ABSTRACT

We develop a method to identify individuals with abnormal aging rates based on DNA methylation data. Recent research has shown that while DNA methylation levels could be used to form an epigenetic clock with great prediction accuracy, there are significant differences on individual level if different models are used to form the clock. We analyze the limitations of current methods that lead to such inconsistencies and develop a method for outlier detection. We argue that conducting outlier detection based on a single model is unconvincing, and thus form a residual path for each individual based on multiple models. We then develop a distance to measure the deviation of each residual path from chronological age and conduct outlier detection based on the distance we define. We demonstrate the performance of our method by simulations, and compare it with the result gained by using a single model. Finally, we apply the methodology to real data.