

1163-68-104

Hannah Guan* (hguan000@gmail.com). *Predicting biological age: Neural network and algorithmic models for high dimensional DNA methylation data.*

While the traditional formula of DNA methylation aging is based on linear models, few works have explored the effectiveness of neural network models, which can learn more complex relationships from data. DNA methylation data usually consists of hundreds of thousands of feature space, which is much greater than the number of biological samples. This introduces the problem of overfitting which leads to a poor generalization of the neural network model. We propose a neural network model called Correlation Pre-Filtered Neural Network (CPFNN). CPFNN uses Spearman Correlation to pre-filter the features before feeding them into a neural network. We compare CPFNN with the Statistical Regression models (e.g, Horvath, and Hannum's formula), the Basic Neural Network, and the Dropout Neural Network. CPFNN outperforms these models by at least 1 year in terms of mean average error (MAE), with an MAE of 2.7 years. We also test for association between the epigenetic age using CPFNN with Schizophrenia and Down syndrome ($p=0.024$ and <0.001 , respectively). We discover that for a large number of candidate features, such as in genome-wide DNA methylation data, a key factor in improving prediction accuracy is how to appropriately weight features that are highly correlated with the outcome of interest. (Received August 15, 2020)