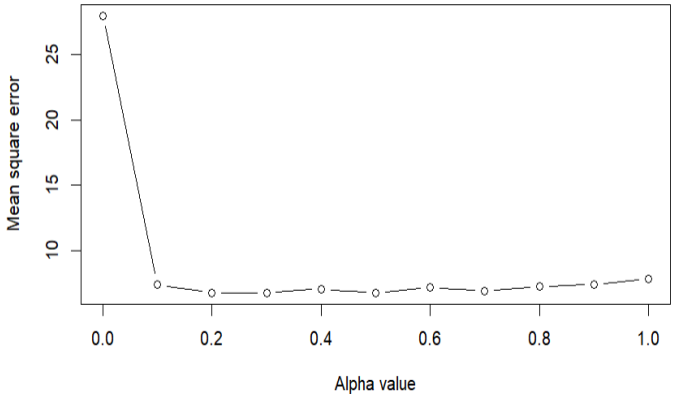


Fig 8: Plot of MSE of elastic-net models with different alpha values

The most accurate model used an alpha value of 0.5 (figure 8). An alpha value of 0.5 meaning the model's loss function was half a ridge regression penalty and half a lasso penalty. The optimal model achieved a MSE of 6.784, a standard error of 2.605 and an R squared coefficient of 0.975. It used 102 different methylation sites to predict age. The predictions and real age values are visually presented in figure 9. The model was very accurate at predicting young people between the ages 10 to 20, however it was marginally less accurate at predicting older people between ages 35 to 60.

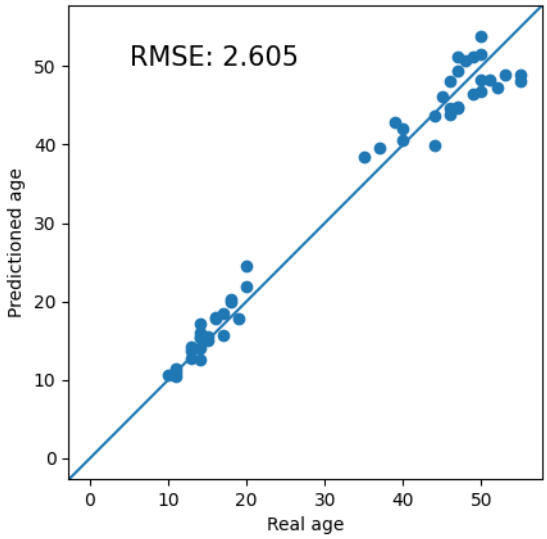


Fig 9: Elastic-net predicted age and real age plot

Comparing the models

The three models' age predictions were plotted together on the same graph and against the real age data (figure 10). The models varied in their accuracy for the different age groups in the test data.

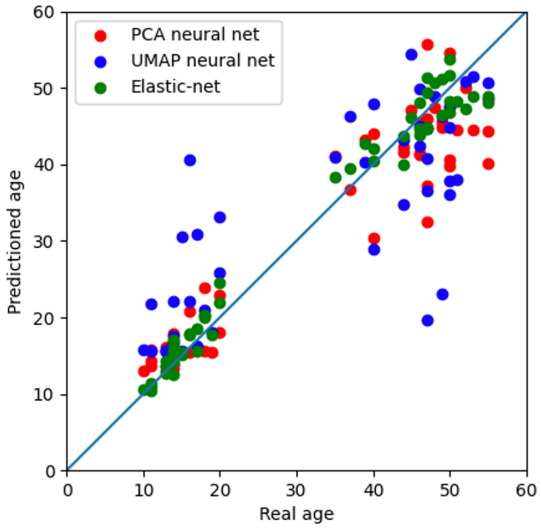


Fig 10: Plot of all the predictions made by the best model of each type

The data contains two major age groups in the data with a young group and an old group. The PCA neural network and elastic-net were both able to accurately predict the age of younger people as shown in figure 10. The UMAP trained neural network did not predict the age of the young cohort accurately with most predictions being older than their real age. The accuracy of both neural networks reduced significantly when predicting the age of the older cohort. The elastic-net model outperformed both neural network models with very accurate predictions in the older cohort, and with only a small number of predictions having any significant error.

Model implementation

The neural network models allowed customisation of the models through the optimisation of multiple parameters. Optimising all the parameters however significantly increased the implementation and training time. The elastic model did not allow for customisation apart from giving an alpha value for the ratio of the two penalties and a lambda value for the magnitude of the penalties. The elastic net model was significantly easier to implement and train when compared to the two neural network models. The elastic net regression only needed 11 models tested to choose an accurate model. The two neural networks needed 100+ models each tested before the final model was chosen.

The neural network models also needed significant data preprocessing to reduce and standardise the data. The preprocessing of the data therefore increased the time taken to implement the neural network models. Conversely the elastic-net model did not need any data pre-processing asides from replacing missing data. The elastic-net model was therefore significantly easier and faster to implement compared to the neural network models.

Reproducible results

The reproducibility of results is very important in the development and validation of research results. The reproducibility of results allows ensures integrity and transparency of research findings. A significant issue with the creation of the ‘Keras’ neural network models in our research was the lack of reproducible results even after setting the seed for all random number generation. Reproducible results were not achieved in this paper for the neural network models as the neural network parameters trained slightly different every time. In this paper to simulate reproducible results the final neural networks were saved as ‘Keras’ model files and loaded in from file instead of training the model each time the python script was ran. Alternately the elastic-net model achieved reproducible results through setting a seed which ensured the same model was trained every time.

Dimensionality reduction comparison

When comparing the dimensionality reduction techniques, PCA performed significantly better then UMAP on the methylation data examined. Initially the PCA plot was able to identify an outlier which once removal improved the predictions of both neural network models. The UMAP plot did not visually separate the outlier and grouped it with the other data. The PCA neural network performed accurately with a standard error of 5.38 which is significantly less than the UMAP neural networks standard error of 9.08. In the testing of the two dimensionality reduction techniques the UMAP reduction provided no advantage compared to the standard PCA reduction showing worse performance.

Conclusion

There are many approaches to age estimation from methylation data. Three models: elastic-net PCA trained neural network and UMAP trained neural network, were trained to predict age using methylation data. There were significant differences in the predictive abilities of the different models. The elastic-net regression produced the most accurate predictions with a standard error of 2.605 years. The PCA trained neural network performed second best with a moderate standard error of 5.378 years. The UMAP trained neural network performed poorly with a large standard error of 9.079 years. The optimal predictive ability of elastic-net is consistent with existing literature as most ‘epigenetic clock’ models in published papers used an elastic-net regression for their final model. The analysis of the different age estimation methods provided insight into the relative performance of different models and the strengths and weaknesses of different approaches.

The implementation of the elastic-net model was efficient as it did not require a large amount of data preprocessing. The small number of parameters in the elastic-net model was advantageous as it simplified the training and optimisation of the elastic-net model. Testing the elastic-net model was easy as the elastic-net allowed seeding of model training, allowing for reproducible results.

The model produced using elastic-net was very interpretable as all the 102 parameters used to predict age relate to specific methylation levels at sites around the body. The interpretability of the elastic-net model is significant as it allows the model to show which methylation sites around the body are most correlated to age. The elastic-net model can therefore be applied to other area of research into methylation at specific location in the body.

In the future, the research could be expanded to include a larger number of samples and including all ages in the data.

References

Tammen, Stephanie A., Simonetta Friso, and Sang-Woon Choi. 2013. “Epigenetics: The Link between Nature and Nurture.” Molecular Aspects of Medicine 34 (4): 753–64. https://doi.org/10.1016/j.mam.2012.07.018.

Anant Dadu, Vipul K Satone, Rachneet Kaur, Mathew J Koretsky, Hirotaka Iwaki, Yue A Qi, Daniel M Ramos, et al. 2023. “Application of Aligned-UMAP to Longitudinal Biomedical Studies.” Patterns 4 (6): 100741–41. https://doi.org/10.1016/j.patter.2023.100741.

Field, Adam E., Neil A. Robertson, Tina Wang, Aaron Havas, Trey Ideker, and Peter D. Adams. 2018. “DNA Methylation Clocks in Aging: Categories, Causes, and Consequences.” Molecular Cell 71 (6): 882–95. https://doi.org/10.1016/j.molcel.2018.08.008.

Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, et al. 2013. “Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates.” Molecular Cell 49 (2): 359–67. https://doi.org/10.1016/j.molcel.2012.10.016.

King-Batoon, Audrey, Joanna M. Leszczynska, and Catherine B. Klein. 2008. “Modulation of Gene Methylation by Genistein or Lycopene in Breast Cancer Cells.” Environmental and Molecular Mutagenesis 49 (1): 36–45. https://doi.org/10.1002/em.20363.

Powell, Joseph E., Anjali K. Henders, Allan F. McRae, Anthony Caracella, Sara Smith, Margaret J. Wright, John B. Whitfield, et al. 2012. “The Brisbane Systems Genetics Study: Genetical Genomics Meets Complex Trait Genetics.” Edited by Dan E. Arking. PLoS ONE 7 (4): e35430. https://doi.org/10.1371/journal.pone.0035430.

Yousef El-Laham, Elizabeth Fons, Dillon Daudert, and Svitlana Vyetrenko. 2024. “Augment on Manifold: Mixup Regularization with UMAP,” April. https://doi.org/10.1109/icassp48485.2024.10446585.

Moore, Lisa D, Thuc Le, and Guoping Fan. 2012. “DNA Methylation and Its Basic Function.” *Neuropsychopharmacology* 38 (1): 23–38. https://doi.org/10.1038/npp.2012.112.

Kader, Farzeen, Meenu Ghai, and Leah Maharaj. 2018. “The Effects of DNA Methylation on Human Psychology.” *Behavioural Brain Research* 346 (July): 47–65. https://doi.org/10.1016/j.bbr.2017.12.004.

Levine, Morgan E., Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, et al. 2018. “An Epigenetic Biomarker of Aging for Lifespan and Healthspan.” *Aging (Albany NY)* 10 (4): 573–91. https://doi.org/10.18632/aging.101414.

Horvath, Steve. 2013. “DNA Methylation Age of Human Tissues and Cell Types.” *Genome Biology* 14 (10): R115. https://doi.org/10.1186/gb-2013-14-10-r115.

Zou, Xueqing . n.d. “DNA Methylation and Hydroxymethylation.” Www.ks.uiuc.edu. Accessed September 13, 2024. https://www.ks.uiuc.edu/Research/methylation/.

Gal, Yarin, and Zoubin Ghahramani. (2016). "A theoretically grounded application of dropout in recurrent neural networks." *Advances in neural information processing systems* 29

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014 "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15, no. 1 (2014): 1929-1958.

Roy, Ritwick. 2022. “Neural Networks: Forward Pass and Backpropagation.” Medium. November 17, 2022. https://towardsdatascience.com/neural-networks-forward-pass-and-backpropagation-be3b75a1cfcc.

Brownlee, Jason. 2021. “Difference between Backpropagation and Stochastic Gradient Descent.” Machine Learning Mastery. January 31, 2021. https://machinelearningmastery.com/difference-between-backpropagation-and-stochastic-gradient-descent/.

Sarita. 2023. “Basic Understanding of Neural Network Structure.” Medium. October 3, 2023. https://medium.com/@sarita_68521/basic-understanding-of-neural-network-structure-eecc8f149a23.

Zou, Hui, and Trevor Hastie. 2005. “Regularization and Variable Selection via the Elastic Net.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Al-Faiz, Mohammed Z, Ali Abdulhafidh Ibrahim, and Sarmad M Hadi. 2019. “The Effect of Z-Score Standardization (Normalization) on Binary Input Due the Speed of Learning in Back-Propagation Neural Network.” *Iraqi Journal of Information and Communication Technology* 1 (3): 42–48. https://doi.org/10.31987/ijict.1.3.41.