



Application of AI in biological age prediction

Dawei Meng¹, Shiqiang Zhang¹, Yuanfang Huang¹,
 Kehang Mao¹ and Jing-Dong J. Han

Abstract

The development of anti-aging interventions requires quantitative measurement of biological age. Machine learning models, known as “aging clocks,” are built by leveraging diverse aging biomarkers that vary across lifespan to predict biological age. In addition to traditional aging clocks harnessing epigenetic signatures derived from bulk samples, emerging technologies allow the biological age estimating at single-cell level to dissect cellular diversity in aging tissues. Moreover, imaging-based aging clocks are increasingly employed with the advantage of non-invasive measurement, making it suitable for large-scale human cohort studies. To fully capture the features in the ever-growing multi-modal and high-dimensional aging-related data and uncover disease associations, deep-learning based approaches, which are effective to learn complex and non-linear relationships without relying on pre-defined features, are increasingly applied. The use of big data and AI-based aging clocks has achieved high accuracy, interpretability and generalizability, guiding clinical applications to delay age-related diseases and extend healthy lifespans.

Addresses

Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Center for Quantitative Biology (CQB), Peking University, Beijing 100871, China

Corresponding author: Han, Jing-Dong J. (jackie.han@pku.edu.cn)

¹ These authors contributed equally to this work.

Current Opinion in Structural Biology 2024, 85:102777

This review comes from a themed issue on **Artificial Intelligence (AI) Methodologies in Structural Biology (2024)**

Edited by Tero A. Aittokallio and Evandro Fei Fang

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online xxx

<https://doi.org/10.1016/j.sbi.2024.102777>

0959-440X/© 2024 Elsevier Ltd. All rights reserved.

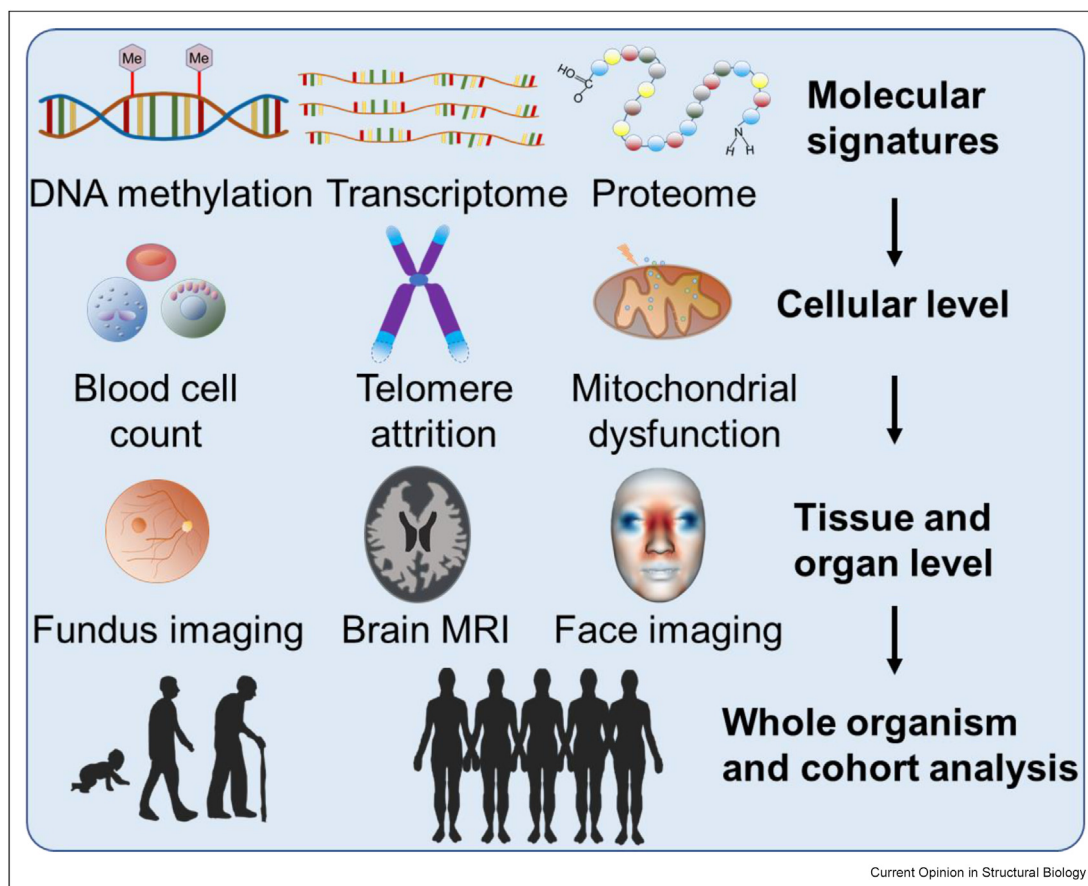
Introduction

Chronological age has been recognized as the single largest risk factor for human diseases, including cancer, cardiovascular disease and neurodegeneration. During the aging process, a variety of molecular and cellular hallmarks have been proposed, such as epigenetic alterations, telomere attrition, upregulated chronic

inflammation, loss of proteostasis, mitochondrial dysfunction, cellular senescence [1] (Figure 1). Biological age estimation is a challenging task involving the systematic measurement of key aging biomarkers, mortality modeling, health status and disease risk evaluation. Accurately predicting biological age is essential for the early detection of age-related diseases, treatment monitoring of clinical interventions, development of precise and personalized medicines and is also beneficial for the understanding of aging process itself. The most commonly used method to predict biological age is to employ machine learning algorithms to develop an aging clock based on these varying biomarkers. Furthermore, by comparing the actual chronological age with the predicted age, the age-independent part of the difference obtained is called Δ age or AgeDiff, which can be used to estimate the rate of biological aging [2] (Figure 2a). DNA methylation changes, as a hallmark of aging, have been harnessed to develop the epigenetic aging clocks, which are among the first and most widely used aging clocks [3]. Since the development of epigenetic aging clocks, numerous aging clocks have been built to predict biological aging using various molecular and cellular biomarkers, particularly omics-based approaches, such as transcriptomic clocks [4], proteomic clocks [5], and metabolomic clocks [6]. Recently, advances in single-cell isolation and barcoding technologies have enabled measuring biological age at single-cell resolution, allowing the cell-type-specific profiling of epigenomic, transcriptomic and proteomic changes during aging [7,8]. Moreover, aging clocks that make use of non-invasive techniques, such as brain structural MRI [9,10] or human facial imaging [11,12], have achieved high accuracy and gained wide acceptance. These approaches enable the analysis of large-scale human cohorts and the correlation of biological age with environmental factors and disease risk and prevalence.

To model biological aging measured by different biomarker metrics, a variety of artificial intelligence (AI) methods have been employed, including traditional machine learning and more recent developed deep learning approaches. Traditionally, linear regression models are widely used to build aging clocks. Due to the large dimensionality of omics sequencing data, standard linear models perform poorly and fail to model the complex distributions of sequencing data. To address the limitations, penalized linear regression models,

Figure 1



Systemic profiling of aging biomarkers and features used to estimate biological aging.

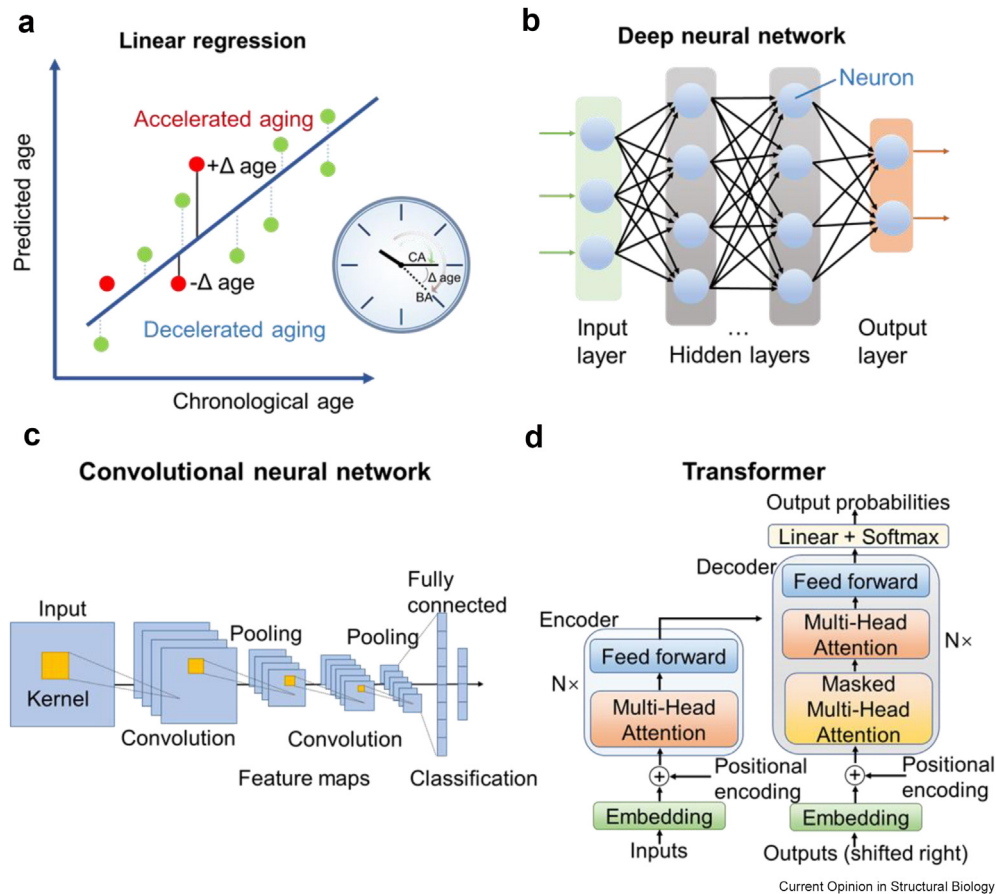
including ridge, lasso and elastic net, are commonly used to build aging clocks using high-throughput omics data such as DNA methylation, transcriptomic and proteomic data. The advantages of linear regression models are lightweight, simplicity and interpretability.

With the advances of AI, deep neural networks (DNN) are proving to be a powerful tool to build aging clocks, particularly for high-dimensional omics data and large-scale medical imaging datasets. DNN architectures are composed of multiple hidden layers between the input and output layer (Figure 2b). Each layer contains a given number of nodes (neurons) that can perform non-linear transformations on the inputs from the prior layer [13]. Typical DNNs include convolutional neural networks (CNN), multi-layer perceptron (MLP) and recurrent neural networks (RNN). CNNs are widely used for computer vision (CV) and medical imaging analysis. CNN architectures consist of multiple convolutional layers and pooling layers followed by a downstream classifier. The convolutional layer uses a kernel (filter) to extract structures and features from the input images

(Figure 2c). With multiple convolution operations, CNNs are able to capture high-level representation of the input images. Recently, Transformer based architectures are revolutionizing natural language processing (NLP) and computer vision [14]. A Transformer is a novel deep learning architecture that adopts the self-attention, which is an attention mechanism that calculates the attention weights between each element in an input sequence, allowing the model to find the long-term dependencies and relationships in sequential data (Figure 2d). Popular Transformer-based models include AlphaFold2, BERT and GPT. Due to stronger modeling capabilities and greater scalability with respect to large models and large datasets, the Transformers are increasingly employed in medical imaging classification, omics data analysis and large model building.

In this review, we explore the application of artificial intelligence to build aging clocks. Specifically, we introduce how machine learning models can help the construction of aging clocks utilizing bulk omics data or single-cell sequencing data, as well as non-invasive aging

Figure 2



Machine learning and deep learning architectures used to build aging clocks. (a) Linear regression based aging clocks to estimate biological aging. BA – biological age, CA – chronological age. (b–d) General architectures for deep neural network (DNN, B), convolutional neural network (CNN, C) and Transformer (D).

biomarkers, with a particular focus on newly developed deep learning models. We also discussed the potential challenges and prospects for deep learning models to estimate biological aging.

Aging clocks built by machine learning models using bulk omics data

DNA methylation clocks are among the earliest and most widely used aging clocks, known for their high accuracy. As early as 2011, Bocklandt et al. accurately predicted human age (MAE = 5.2) using array-based DNA methylation data with Lasso linear regression model [3]. After that, several methylation aging clock studies were constructed based on linear models such as elastic net [15–21]. Currently, there are four widely used DNA methylation aging clocks: the Horvath clock [16], Hannum clock [15], Levine clock [22], and GrimAge clock [23]. Recently, taking advantage of the rapid development of machine learning and artificial intelligence, Levy et al. built MethylNet, a deep

learning-based aging clock, and the MAE of age prediction reached 3.0. Moreover, MethylNet is not limited to prediction, but can also be used for classification, regression and multi-output tasks [24]. Galkin et al. introduced another DNA methylation aging clock based on DNN, named DeepMAge, further enhancing the accuracy of DNA methylation-based aging clocks (MedAE = 2.8 from an Independent validation) [25]. Camillo et al. built AltumAge based on a deep learning model, which achieved a MAE of 2.15 for predicting age using all 20,318 methylation sites, significantly better than other traditional models such as ElasticNet [26]. These results highlight the potential of utilizing advanced techniques to improve the precision and effectiveness of DNA methylation-based aging clocks.

The aging clocks based on other data have also been quickly established. In 2015, Peters et al. constructed the first transcriptome aging clock (MAE = 7.77) using multiple linear regression based on the gene expression

data of 14,983 individuals' whole blood [4]. Subsequently, Fleischer et al., Mamoshina et al. and Xia et al., in 2018, 2018 and 2020 respectively, also built linear transcriptome aging clocks based on human dermal fibroblasts, muscle and PBMC gene expression data (MedAE = 4.0, MAE = 6.24 and MAE = 5.68) [12,27,28]. Holzscheck et al. developed an artificial neural network built from well-described biological pathways that can predict age from gene expression data in skin tissue (MAE = 5.51), while revealing the aging state of the pathway that contribute to the prediction [29]. Urban et al. first constructed a transformer-based model [30], named Precious1GPT, for age prediction using DNA methylation and transcriptome data. However, compared to previous linear models, no significant improvement in accuracy was observed (MAE = 4.22 for methylation, MAE = 6.29 for gene expression and MAE = 5.62 for combined).

Protein homeostasis has long been recognized as one of the hallmarks of aging [1]. Protein levels can also be used to predict age. In 2014, Krišić et al. first established an aging clock based on the measurement of IgG glycosylation in human plasma (MAE = 9.7) [31]. In 2021, Sayed et al. developed an inflammatory aging clock (iAge) based on deep neural network method called guided auto-encoder using the blood immunome of 1001 individuals [32]. In 2018, Tanaka et al. created an aging clock utilizing elastic net regression models for the plasma proteome as a whole, rather than focusing on a specific protein class ($r = 0.94$) [5]. In 2019, Lehallier et al. achieved more accurate age predictions using lasso model and a larger number of plasma proteins ($n = 373$) (PCC = 0.97). Furthermore, their research creatively discovered that the number of age-associated proteins does not change steadily with age but fluctuates, exhibiting peaks at the ages of 34, 60, and 78 [33]. The following year, they refined their model to higher accuracy by further combining different proteins (MAE = 2.44) [34].

In addition, there are aging clocks based on other types of data. In 2020, Galkin et al. developed human gut microbiome aging clock based on taxonomic profiling and deep learning for predicting host age [35]. They found that the architecture of deep neural networks (MAE = 5.91) performed better compared to linear models like elastic net. Chen et al. developed a microbiome aging clock with an ensemble model with multiple heterogeneous algorithms namely multi-view learning, which combined species and pathway profiles and showed acceptable accuracy (MAE = 8.33) [36].

Aging heterogeneity uncovered by single-cell aging clocks

The advancement of single-cell high-throughput omics technologies enabled the scientific community to

decode the aging hallmarks at single-cell resolution. In this part, we summarized the application of AI methods on bulk or single-cell omics to estimate biological aging of individual cells.

In 2017, our group developed an "iCpSc" package, which predicts differentiation time for individual mouse embryonic stem cells (mESC) single cells utilizing matched cell population RNA-seq (cpRNA-seq) as a reference [37]. By leveraging partial least squares regression, we demonstrated that the first two components of the model suffice to construct a linear model with satisfactory performance. Additionally, Singh et al. introduced GERAS (Genetic Reference for Age of Single-cell), a machine learning age classifier capable of predicting age stages of individual cells based on their transcriptomes [38]. GERAS demonstrated over 90 % accuracy in classifying zebrafish and human pancreatic cell ages. Bulteau et al. presented another computation method named real-age prediction from transcriptome staging on reference (RAPToR) [39]. RAPToR utilizes existing time-series bulk transcriptome data as a reference to estimate real age of a sample through Independent Component Analysis (ICA) or Principal Component Analysis (PCA) based maximum likelihood estimation. The study showcased RAPToR's capability in predicting age of not only individual animal, but also dissected tissue and single-cell data. An alternative approach has been proposed by Buckley et al., they developed cell-type-specific aging clocks to predict both chronological age and biological age (neural stem cell proliferation capacity) in the six most abundant cell types within subventricular zone neurogenic region [7]. Leveraging scRNA-seq data from 28 mice spanning 26 different age stages, they observed varying age-sensitivity among the six cell types. Moreover, the results showed that lasso models trained on bootstrap-sampled meta cells outperformed all other models including single cell models. Notably, these clocks were successfully employed to evaluate the effects of anti-aging interventions such as exercise and heterochronic parabiosis. Similarly, Lu et al. developed a human CD8+ T cell-specific aging clock using a fitted mixed-effect elastic net model [40]. By utilizing scRNA-seq data from both cross-sectional and longitudinal samples, the model accurately predicted the age of individual cells based on their transcriptomic features. Moreover, the researchers uncovered a close association between cell age, cell differentiation, and mutation burden. Instead of building aging clock based on gene expression profile, our group constructed a single cell aging clock based on cell type composition using a partial least square regression (PLSR) linear model [41]. The clock, developed from scRNA-seq data from PBMC samples, revealed supercentenarians (SCs) have delayed age-related changes in cell composition and gene expression compared to clock model expected values.

Epigenetic clocks offer another avenue for measuring single-cell aging. In 2021, Trapp *et al.* pioneered the development of the first epigenetic single-cell aging clock, named scAge [8]. This novel approach infers the biological age of individual cells from bulk data using maximum likelihood estimation. Apart from DNA methylation modifications, ribosomal DNA methylation can also be an indicator of aging. A multi-tissue ribosomal DNAm (rDNAm) clock was constructed using bulk DNAm data [42]. Moreover, the authors applied this clock to predict single-cell age from single-cell methylation sequencing data by converting single cell sequencing data into pseudo bulk sequencing data. In a preprint from May 2022, a mitotic age predictor called EpiTrace was developed utilizing single-cell chromatin accessibility sequencing (scATAC) data [43]. By quantifying the opened fraction of clock-like differential methylation loci (ClockDML) in single cells, EpiTrace provides accurate age prediction.

In general, there are two main strategies of estimating the age of a single cell: inferring age from either omics data of bulk tissue/purified cell population or from single-cell omics data. Datasets derived from bulk tissue offer enhanced consistency compared to those from single cells. However, bulk tissue and even purified cell populations amalgamate molecular changes across diverse tissue compositions and cell states, masking the authentic gene distribution inherent to a specific cell type. The proliferation of single-cell omics data in recent years has provided the avenue to construct aging clocks directly from single-cell omics data. This facilitates the investigation into the distinct contributions of cell types to aging and rejuvenation process. Nonetheless, it imposes higher demands on data processing and model development due to the elevated noise and heterogeneity intrinsic to single-cell omics data.

Deep learning-based aging clocks using non-invasive biomarkers of aging

Non-invasive aging clocks are a type of aging clock that uses data collected from non-invasive examinations to evaluate the biological status of aging. These examinations can include facial images, biomedical images (e.g., MRI) and questionnaires. Using deep learning models and artificial intelligence, these clocks provide accurate estimations of chronological age with minimal error rates. In view of the convenience and cost-effectiveness of the collection methods, non-invasive examinations are more applicable for routine medical check-ups and mass screening than invasive examinations.

Taken facial images first, our group published the first 3D facial image-based age predictor in 2015 which has a MAE of 6 years and extended the facial age prediction to a new AI-based model by collecting 3D facial images from a cohort of ~5000 Han Chinese [44]. New facial aging clock achieved MAE of 2.8 years to estimate

chronological and perceived age (the age that a person is visually estimated to be on the basis of face), respectively named FaceCnnAge and FaceCnnPerceivedAge, and revealed a high heterogeneity of aging rate in middle age, providing evidence for the best time for anti-aging interventions. Besides, causal inference test between lifestyles and aging rate uncovered potential lifestyle factors, such as smoking or yogurt consumption that exacerbate or mitigate the aging rate [12].

Besides facial images, PhotoAgeClock is developed to predict chronological age with deep neural network through a single eye corner image and overall 8414 images of eye corners reach an MAE of 2.3 years for model error [45]. However, there lacks biological relevance for this aging clock. Furthermore, fundus image, involving photographing the rear of an eye and including main structures of retina, have been applied to construct retinal aging clocks with deep learning models. Zhu *et al.* built a CNN-based retinal aging clock with MAE of 3.55 years and showed that each one year increase in the retinal age gap (retinal age predicted minus chronological age) was associated with a 2 % increase in risk of all-cause mortality [46]. More accurate retinal aging clock has been obtained by Ahadi *et al.* and was reinforced through GWAS hits by identifying candidate genes associated with several age-related functions. The top GWAS locus was further validated via knockdown of the fly homolog, *Alk*, which slowed age-related decline in vision in flies, thus strengthening the biological foundation of retinal aging clocks [47].

Other non-invasive examinations also include biomedical images of MRIs. In 2017, Cole *et al.* generated 'brain-predicted age' on brain MRI data using 3D-CNN networks with MAE of 4.16 years, which is better compared to conventional machine learning method Gaussian Process Regression (GPR) [9]. Recently, Yin *et al.* used an interpretable 3D-CNN network to estimate MRI-derived brain age and reached MAE around 2.3 years [10]. Their models provided detailed anatomic maps of brain aging patterns that reveal neurocognitive trajectories in adults with mild cognitive impairment and Alzheimer's disease. They demonstrated that significant associations between biological age and early signs of Alzheimer's disease and provided insight into early identification of individuals at high risk of Alzheimer's disease.

In conclusion, non-invasive aging clocks represent a transformative paradigm in aging research, enabling precise and accessible assessments of the biological status of aging. The implementation of deep learning models and artificial intelligence has opened unprecedented opportunities for advancing preventive and personalized anti-aging strategies, as well as facilitating early detection of age-related health risks on a large population scale. Continued research and validation

efforts will further enhance the scientific significance and applicability of non-invasive aging clocks in diverse clinical and public health settings.

Challenges and future directions

Biological aging is a systemic and complex process, with diverse molecular, cellular and organ-level changes. Thus, large numbers of machine learning models were built to estimate biological age leveraging different biomarkers of aging. Traditional machine learning methods are still widely used to build the epigenetic, transcriptomic and other omics-based aging clocks due to their less data intensive, better model interpretability and computational simplicity. On the other hand, DNN-based models are quickly dominating other types of aging clocks, particularly those based on imaging and multi-modal data. In addition, AI models, especially deep learning-based models, are growing extremely large and complex in order to obtain higher accuracy, robustness and generalization in handling big data.

Here, we listed the state-of-the-art AI models for biological aging prediction, with an emphasis on DNN-based methods (Table 1) (Table 2).

However, there are still challenges for the wide use of deep learning in biological aging estimation.

Multi-modal data integration

Considering the multifaceted nature of aging, a single modality of data may be insufficient to capture the

complexity of biological aging. Using multi-modal data, including omics data, imaging data, electronic records and clinical tests, to generate a composite aging clock may better represent biological aging, improve model interpretability and more accurately reflect the pace of biological aging [50,51]. Recently, advances in single-cell multiomics sequencing technologies have enabled the simultaneous collection of multi-modal datasets, providing opportunities to systematically explore the cellular diversity and heterogeneity in aging tissues.

Deep learning black box

One of the major challenges for deep learning-based aging clocks is model interpretability. The “Black Box” nature of deep learning has generated widespread concerns when it comes to clinical applications. To address the issue of model interpretability, a variety of algorithms have been developed in order to explain the model’s predictions and identify the important factors contributing to the model’s decisions, such as CAM (classification activation map), SHAP (SHapley Additive exPlanations) and attention maps for Transformer-based models. For example, SHAP, a model interpretation method derived from game theory, explains individual predictions by computing the contribution of each feature to the prediction and has been widely used to explain the output of aging clocks based on DNN models [24,26]. Lucas Camillo and colleagues extracted the most important CpG sites based on the SHAP values from the neural network-based epigenetic clock, AltumAge, and found that the top-ranking CpG sites are

Table 1

AI models used to build human aging clocks.

| Data type | Sample size | Model architecture | MAE | PCC | Aging clocks | Year |
|-------------------------|-------------|------------------------------|------|------|---------------------------------|------|
| 2D facial images | 55,134 | CNN | 2.93 | N.A. | Facial Age [11] | 2018 |
| 3D facial images | ~5000 | Ensemble of CNN | 2.79 | 0.96 | FaceCnnAge [12] | 2020 |
| Blood tests | 62,419 | Ensemble of DNN | 5.55 | 0.91 | Hematological aging clock [48] | 2016 |
| Blood immunome | 1001 | GAE | 15.2 | N.A. | iAge [32] | 2021 |
| Blood plasma proteins | 3301 | LASSO | 2.44 | 0.96 | Proteomic aging clock [34] | 2020 |
| Chest X-ray image | 116,035 | CNN | N.A. | N.A. | CXR-Age [49] | 2021 |
| DNAm | 9926 | ElasticNet | N.A. | 0.71 | PhenoAge [22] | 2018 |
| DNAm | 6395 | ElasticNet | N.A. | 0.82 | GrimAge [23] | 2019 |
| DNAm | 9500 | VAE | 3.0 | 0.98 | MethylNet [24] | 2020 |
| DNAm | 4930 | DNN | 3.21 | 0.98 | DeepMAge [25] | 2021 |
| DNAm | 13505 | DNN | 2.15 | 0.98 | AltumAge [26] | 2022 |
| Eye corner images | 8414 | Xception | 2.3 | 0.96 | PhotoAgeClock [45] | 2018 |
| Fundus images | 19,200 | Xception | 3.55 | 0.81 | Retinal age [46] | 2023 |
| Fundus images | 271,581 | Inception | 2.86 | 0.95 | EyeAge [47] | 2023 |
| Gut metagenomics | 3663 | DNN | 5.91 | 0.53 | Microbiome aging clock [35] | 2020 |
| Structural MRI | 2001 | CNN | 4.16 | 0.96 | BrainAge [9] | 2017 |
| Structural MRI | 5851 | CNN | 2.23 | 0.95 | BrainAge [10] | 2023 |
| Transcriptome | 133 | Linear discriminant analysis | 7.7 | 0.9 | Transcriptomic aging clock [27] | 2018 |
| DNAm and transcriptome | 33,745 | Transformer | 5.62 | 0.90 | Precious1GPT [30] | 2023 |
| Transcriptomic pathways | 887 | DNN | 5.51 | N.A. | Transcriptomic aging clock [29] | 2021 |

Table 2**Key terms and definitions used in the article.**

| Abbreviation | Term | Definition |
|--------------|---|--|
| CA | Chronological age | The actual time an organism has been alive after birth. |
| BA | Biological age | A model to estimate an organism's physiological age based on the measurement of age-dependent biomarkers. |
| AgeDiff | Age difference | The difference between biological age and chronological age. |
| MRI | Magnetic resonance imaging | A non-invasive medical imaging technique using strong magnetic to generate three dimensional detailed anatomical images. |
| Lasso | Least absolute shrinkage and selection operator | A regularized linear regression model that uses an L1 penalty term on regression parameters. |
| PLSR | Partial least squares regression | A linear regression model that reduces the predictors into a smaller set of uncorrelated components before performing least squares regression. |
| DNN | Deep neural network | A kind of neural network that consists of multiple hidden layers between the input and output layer. |
| CNN | Convolutional neural network | A subtype of neural network with multiple convolutional layers to extract features from the input images, followed by pooling layers and fully connected layers to achieve image detection and classification. |
| Transformer | Transformer architecture | A neural network architecture that makes use of self-attention mechanisms to attend to different positions in the input sequence simultaneously. |
| SHAP | SHapley Additive ExPlanation | An approach for model interpretation, based on game theory to assign credit for a model's prediction to each feature. |
| MAE | Mean absolute error | A metric used to evaluate the performance of a model's prediction, which is calculated as the sum of absolute errors divided by the sample size. |
| MedAE | Median absolute error | A metric used to evaluate the performance of a model, which is the median value of the absolute errors. |

highly associated with gene regulatory regions in the genome [26]; Grad-CAM has been employed to localize the anatomy contributing to CXR-Age [49]. Furthermore, an increasing number of deep-learning based models have incorporated interpretable features in model design [30,52]. For example, by transforming high-dimensional sparse expression space to low-dimensional pathway-level embeddings, TOSICA, a Transformer based model, can also achieve interpretability through calculating the attention weights between classifier token (CLS) and pathway tokens [52].

Intervention guidance and disease relevance

The ultimate goal of building aging clocks is to provide clinical guidance for anti-aging interventions and predict disease risk. The traditional linear models have made it possible to predict biological aging using a limited number of molecular markers, such as DNA methylation sites, protein and RNA molecules. However, challenges in dissecting the causative factors have limited their clinical intervention applications. For deep learning-based models, the effector genes and factors are difficult to retrieve from the model due to the over-parameterized “black box” nature. To tackle these problems, Cumplido-Mayoral *et al.* developed a brain-age clock from structural neuroimaging data and found that brain-age delta is associated with specific neuro-degeneration and cerebrovascular disease biomarkers [53]; Xia *et al.* built a causal-inference network to infer the impact of molecular mediators in the blood and

lifestyles on human facial-aging rate, guiding the lifestyle choices and anti-aging interventions [12].

The growing large datasets and multi-modal high-dimensional data collected from aging biomarkers necessitates the development of more powerful AI models to integrate, process and analyze. Deep learning-based aging clocks, particularly CNN and Transformer-based architectures, have demonstrated to be powerful tools in estimating biological aging for bulk and single-cell omics data and non-invasive imaging data. We expect these aging clocks will be able to assist in evaluating human health status, predicting disease risk and guiding the personalized interventions of human aging in the near future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by grants from China Ministry of Science and Technology (2020YFA0804000) and National Natural Science Foundation of China (92049302, 32088101, 32330017) to J.D.J.H. and CLS

postdoctoral fellowship to D.M. We apologize to colleagues whose works are omitted in the review due to space limit.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G: **Hallmarks of aging: an expanding universe.** *Cell* 2023, **186**: 243–278.
This review systematically proposed 12 hallmarks of aging, including genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis et al. These hallmarks are based on 3 criteria: (1) age-associated manifestation, (2) accentuating hallmark cause acceleration of aging, (3) potential of therapeutic interventions.
2. Rutledge J, Oh H, Wyss-Coray T: **Measuring biological age using omics data.** *Nat Rev Genet* 2022, **23**:715–727.
This review comprehensively introduced how omics data, such as epigenomic, transcriptomic, proteomic and metabolomic data, can be leveraged with machine learning to build aging clocks for the biological age prediction.
3. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E: **Epigenetic predictor of age.** *PLoS One* 2011, **6**, e14821.
4. Peters MJ, Joeannes R, Pilling LC, Schurmann C, Conneely KN, Powell J, Reinmaa E, Sutphin GL, Zhernakova A, Schramm K, et al.: **The transcriptional landscape of age in human peripheral blood.** *Nat Commun* 2015, **6**:8570.
5. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, Candia J, Zhang P, Cheung F, Fantoni G, et al.: **Plasma proteomic signature of age in healthy humans.** *Aging Cell* 2018, **17**, e12799.
6. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai PC, Gieger C, Schulz H, Erte I, John S, et al.: **Metabolomic markers reveal novel pathways of ageing and early development in human populations.** *Int J Epidemiol* 2013, **42**: 1111–1119.
7. Buckley MT, Sun ED, George BM, Liu L, Schaum N, Xu LCY, Reyes JM, Goodell MA, Weissman IL, Wyss-Coray T, et al.: **Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain.** *Nature Aging* 2023, **3**:121 [+].
This study constructed the first single cell aging clocks from single-cell transcriptomic data. It compared the heterogeneity of aging rate in six different cell types and applied them to measure transcriptomic rejuvenations-exercise and hetero parabiosis.
8. Trapp A, Kerepesi C, Gladyshev V: **Profiling epigenetic age in single cells.** *Innovation in Aging* 2021, **5**:673–673.
9. Cole JH, Poudel RPK, Tsagkrasoulis D, Caan MWA, Steves C, Spector TD, Montana G: **Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker.** *Neuroimage* 2017, **163**:115–124.
10. Yin C, Imms P, Cheng M, Amgalan A, Chowdhury NF, Massett RJ, Chaudhari NN, Chen X, Thompson PM, Bogdan P, et al.: **Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment.** *Proc Natl Acad Sci U S A* 2023, **120**, e2214634120.
11. Wan J, Tan Z, Lei Z, Guo G, Li SZ: **Auxiliary Demographic information assisted age estimation with cascaded structure.** *IEEE Trans Cybern* 2018, **48**:2531–2541.
12. Xia X, Chen X, Wu G, Li F, Wang Y, Chen Y, Chen M, Wang X, Chen W, Xian B, et al.: **Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle.** *Nat Metab* 2020, **2**:946–957.
13. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521**: 436–444.
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I: **Attention is all you need.** *Adv Neural Inf Process Syst* 2017, **30**:30 (Nips 2017).
15. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al.: **Genome-wide methylation profiles reveal quantitative views of human aging rates.** *Mol Cell* 2013, **49**:359–367.
This study demonstrated a groundbreaking approach in utilizing DNA methylation levels to construct a human aging clock, enabling highly accurate prediction of physiological age. The work, alongside Horvath's contributions, serves as a pioneering cornerstone in the field of aging clocks, providing significant guidance for future studies in this area.
16. Horvath S: **DNA methylation age of human tissues and cell types.** *Genome Biol* 2013, **14**.
17. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jockel KH, Erbel R, Muhleisen TW, et al.: **Aging of blood can be tracked by DNA methylation changes at just three CpG sites.** *Genome Biol* 2014, **15**.
18. Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MRP, Deary IJ, Wagner W: **DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy.** *Aging-Us* 2016, **8**:394–401.
19. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD, Zheng SJC, Widschwendter M, Beck S, Teschendorff AE: **Correlation of an epigenetic mitotic clock with cancer risk.** *Genome Biol* 2016, **17**.
20. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, Felton S, Matsuyama M, Lowe D, Kabacik S, et al.: **Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies.** *Aging-Us* 2018, **10**: 1758–1775.
21. Levine ME, Suarez JA, Brandhorst S, Balasubramanian P, Cheng CW, Madia F, Fontana L, Mirisola MG, Guevara-Aguirre J, Wan J, et al.: **Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population.** *Cell Metabol* 2014, **19**: 407–417.
22. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, et al.: **An epigenetic biomarker of aging for lifespan and healthspan.** *Aging-Us* 2018, **10**:573–591.
23. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou LF, Baccarelli AA, Li Y, Stewart JD, et al.: **DNA methylation GrimAge strongly predicts lifespan and healthspan.** *Aging-Us* 2019, **11**:303–327.
24. Levy JJ, Titus AJ, Petersen CL, Chen Y, Salas LA, Christensen BC: **MethylNet: an automated and modular deep learning approach for DNA methylation analysis.** *BMC Bioinf* 2020, **21**:108.
25. Galkin F, Mamoshina P, Kochetov K, Sidorenko D, Zhavoronkov A: **DeepMAGE: a methylation aging clock developed with deep learning.** *Aging and Disease* 2021, **12**: 1252–1262.
26. de Lima Camillo LP, Lapierre LR, Singh R: **A pan-tissue DNA-methylation epigenetic clock based on deep learning.** *npj Aging* 2022, **8**:4.
27. Fleischer JG, Schulte R, Tsai HH, Tyagi S, Ibarra A, Shokhirev MN, Huang L, Hetzer MW, Navlakha S: **Predicting age from the transcriptome of human dermal fibroblasts.** *Genome Biol* 2018, **19**:221.
28. Mamoshina P, Kochetov K, Putin E, Cortese F, Aliper A, Lee WS, Ahn SM, Uhn L, Skjodt N, Kovalchuk O, et al.: **Population specific biomarkers of human aging: a big data study using South Korean, Canadian, and Eastern European patient populations.** *J Gerontol A Biol Sci Med Sci* 2018, **73**:1482–1490.
29. Holzscheck N, Falckenhayn C, Söhle J, Kristof B, Siegner R, Werner A, Schössow J, Jürgens C, Völzke H, Wenck H, et al.: **Modeling transcriptomic age using knowledge-primed artificial neural networks.** *npj Aging and Mechanisms of Disease* 2021:7.
30. Urban A, Sidorenko D, Zagirova D, Kozlova E, Kalashnikov A, Pushkov S, Naumov V, Sarkisova V, Leung GHD, Leung HW, et al.: **Precious1GPT: multimodal transformer-based transfer learning for aging clock development and feature importance**

- analysis for aging and age-related disease target discovery.** *Aging (Albany NY)* 2023, **15**:4649–4666.
31. Kristic J, Vuckovic F, Menni C, Klaric L, Keser T, Beceheli I, Pucic-Bakovic M, Novokmet M, Mangino M, Thaqi K, *et al.*: **Glycans are a novel biomarker of chronological and biological ages.** *Journals of Gerontology Series A-Biological Sciences and Medical Sciences* 2014, **69**:779–789.
 32. Sayed N, Huang Y, Nguyen K, Krejciova-Rajaniemi Z, Grawe AP, Gao T, Tibshirani R, Hastie T, Alpert A, Cui L, *et al.*: **An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging.** *Nat Aging* 2021, **1**:598–615.
- This study developed a new aging clock based on deep learning, iAge, which can monitor chronic inflammatory signals in the human body, so as to timely find individuals who are at increased risk of diseases such as weakness, immune aging and cardiovascular disease, or have important significance for early diagnosis and early intervention of age-related diseases.
33. Lehallier B, Gate D, Schaum N, Nanasi T, Lee SE, Yousef H, Losada PM, Berdnik D, Keller A, Verghese J, *et al.*: **Undulating changes in human plasma proteome profiles across the lifespan.** *Nat Med* 2019, **25**:1843 [–+].
- This study developed a highly predictive plasma proteomic aging clock and showed the relationship between proteomic age differences and many aging characteristics. It also describes a nonlinear pattern of proteomic aging, in which physiological aging does not simply proceed at a constant rate, but has three "accelerating inflection points": 34, 60, and 78 years.
34. Lehallier B, Shokhirev MN, Wyss-Coray T, Johnson AA: **Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging.** *Aging Cell* 2020, **19**.
 35. Galkin F, Mamoshina P, Aliper A, Putin E, Moskalev V, Gladyshev VN, Zhavoronkov A: **Human gut microbiome aging clock based on taxonomic profiling and deep learning.** *iScience* 2020, **23**, 101199.
 36. Chen YT, Wang HC, Lu WW, Wu T, Yuan WW, Zhu JL, Lee YK, Zhao JX, Zhang H, Chen W: **Human gut microbiome aging clocks based on taxonomic and functional signatures through multi-view learning.** *Gut Microb* 2022:14.
 37. Sun N, Yu XM, Li F, Liu DH, Suo SB, Chen WY, Chen SR, Song L, Green CD, McDermott J, *et al.*: **Inference of differentiation time for single cell transcriptomes using cell population reference data.** *Nat Commun* 2017, **8**.
 38. Singh SP, Janjuha S, Chaudhuri S, Reinhardt S, Kränkel A, Dietz S, Eugster A, Bilgin HI, Korkmaz S, Zararsiz G, *et al.*: **Machine learning based classification of cells into chronological stages using single-cell transcriptomics**, vol. 8; 2018.
 39. Bulteau R, Francesconi MJNM: **Real age prediction from the transcriptome with RAPToR** 2021, **19**:969–975.
 40. Lu J, Ahmad R, Nguyen T, Cifello J, Hemani H, Li JY, Chen JG, Li SY, Wang J, Achour A, *et al.*: **Heterogeneity and transcriptome changes of human CD8(+) T cells across nine decades of life.** *Nat Commun* 2022, **13**.
 41. Zhu H, Chen J, Liu K, Gao L, Wu H, Ma L, Zhou J, Liu Z, Han JJ: **Human PBMC scRNA-seq-based aging clocks reveal ribosome to inflammation balance as a single-cell aging hallmark and super longevity.** *Sci Adv* 2023, **9**, eabq7599.
 42. Kerepesi C, Zhang BH, Lee SG, Trapp A, Gladyshev VN: **Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging.** *Sci Adv* 2021:7.
 43. Wang X, Jin W, Wang G, Ju L, Chen F, Qian K, Xiao Y, Zhang Y: **Tracking single cell evolution via clock-like chromatin accessibility.** *bioRxiv*; 2022.
 44. Chen W, Qian W, Wu G, Chen W, Xian B, Chen X, Cao Y, Green CD, Zhao F, Tang K, *et al.*: **Three-dimensional human facial morphologies as robust aging markers.** *Cell Res* 2015, **25**:574–587.
 45. Bobrov E, Georgievskaya A, Kiselev K, Sevastopolsky A, Zhavoronkov A, Gurov S, Rudakov K, Tobar MdPB, Jaspers S, Clemann S, PhotoAgeClock: **Deep learning algorithms for development of non-invasive visual biomarkers of aging.** *Aging (Albany NY)* 2018, **10**:3249.
 46. Zhu Z, Shi D, Guankai P, Tan Z, Shang X, Hu W, Liao H, Zhang X, Huang Y, Yu H, *et al.*: **Retinal age gap as a predictive biomarker for mortality risk.** *Br J Ophthalmol* 2023, **107**:547–554.
 47. Ahadi S, Wilson KA, Babenko B, McLean CY, Bryant D, Pritchard O, Kumar A, Carrera EM, Lamy R, Stewart JM, *et al.*: **Longitudinal fundus imaging and its genome-wide association analysis provide evidence for a human retinal aging clock.** *Elife* 2023:12.
- This study shows not only the high accuracy of retinal aging clock using AI-based models, but also the correlation between aging rate and all-cause mortality. Specially, their gene-knockdown experiment confirms the top GWAS locus could slow down age-related decline.
48. Putin E, Mamoshina P, Aliper A, Korzinkin M, Moskalev A, Kolosov A, Ostrovskiy A, Cantor C, Vijg J, Zhavoronkov A: **Deep biomarkers of human aging: application of deep neural networks to biomarker development.** *Aging (Albany NY)* 2016, **8**: 1021–1033.
 49. Raghu VK, Weiss J, Hoffmann U, Aerts HJWL, Lu MT: **Deep learning to estimate biological age from chest radiographs.** *JACC (J Am Coll Cardiol): Cardiovascular Imaging* 2021, **14**: 2226–2236.
 50. Li J, Xiong M, Fu X-H, Fan Y, Dong C, Sun X, Zheng F, Wang S-W, Liu L, Xu M, *et al.*: **Determining a multimodal aging clock in a cohort of Chinese women.** *Med* 2023, **4**:825–848.
 51. Belsky DW, Caspi A, Arseneault L, Baccarelli A, Corcoran DL, Gao X, Hannon E, Harrington HL, Rasmussen LJ, Houts R, *et al.*: **Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm.** *Elife* 2020, **9**.
 52. Chen J, Xu H, Tao W, Chen Z, Zhao Y, Han J-DJ: **Transformer for one stop interpretable cell type annotation.** *Nat Commun* 2023, **14**:223.
 53. Cumplido-Mayoral I, García-Prat M, Operto G, Falcon C, Shekari M, Cacciaglia R, Milà-Alomà M, Lorenzini L, Ingala S, Meije Wink A, *et al.*: **Biological brain age prediction using machine learning on structural neuroimaging data: multi-cohort validation against biomarkers of Alzheimer's disease and neurodegeneration stratified by sex.** *Elife* 2023, **12**.