Prediction of Biological Age from Longitudinal Data Using Mixed – Effects Linear Regression and Long Short-Term Memory

Maria Magdalena Pradita Eka Kurniawati[1], Anuchate Pattanateepapon[*, 1], Nawanan Theera-Ampornpunt[1],
Boonsong Ongphiphadhanakul[2], Prin Vathesatogkit[2], and Ammarin Thakkinstian[1].

[1]Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital,
Mahidol University, Bangkok, Thailand
[2]Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
*Corresponding author: Email address: anuchate.gab@mahidol.ac.th

Abstract

This study aimed to accurately predict biological age (BA) in the Thai population over time, utilizing predictors are commonly used in clinical routines using a retrospective cohort study from Electric Generating Authority of Thailand (EGAT) 2 dataset. The BA prediction has largely been conducted in developed countries, and many studies have built predictive models using cross-sectional data with nonclinical routine predictors such as telomere length and DNA. This study utilized the appropriate statistical models to evaluate BA in longitudinal data, using Mixed Effects Linear Regression (MELR) and Long Short-Term Memory (LSTM). To evaluate the performance of the models, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used for both MELR with 13 predictors selected by stepwise selection and LSTM with all 17 predictors. Additionally, Bland-Altman limits of agreement were used to assess the closeness between Biological Age (BA) and Chronological Age (CA). Results of MELR and LSTM showed MAE of training and testing sets of 2.489 and 4.483, and 3.611 and 3.997, respectively. LSTM produces more stable performance between the training and test datasets than MELR, but MELR returns the lowest Bland-Altman agreement boundary (mean difference 0.00 years; 95% CI: -5.93, -5.93 years). Both models indicated that the EGAT 2 population was expected to be older than the CA population, based on predictors derived from commonly used clinical routines.

Keywords*: *Biological Age, Longitudinal Data Analysis, Mixed Effects Linear Regression, Long Short-Term Memory

1. Introduction

Nowadays, the world is facing a wave of an aging population. Based on the data from the World Health Organization (WHO), 703 million people aged over 65 years in 2019 will increase rapidly to 1.5 billion in 2050 (United Nations. Department of Economic and Social Affairs, Population Division, 2019, 2020). The aging population is always associated with a high burden of financial security and health cost (Hsu, Huang, & Yupho, 2015). A healthy aging population is one goal to help reduce the burden of costs and illnesses (Christensen, Doblhammer, Rau, & Vaupel 2009; Partridge, Deelen, & Slagboom 2018). For this reason, reliable information about the physiological condition will help to reach this goal (Ferrucci et al., 2020; Higgins-Chen et al., 2021).

For those situations, biological age (BA) can give reliable information about aging, which cannot be provided by chronological age (CA) (Fedintsev et al., 2017; Levine, 2013). While CA is measured by the date of birth (Galkin et al., 2020), BA is a quantitative measure of aging based on biological data having a high correlation with age (Bae et al., 2021; Galkin et al., 2020). So, biological data such as blood tests, DNA, telomere length, clinical examinations, and clinical image were used to estimate BA (Galkin et al., 2020). Numerous studies about BA have been using statistical models and machine learning (Bae et al., 2021; Fedintsev et al., 2017; Jee, & Park, 2017; Levine, 2013; Liu, 2021; Mamoshina et al., 2018; Ng et al., 2020; Zhong et al., 2020).

The statistical models that are widely used to predict BA are multiple linear regression (MLR) (Jee & Park, 2017; Levine, 2013), principal component analysis (PCA) (Jee, & Park, 2017; Levine, 2013; Mamoshina et al., 2018), including klemera and doubal's method (KDM) (Liu, 2021; Ng et al., 2020; Zhong et al., 2020). Levine (2013)'s study is the most cited BA study that estimates BA using MLR, PCA, and KDM with ten different predictors to estimate BA. Another study conducted uses a mixed–effect linear regression (MELR) model using six BAs (telomere length, DNA methylation age, cognitive function age, functional aging index (FAI), physiological age, and frailty index (FI); this study is one of the longitudinal studies on BA (Li et al., 2020). Most study predictors, such as telomere length and DNA methylation age, cannot be found easily in the clinical routine.

In the machine learning area, several models have been applied from deep neural network (DNN), 1D convolutional neural network (1D – CNN), 2D – CNN, 3D – CNN, convolutional long short–term memory (ConvLSTM), and CNN + long short – term memory (LSTM) to estimate BA based on anthropometry, biomarkers, and physical activity. Previous research usually uses 1D – CNN as a based model, assuming all data in previous studies are complete data without any missed time steps (Rahman et al., 2021). With the exception of machine learning models utilized in estimating BA (Shahid, Zameer, & Muneeb, 2020), other models such as LSTM have not been explored yet. LSTM is a suitable model for processing sequential data, such as longitudinal data, which can accommodate missing time-steps (Hochreiter, & Schmidhuber, 1997; Shahid et al., 2020).

In the previous studies, BA will be predicted using several predictors (i.e., blood test, behavior test, physical examination, and other laboratory tests) as independent variables. CA was commonly used as the dependent variable in the equations to predict BA (Rahman et al., 2021; Jee, & Park, 2017; Levine, 2013; Park, Cho, Kwon, & Lee, 2009). Previous MLR models assume that BA is the same as the predicted CA. In other words, BA will be linearly constructed with CA (Jia, Zhang, & Chen, 2017; Zhong et al., 2020).

Nevertheless, most research on BA has been conducted in Western countries with better healthcare systems and higher life expectancy than in Asian countries, such as the United States (Rahman et al., 2021), Canada (Mitnitski, Howlett, & Rockwood, 2017) and Russia (Pyrkov et al., 2018). Similar to the Asian region, BA research is mostly carried out in developed countries such as South Korea (Park et al., 2009), China (Liu, 2021), and Singapore (Zhong et al., 2020). Limited research exists to evaluate the use of machine learning for longitudinal data analysis, particularly in terms of its ability to address multicollinearity issues.

This study aimed to investigate the effects of longitudinal predictors on BA in the Thai population. MELR and LSTM were used to identify the most important factors associated with BA over time, with consideration given to commonly used clinical predictors such as age, gender, lifestyle, and health conditions. The results of this research would provide insights into the factors influencing BA in the Thai population, which can be used to inform public health policy and practice.

2. Objectives

To develop mixed – effect linear regression (MELR) and long short – term memory (LSTM) to estimate biological age in the Thai population using commonly used clinical predictors (demographics, biomarkers, physical examination, and behavior features) in the longitudinal data.

3. Materials and Methods

This study is a retrospective cohort study known as the Electric Generating Authority of Thailand (EGAT). This study used EGAT 2 (EGAT 2/2, 2/3, 2/4, and 2/5) as a longitudinal dataset. The first survey (EGAT 2/2) was conducted in 2008 and is repeated every five years. Only participants between 30 - 90 years without cancer history would be included, and participants with missing age data would be excluded. The total number of participants was 2,844 and the total data was 8,113 observations, but not all participants took all the surveys. Consequently, the data was divided into 80% training data and 20% as testing dataset (Mamoshina et al., 2019). The training dataset was used to build statistical and machine-learning models. In addition, dataset testing was needed to inspect model's performance. Technically, it helps to see which models are overfitting, fit or underfitting (Joseph, 2022).

Predictors to estimate BA consist of demographic (sex), biomarkers (high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (Tg), total cholesterol (Chol), creatinine (Cre), albumin (Alb), blood urea nitrogen (BUN), fasting Blood Glucose (Glu), alkaline phosphatase (ALP), hemoglobin (Hgb)), physical examination (systolic blood pressure (SBP), diastolic blood pressure (DBP), waist circumference (WC), body mass index (BMI)), and behavioral predictors (alcohol, smoking). However, the stepwise selection was carried out as a feature selection method for MELR model. Feature selection dropped BMI, DBP, Cre, and ALP and left with only 13 predictors that would be included in MELR model. LSTM model will use all predictors without a feature selection process. The evaluation metrics in this study are the mean absolute error (MAE), the root means square error (RMSE), and the Bland - Altman limit of agreement with a mean difference (CA - BA) at 95% confidence intervals (CI) to illustrate the closeness between BA and CA. MELR was developed using R version R-4.1.0. LSTM was also developed using Python version 3.8.13. The Ethics Committee of the Faculty of Medicine Ramathibodi Hospital, Mahidol University, approved this study (COA. MURA2022/304).

4. Results and Discussion

The baseline characteristics of the EGAT 2 participants are shown in **Error! Reference source not found.**. Most participants were male (74.4%) with a mean age of 53.3 years. The mean WC was 87.9 cm and most participants smoked (54.9%). Mean SBP and DBP were recorded as 128.8 mmHg and 81.0 mmHg, respectively, with blood pressure measurements predominantly falling within the normal range.

MELR could predict BA with MAE and RMSE on the training dataset of 2.489 and 3.026, respectively. Table 1 shows the coefficients of all variables in the MELR model, which indicate that some variables, such as Chol, Hgb, and Alcohol have a negative direction. When the MELR model was used to calculate BA in the testing dataset, the MAE and RMSE became 4.483 and 5.585, respectively, as shown in Table 3. Despite having a nearly two-year gap between the training and testing periods, the mean difference of the lowest Bland-Altman agreement boundary remains highly similar across both datasets. For the training dataset, the mean difference (95% CI) is 0.00 (-5.93, 5.93) years; and for the testing dataset, the mean difference (95% CI) is -0.02 (-10.97, 10.93) years.

Table 1 Baseline characteristics of patients

Observation = 8,113 , Participants = 2,844

No	Variables	Range (Min-Max)	Mean (SD) or %	Missing (%)
	Demographic			
1	Age (year)	31 - 74	53.3 (6.77)	0
2	Sex (%)	0 - 1	74.7%	0
	Physical Examination			
3	Systolic blood pressure (SBP) mmHg	80 - 230	128.8 (18.4)	0.73
4	Diastolic blood pressure (DBP) mmHg	45 - 134	81.0 (10.8)	0.73
5	Waist circumference (WC) cm	57 - 135	87.9 (10.1)	0.89
6	Body mass index (BMI)kg	13.06 - 46.18	24.9 (3.8)	0.67
	Behavior			
7	Alcohol (%)	0 - 1	31.4%	0.60
8	Smoking (%)	0 - 1	54.9%	0.58
	Laboratories / Biomarkers			
9	High-density lipoprotein (HDL) mmol/L	16 - 170	53.7 (14.4)	0.83
10	Low-density lipoprotein (LDL) mmol/L	15.4 - 442	148.0 (40.2)	2.08
11	Triglycerides (Tg) mg/dL	27 - 1362	151.1 (104.4)	0.84
12	Total cholesterol (Chol) mg/dL	85 - 553	208.2 (62.0)	0.83
13	Creatinine (Cre) µmol/L	0.39 - 13.61	1.0 (0.3)	0.84
14	Albumin (Alb) g/dL	1.95 - 5.73	4.7 (0.3)	27.04
15	Blood urea nitrogen (BUN) mg/dL	3 - 66.9	12.6 (3.6)	27.04
16	Fasting Blood Glucose (Glu) mmol/L	56 - 596	101.4 (29.4)	0.36
17	Alkaline phosphatase (ALP) IU/L	18 - 400	73.8 (22.8)	27.04
18	Hemoglobin (Hgb) g/dL	6.40 - 20.5	14.2 (1.5)	27.09

LSTM showed different results from MELR model. Based on model performance, LSTM model showed stable performance between the training and test datasets. The MAE and RMSE of the training dataset are 3.611 and 4.539, respectively; while MAE and RMSE of the test dataset were 3.997 and 4.925, respectively (Table 3). The models showed good performance on both datasets. In addition, the mean Bland-Altman agreement limits in the training and test datasets are very different, but the lower and upper bounds look similar. For the training dataset, the mean difference (95% CI) was 0.79 (-7.97, 9.55) years; and for the test dataset, the mean difference (95% CI) was -0.36 (-9.99, 9.28) years.

Table 1 Results from MELR for BA prediction

No	Variable Name	Estimate	SE	p-value	95% CI	
					Lower	**Upper**
	Intercept	27.630	1.690	< 2e-16*	24.230	31.061
1	Systolic blood pressure (SBP)	0.107	0.004	< 2e-16*	0.100	0.115
2	Waist circumference (WC)	0.089	0.010	< 2e-16*	0.069	0.108
3	Low-density lipoprotein (LDL)	0.039	0.002	< 2e-16*	0.035	0.042
4	Total cholesterol (Chol)	-0.063	0.001	< 2e-16*	-0.065	-0.060
5	Blood urea nitrogen (BUN)	0.210	0.021	< 2e-16*	0.169	0.251
6	High-density lipoprotein (HDL)	0.095	0.006	< 2e-16*	0.083	0.107
7	Triglycerides (Tg)	0.009	0.001	< 2e-16*	0.007	0.011
8	Fasting Blood Glucose (Glu)	0.016	0.002	0.000*	0.011	0.021
9	Albumin (Alb)	1.310	0.281	0.000*	0.753	1.865
10	Alcohol	-0.705	0.170	0.000*	-1.037	-0.371
11	Sex	0.308	0.297	0.301	-0.275	0.890
12	Hemoglobin (Hgb)	-0.383	0.068	0.000*	-0.516	-0.250
13	Smoking	0.715	0.200	0.000*	0.323	1.108

Note: *p < 0.01

Table 2 The MAE and RMSE from MELR and LSTM on training and testing set

No	Model	MAE		RMSE	
		Training	**Testing**	**Training**	**Testing**
1	MELR	2.489	4.483	3.026	5.585
2	LSTM	3.611	3.997	4.539	4.925

MELR showed extraordinary results, such as the inverse relationship between alcohol, Hgb, and Chol with biological age. Most studies have found that alcohol supports aging, but limited research suggests it may also have health benefits. Alcohol has a low effect on cardiovascular disease (Brien et al., 2011; Ronksley et al., 2011; Schou et al., 2017) and all causes of mortality (White, 1999). Furthermore, it is possible to have a high Hgb and support being a younger BA, as a previous study showed that a high Hgb supports better physical performance (Corona et al., 2022; Hirani et al., 2016). In addition, previous research also showed an inverse relationship between high Chol with CVD in the elderly group (Yi, Yi, & Ohrr, 2019). The other study also found that high Chol has a low correlation with all–cause mortality (Schatz et al., 2001).

MAE and RMSE were used as evaluation metrics to see which model has model performance and prediction between CA and BA. Based on MAE and RMSE, the performance of the LSTM model was compared to that of MELR model. LSTM model showed lower MAE and RMSE in the training and testing dataset than MELR model. LSTM models displayed a more stable performance than MELR models. MAE of LSTM is 3.611 and 3.997 on the training and testing dataset. RMSE of LSTM model was 4.539 and 4.925 on the training and testing dataset. Meanwhile, the MELR model had a lower MAE of 2.489 and a RMSE of 3.026 in the training dataset. Then in the testing dataset, MELR showed a MAE of 4.483 and a RMSE of 5.585, which were over two years of error difference and occured model overfitting problem. Based on the previous conditions, the performance of the LSTM model was better than the MELR model.

Compared with previous studies, the MAE in this study was smaller than the previous studies. Modern machine learning models were used in three studies to report a difference in MAE showing more than five years and RMSE showing more than ten years (Rahman et al., 2021; Rahman, & Adjeroh, 2019). Nonetheless, the prediction error in this study was still acceptable compared to those studies.

Bland - Atlman's 95% CI in MELR (95% CI: -10.971, 10.929) was more comprehensive than LSTM (995% CI: -9.987, 9.276). 95% CI in both models was an extensive range but still acceptable compared to the previous study (Jee, & Park, 2017). Bland–Altman plots on both models predicted BA to be older than their CA in the test data set. This could be seen from the mean difference of -0.21 in the MELR model and -0.36 in the LSTM model. These findings indicated that the EGAT population has older Bas than their Cas (negative mean difference means that the predicted value is older than the true value).

5. Conclusion

The predictions generated from this study's MELR and LSTM models suggested that the EGAT 2 population was expected to be older than their CA. The LSTM model performed better than the MELR model with more stable MAE and RMSE values across training and testing sets showing less overfitting. LSTMs could be generalized when working with test and training datasets. In addition, the predictors used in this study were derived from easily accessible clinical routines, and this study also used the benefit of longitudinal data to calculate BA, which could be applied to patient follow-up care.

Future studies on BA in Thai populations are still needed to provide insight into how BA is in Southeast Asian people. The predictions of BA can be applied for survival analysis in the EGAT 2 population, such as the analysis of non-communicable disease case survival (hypertension, diabetes, and cancer).

6. Acknowledgements

This research is a part of the Master of Science in Data Science for Healthcare and Clinical Informatics (International Program) undertaken by Maria Magdalena Pradita Eka Kurniawati, at the Faculty of Medicine Ramathibodi Hospital, Mahidol University, Thailand. This research is also supported by Mahidol University, Pierre Fabre Foundation, and AeHIN Scholarships for Digital Health Excellence.

7. References

Bae, C. Y., Im, Y., Lee, J., Park, C. S., Kim, M., Kwon, H., ... & Kim, J. (2021). Comparison of Biological Age Prediction Models Using Clinical Biomarkers Commonly Measured in Clinical Practice Settings: AI Techniques Vs. Traditional Statistical Methods. *Frontiers in Analytical Science*, *1*, 709589 https://doi.org/10.3389/frans.2021.709589

Brien, S. E., Ronksley, P. E., Turner, B. J., Mukamal, K. J., & Ghali, W. A. (2011). Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ*, *342*, d636. https://doi.org/10.1136/bmj.d636

Christensen, K., Doblhammer, G., Rau, R., & Vaupel, J. W. (2009). Ageing populations: the challenges ahead. *The Lancet*, *374*(9696), 1196-1208. https://doi.org/10.1016/S0140-6736(09)61460-4

Corona, L. P., Andrade, F. C. D., da Silva Alexandre, T., de Brito, T. R. P., Nunes, D. P., & de Oliveira Duarte, Y. A. (2022). Higher hemoglobin levels are associated with better physical performance among older adults without anemia: a longitudinal analysis. *BMC geriatrics*, *22*(1), 1-9. https://doi.org/10.1186/s12877-022-02937-4

Fedintsev, A., Kashtanova, D., Tkacheva, O., Strazhesko, I., Kudryavtseva, A., Baranova, A., & Moskalev, A. (2017). Markers of arterial health could serve as accurate non-invasive predictors of human biological and chronological age. *Aging*, *9*(4), 1280-1292. https://doi.org/10.18632/aging.101227

Ferrucci, L., Gonzalez-Freire, M., Fabbri, E., Simonsick, E., Tanaka, T., Moore, Z., ... & de Cabo, R. (2020). Measuring biological aging in humans: A quest. *Aging cell*, *19*(2), e13080-e13080. https://doi.org/10.1111/acel.13080

Galkin, F., Mamoshina, P., Aliper, A., de Magalhães, J. P., Gladyshev, V. N., & Zhavoronkov, A. (2020). Biohorology and biomarkers of aging: Current state-of-the-art, challenges and opportunities. *Ageing Research Reviews*, *60*, 101050. https://doi.org/10.1016/j.arr.2020.101050

Higgins-Chen, A. T., Thrush, K. L., & Levine, M. E. (2021). Aging biomarkers and the brain. *Seminars in Cell & Developmental Biology*, *116*, 180-193. https://doi.org/10.1016/j.semcdb.2021.01.003

Hirani, V., Naganathan, V., Blyth, F., Le Couteur, D. G., Seibel, M. J., Waite, L. M., ... & Cumming, R. G. (2016). Low hemoglobin concentrations are associated with sarcopenia, physical performance, and disability in older Australian men in cross-sectional and longitudinal analysis: the Concord Health and Ageing in Men project. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *71*(12), 1667-1675. https://doi.org/10.1093/gerona/glw055

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hsu, M., Huang, X., & Yupho, S. (2015). The development of universal health insurance coverage in Thailand: Challenges of population aging and informal economy. *Social Science & Medicine*, *145*, 227-236. https://doi.org/10.1016/j.socscimed.2015.09.036

Jee, H., & Park, J. (2017). Selection of an optimal set of biomarkers and comparative analyses of biological age estimation models in Korean females. *Archives of Gerontology and Geriatrics*, *70*, 84-91. https://doi.org/10.1016/j.archger.2017.01.005

Jia, L., Zhang, W., & Chen, X. (2017). Common methods of biological age estimation. *Clinical interventions in aging*, *12*, 759-772. https://doi.org/10.2147/CIA.S134921

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(4), 531-538. https://doi.org/10.1002/sam.11583

Levine, M. E. (2013). Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age?. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *68*(6), 667-674. https://doi.org/10.1093/gerona/gls233

Li, X., Ploner, A., Wang, Y., Magnusson, P. K., Reynolds, C., Finkel, D., ... & Hägg, S. (2020). Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *eLife*, *9*, e51507. https://doi.org/10.7554/eLife.51507

Liu, Z. (2021). Development and Validation of 2 Composite Aging Measures Using Routine Clinical Biomarkers in the Chinese Population: Analyses From 2 Prospective Cohort Studies. *The Journals of Gerontology: Series A*, *76*(9), 1627-1632. https://doi.org/10.1093/gerona/glaa238

Mamoshina, P., Kochetov, K., Putin, E., Cortese, F., Aliper, A., Lee, W. S., ... & Zhavoronkov, A. (2018). Population Specific Biomarkers of Human Aging: A Big Data Study Using South Korean, Canadian, and Eastern European Patient Populations. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *73*(11), 1482-1490. https://doi.org/10.1093/gerona/gly005

Mamoshina, P., Kochetov, K., Cortese, F., Kovalchuk, A., Aliper, A., Putin, E., ... & Zhavoronkov, A. (2019). Blood biochemistry analysis to detect smoking status and quantify accelerated aging in smokers. *Scientific reports*, *9*(1), 1-10. https://doi.org/10.1038/s41598-018-35704-w

Mitnitski, A., Howlett, S. E., & Rockwood, K. (2017). Heterogeneity of human aging and its assessment. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *72*(7), 877-884. https://doi.org/10.1093/gerona/glw089

United Nations. Department of Economic and Social Affairs, Population Division. (2019). *World population prospects 2019*. Highlights. Retrieved form https://www.ined.fr/fichier/s_rubrique/29368/wpp2019_10.key.findings_embargoed.version.en.pdf

Ng, T. P., Zhong, X., Gao, Q., Gwee, X., Chua, D. Q. L., & Larbi, A. (2020). Socio-environmental, lifestyle, behavioural, and psychological determinants of biological ageing: the Singapore longitudinal ageing study. *Gerontology*, *66*(6), 603-613. https://doi.org/10.1159/000511211

Park, J., Cho, B., Kwon, H., & Lee, C. (2009). Developing a biological age assessment equation using principal component analysis and clinical biomarkers of aging in Korean men. *Archives of Gerontology and Geriatrics*, *49*(1), 7-12. https://doi.org/10.1016/j.archger.2008.04.003

Partridge, L., Deelen, J., & Slagboom, P. E. (2018). Facing up to the global challenges of ageing. *Nature*, *561*(7721), 45-56. https://doi.org/10.1038/s41586-018-0457-8

Rahman, S. A., Giacobbi, P., Pyles, L., Mullett, C., Doretto, G., & Adjeroh, D. A. (2021). Deep learning for biological age estimation. *Briefings in bioinformatics*, *22*(2), 1767-1781. https://doi.org/10.1093/bib/bbaa021

United Nations. Department of Economic and Social Affairs, Population Division. (2020). *World Population Ageing 2019*. Highlights. https://doi.org/10.18356/9df3caed-en

Pyrkov, T. V., Slipensky, K., Barg, M., Kondrashin, A., Zhurov, B., Zenin, A., ... & Fedichev, P. O. (2018). Extracting biological age from biomedical data via deep learning: too much of a good thing?. *Scientific reports*, *8*(1), 1-11. https://doi.org/10.1038/s41598-018-23534-9

Rahman, S. A., & Adjeroh, D. A. (2019, November). Estimating biological age from physical activity using deep learning with 3D CNN. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1100-1103). IEEE.

Rahman, S. A., & Adjeroh, D. A. (2019). Deep learning using convolutional LSTM estimates biological age from physical activity. *Scientific reports*, *9*(1), 1-15. DOI: 10.1038/s41598-019-46850-0

Ronksley, P. E., Brien, S. E., Turner, B. J., Mukamal, K. J., & Ghali, W. A. (2011). Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ, 342*, d671. https://doi.org/10.1136/bmj.d671

Schou, A. L., Mølbak, M. L., Schnor, P., Grønbæk, M., & Tolstrup, J. S. (2017). Alcohol consumption, smoking and development of visible age-related signs: a prospective cohort study. *J Epidemiol Community Health*, *71*(12), 1177-1184. https://doi.org/10.1136/jech-2016-208568

Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, *140*, 110212. https://doi.org/10.1016/j.chaos.2020.110212

Schatz, I. J., Masaki, K., Yano, K., Chen, R., Rodriguez, B. L., & Curb, J. D. (2001). Cholesterol and all-cause mortality in elderly people from the Honolulu Heart Program: a cohort study. *The lancet*, *358*(9279), 351-355. https://doi.org/10.1016/S0140-6736(01)05553-2

Zhong, X., Lu, Y., Gao, Q., Nyunt, M. S. Z., Fulop, T., Monterola, C. P., ... & Ng, T. P. (2020). Estimating biological age in the Singapore longitudinal aging study. *The Journals of Gerontology: Series A*, *75*(10), 1913-1920. https://doi.org/10.1093/gerona/glz146

White, I. R. (1999). The Level of Alcohol Consumption at Which All-Cause Mortality Is Least. *Journal of Clinical Epidemiology*, *52*(10), 967-975. https://doi.org/10.1016/S0895-4356(99)00076-1

Yi, S. W., Yi, J. J., & Ohrr, H. (2019). Total cholesterol and all-cause mortality by sex and age: a prospective cohort study among 12.8 million adults. *Scientific reports*, *9*(1), 1596. https://doi.org/10.1038/s41598-018-38461-y