

Received: 2018.09.19

Accepted: 2018.10.16

Published: 2019.03.17

# Sociodemographic Indicators of Health Status Using a Machine Learning Approach and Data from the English Longitudinal Study of Aging (ELSA)

Authors' Contribution:

Study Design A  
Data Collection B  
Statistical Analysis C  
Data Interpretation D  
Manuscript Preparation E  
Literature Search F  
Funds Collection G

ABCDEF 1 **Worrawat Engchuan**  
ABDF 2,3 **Alexandros C. Dimopoulos**  
AEF 2,4,5,6 **Stefanos Tyrovolas**  
BDEF 7,8 **Francisco Félix Caballero**  
DEF 4,5,6 **Albert Sanchez-Niubo**  
ABCDEFG 9 **Holger Arndt**  
EF 6,10,11 **Jose Luis Ayuso-Mateos**  
DEF 4,5,6 **Josep Maria Haro**  
DEF 12 **Somnath Chatterji**  
ABDEFG 2,13 **Demosthenes B. Panagiotakos**

1 The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada  
2 Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece  
3 Department of Informatics and Telematics, School of Digital Technology, Harokopio University, Athens, Greece  
4 Research, Innovation and Teaching Unit, Institut de Recerca Sant Joan de Déu, Carrer Dr. Antoni Pujadas, Barcelona, Spain  
5 Parc Sanitari Sant Joan de Déu, Universitat de Barcelona, Fundació Sant Joan de Déu, Sant Boi de Llobregat, Barcelona, Spain  
6 Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain  
7 Department of Preventive Medicine and Public Health, and Microbiology, Universidad Autónoma de Madrid, Madrid, Spain  
8 Instituto de Salud Carlos III, CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain  
9 Spring Techno GmbH and Co. KG, Bremen, Germany  
10 Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria Princesa (IP), Madrid, Spain  
11 Department of Psychiatry, Universidad Autónoma de Madrid, Madrid, Spain  
12 Health Metrics and Measurement, World Health Organization, Geneva, Switzerland  
13 Faculty of Health, University of Canberra, Canberra, Australia

**Corresponding Author:** Demosthenes B. Panagiotakos, e-mail: [dbpanag@hua.gr](mailto:dbpanag@hua.gr)

**Source of support:** The ATHLOS project has received funding from the European Union Horizon 2020 Research and Innovation Program under grant agreement No. 635316 (EU HORIZON2020-PHC-635316)

**Background:** Studies on the effects of sociodemographic factors on health in aging now include the use of statistical models and machine learning. The aim of this study was to evaluate the determinants of health in aging using machine learning methods and to compare the accuracy with traditional methods.

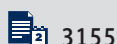
**Material/Methods:** The health status of 6,209 adults, age <65 years (n=1,585), 65–79 years (n=3,267), and >80 years (n=1,357) were measured using an established health metric (0–100) that incorporated physical function and activities of daily living (ADL). Data from the English Longitudinal Study of Ageing (ELSA) included socio-economic and sociodemographic characteristics and history of falls. Health-trend and personal-fitted variables were generated as predictors of health metrics using three machine learning methods, random forest (RF), deep learning (DL) and the linear model (LM), with calculation of the percentage increase in mean square error (%IncMSE) as a measure of the importance of a given predictive variable, when the variable was removed from the model.

**Results:** Health-trend, physical activity, and personal-fitted variables were the main predictors of health, with the %IncMSE of 85.76%, 63.40%, and 46.71%, respectively. Age, employment status, alcohol consumption, and household income had the %IncMSE of 20.40%, 20.10%, 16.94%, and 13.61%, respectively. Performance of the RF method was similar to the traditional LM (p=0.7), but RF significantly outperformed DL (p=0.006).

**Conclusions:** Machine learning methods can be used to evaluate multidimensional longitudinal health data and may provide accurate results with fewer requirements when compared with traditional statistical modeling.

**MeSH Keywords:** **Artificial Intelligence • Data Interpretation, Statistical • Decision Support Techniques • Socioeconomic Factors**

**Full-text PDF:** <https://www.medscimonit.com/abstract/index/idArt/913283>



## Background

As the global population is getting older, the study of health and aging has become increasingly important, particularly in planning current and future healthcare resources. People of 65 years of age and older now represent an increasing proportion of the population, particularly in Europe, Asia, and the USA [1–6]. Changing age demographics poses a dramatic shift towards an increased health burden of non-communicable diseases and disability [7,8]. Therefore, a current public health challenge is to identify health-related factors and to understand how to maintain a healthy life with increasing age. Sociodemographic factors, which include the employment status, household income, level of education, marital status, and social support are recognized major determinants of many health outcomes, including healthy aging [9,10].

Several analytical models have been proposed to evaluate healthy aging in relation to lifestyle characteristics, as well as biological, genetic and clinical factors, based on classical statistical hypothesis testing [11–13]. However, residual confounding and unexplained health risks are a common problem in almost all these hypothesis-derived models. Recently, the use of health informatics has received increasing attention as it allows for collection and analysis of large amounts of data and can extract data on patterns of risk that are free from the strict methodological assumptions of traditional statistical modeling methods [14,15]. In particular, machine learning offers a data-driven approach to allow for the analysis of patterns of health-associated variables and can provide insight into data without forming an *a priori* defined hypothesis regarding the involved variables [16–19]. There are several machine learning algorithms that are used to analyze health data, including support vector machine, decision tree, random forest (RF), the linear model (LM), and more recently, deep learning (DL) [20]. Choosing a machine algorithm for a particular analytical problem is important, as these models have rarely been compared before in terms of their efficiency and accuracy.

The present study was part of the Ageing Trajectories of Health: Longitudinal Opportunities and Synergies (ATHLOS) project (<http://athlosproject.eu/>). The aims of this study were to evaluate the sociodemographic determinants of healthy aging using three machine learning methods, RF, DL, and LR, and to compare these methods in terms of their efficiency. The working dataset was the English Longitudinal Study of Ageing (ELSA), which includes six waves of longitudinal data from 6,209 adults, from between 2002–2012 [21,22]. A previously developed and validated health metric of aging was used as the outcome, which was developed based on characteristics including physical function, activities of daily living (ADL), and instrumental activities of daily living (IADL) [21,23].

## Material and Methods

This study was conducted according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement [24].

### Working dataset

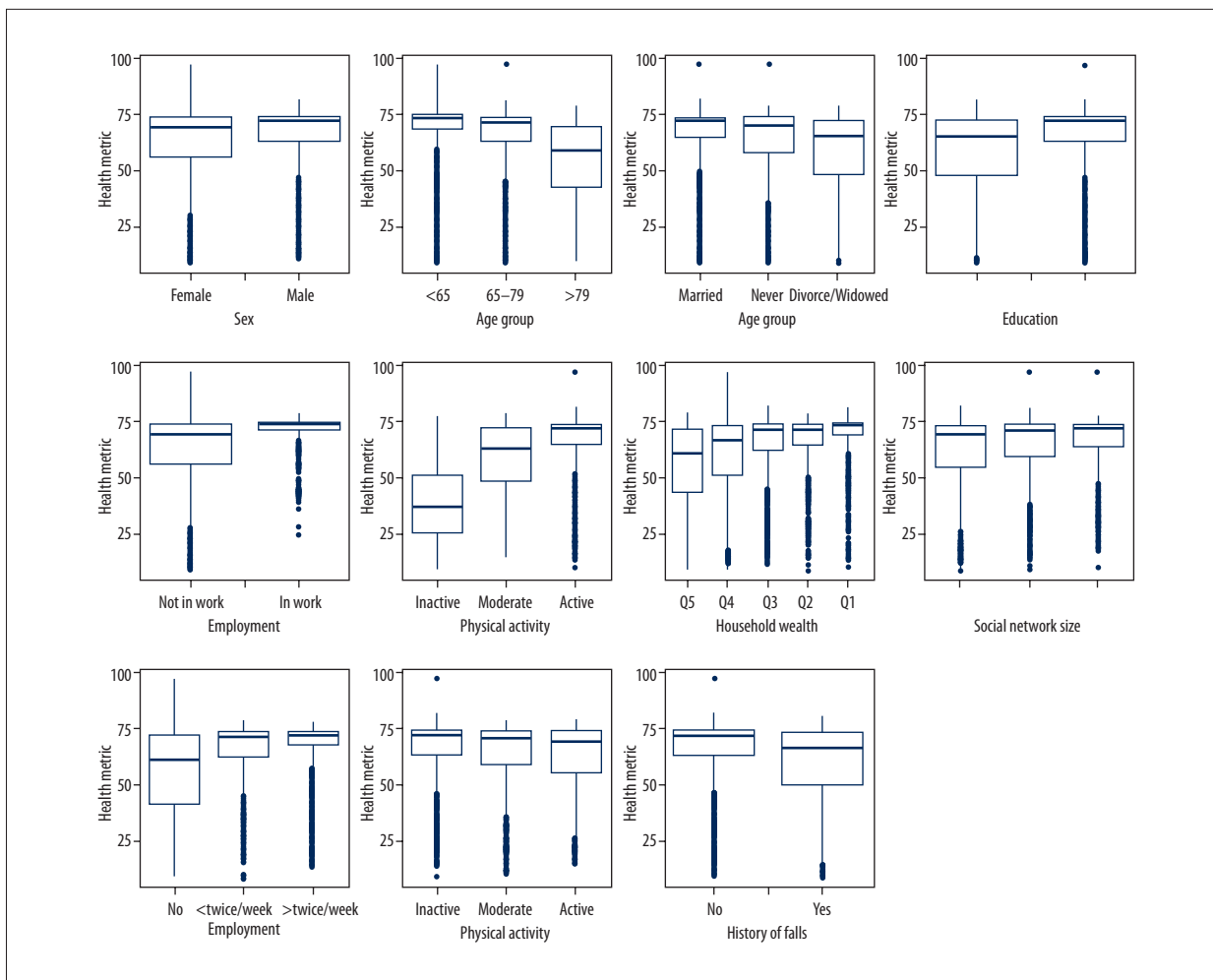
The English Longitudinal Study of Ageing (ELSA) dataset was used to test the research hypothesis, which included six waves of longitudinal data from 6,209 adults [21]. Only subjects who had at least five waves of longitudinal data were used. The number of individuals included in the study analysis was 6,209. The health metric in the last wave of data collection for each subject was treated as the target health metric to be predicted.

The sociodemographic indicators used included; gender, men (n=2,638) and women (n=3,571); age group <65 years (n=1,585), 65–79 years (n=3,267), >80 years (n=1,357); quintiles of household income; formal education (yes or no); marital status, (married, single, and other); smoking history (never smoked, former smoker, current smoker); alcohol consumption (non-drinker, drinking alcohol twice weekly or less, drinking a regular glass of wine glass or equivalent of 12% alcohol more than twice weekly); physical activity (inactive, moderately active, and active); employment status (employed or unemployed); size of social network (number of relatives and close relationships, small <5 people, moderate 5–9 people, large >9 people).

In addition, to sociodemographic data, a history of falls was also incorporated as a possible confounder. More information on discretization of each characteristics can be found in a previous paper [21]. The health metric was developed using item response theory (IRT) using questionnaire data on individual physical functionality, activities of daily living (ADL), and instrumental activities of daily living (IADL), as previously described [23]. The range of the metric was set to 0–100. Figure 1 shows the distribution of the health metric, stratified according to each of the studied sociodemographic characteristics. The increasing trend of the health metric over the increasing value of the sociodemographic characteristic, suggests a positive relationship.

### Feature engineering to develop the new predictors of the personal-fitted variable and the health trend variable

Feature engineering is the process of generating a new predictor that is higher order and more meaningful than existing ones [25]. The sociodemographic characteristics and health metrics from previous four waves of data were used to generate two new predictors by modeling the relationship between predictors, and also the trend of the health metric. The first type of new predictor was a personal-fitted variable, which



**Figure 1.** Global health metric distribution stratified by each value in each predictor. The boxplots show the distribution of the health metrics stratified by the unique value of the predictors. The difference in the distribution within the predictor suggests a relationship between the predictor and the health metric.

captured the relationship between the sociodemographic data and the health metric for each subject. An individual health metric prediction model was built using data from the previous waves, as shown in model 1. The variable was then generated as a prediction from the model on the current wave of data. A linear model (LM) was used to generate the new variable (model 1).

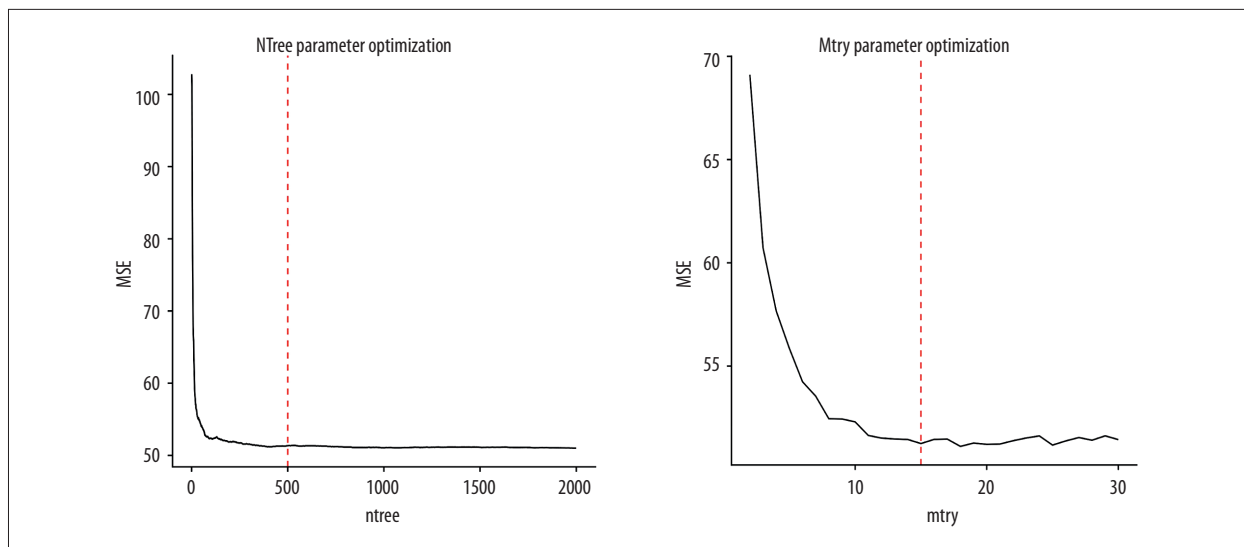
health metric ~ gender + age group + quintile of household wealth + formal education + marital status + falls + smoking behavior + alcohol consumption + physical activity + employment + size of social network (model 1)

The notation of the linear model (model 1) is described as the regressing gender, age group, quintile of household wealth, formal education, marital status, fall, smoking behavior, alcohol consumption, physical activity, employment and size of social network on health metric.

The second type of new predictor was the health trend variable, which captured the trend of the health metric for an individual subject. A health metric prediction model was built by fitting the health metric with the time from the previous four waves of data, as shown in model 2. The variable was generated as a prediction of the model at the current time point. Another linear model was used to generate the variable (model 2).

health metric ~ time (model 2)

In total, there were 13 predictors, of which 11 were sociodemographic determinants, one personal-fitted variable, and one trend variable.



**Figure 2.** Parameters optimization by grid search. The left panel represents the mean squared error (MSE) change over different values of the number of trees used to build a random forest model (ntree). The right panel represents the MSE change for the different values of the number of predictors randomly picked at each branch of the tree (mtry). The optimal parameters (ntree=500) (mtry=15) were used in the final model.

### Health metric prediction using random forest (RF), linear model (LM) and deep learning (DL) models

Random forest (RF), linear model (LM), and deep learning (DL) were applied to evaluate sociodemographic determinants of the health metric. RF is an ensemble learner of decision tree (DT), which identifies a predictor by majority data. RF was implemented in this study by using the RF library of R [26]. Two RF parameters were ntree, the number of trees used to build a random forest model, and mtry, the number of predictors randomly picked at each branch of the tree. Both ntree and mtry were optimized in the training data by undertaking grid search (Figure 2). The optimal parameters (ntree=500, and mtry=15) were used in the final model.

LM is a statistical approach that builds a learning model by fitting beta coefficients of the predictors on linear relationships between them and a target class. LM is a simple and fast approach to build a model. However, as its name suggests, this approach only performs well when the problem has a linear relationship. The R statistical function, lm, of the R Stats Package was used to perform linear regression in this study [27].

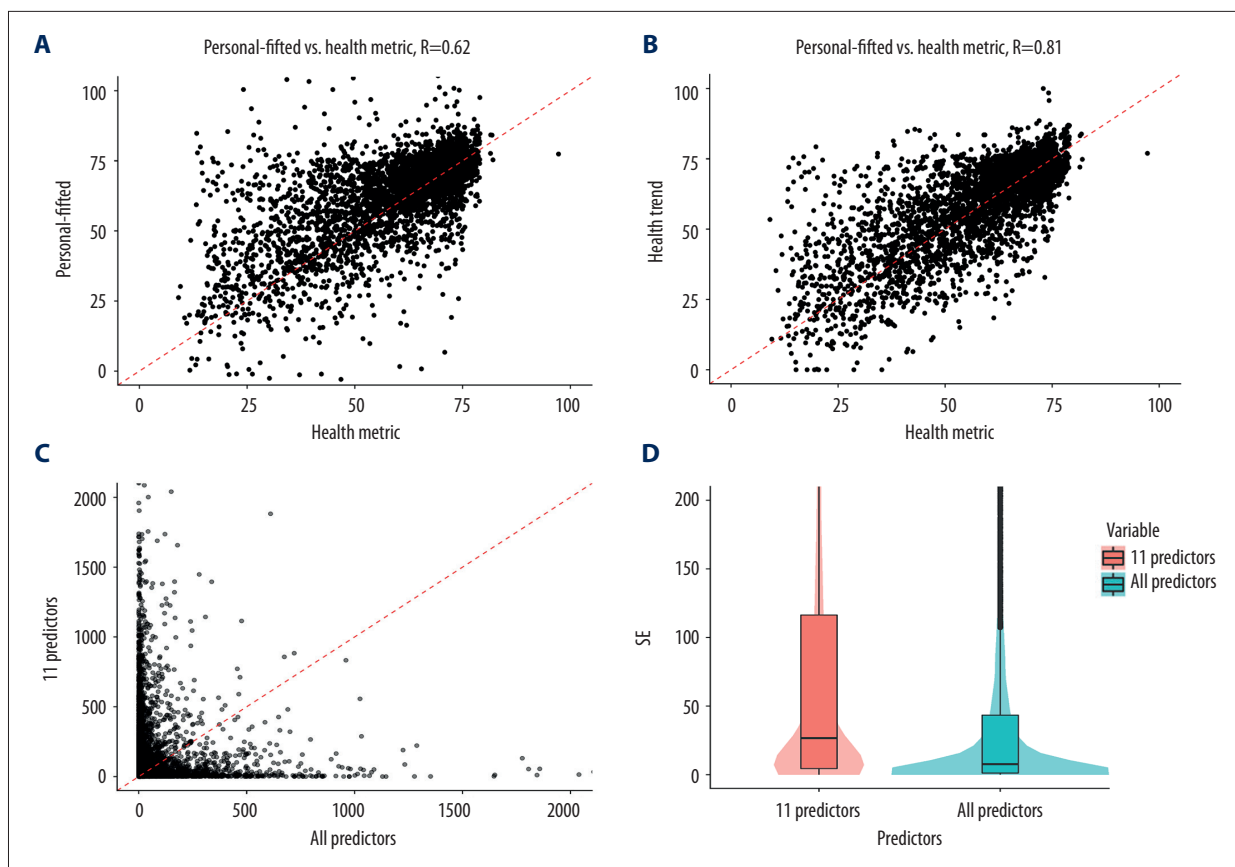
DL is an established algorithm that mimics a biological neural network. Multiple cascading layers of neurons are connected and pass information from one layer to the next by transforming and extracting new predictors. DL has been shown to be superior to other algorithms in many applications [20]. This study implemented DL using keras library in R [28]. The structure of the DL model consisted of five layers (13 neurons with relu activation function, 10% neurons dropout, 5 neurons with linear

activation function, 5% neurons dropout, 1 output neuron). The parameters used to build the model were optimized using the Adam optimization algorithm using a batch size of 128 (default). The size of the validation set was 20% of the training set.

The performance assessment was undertaken using 10-fold cross-validation to robustly estimate the performance by using ten equal random subsets, or folds, of the whole dataset. One of the folds was used as a test set and the remainder were used as a training set. The process was repeated ten times until all folds were used as a test set, then the final result was reported.

### Assessment of predictor performance

Understanding which predictors in the model are driving the performance and how they are doing so can help to adjust the model, so that the performance can be improved. The effect of each predictor was assessed by obtaining the standardized coefficient of the LM model by scaling each predictor using the scale function of the base library in R, and the variable importance of RF was assessed using the importance function of the randomForest library in R [26]. The magnitude of the standardized coefficient of the LM model represented the effect size of a predictor on the health metric, while the sign represented the direction of the relationship. The variable importance of RF assessed the effect of a predictor by removing the predictor from every single tree in the forest and measuring how the accuracy changed. The effect was reported as the percentage increase in mean square error (%IncMSE), where more important predictors have higher %IncMSE [26].



**Figure 3.** Contribution of historical, personal-fitted, and health trend features. The scatter plots (A–C) illustrate the relationship between the personal-fitted predictor (A), health trend predictor (B), the prediction from 11 predictors (C), and the health metric. The boxplot (D) shows the squared errors (SE) from 11 predictors and all predictors.

## Results

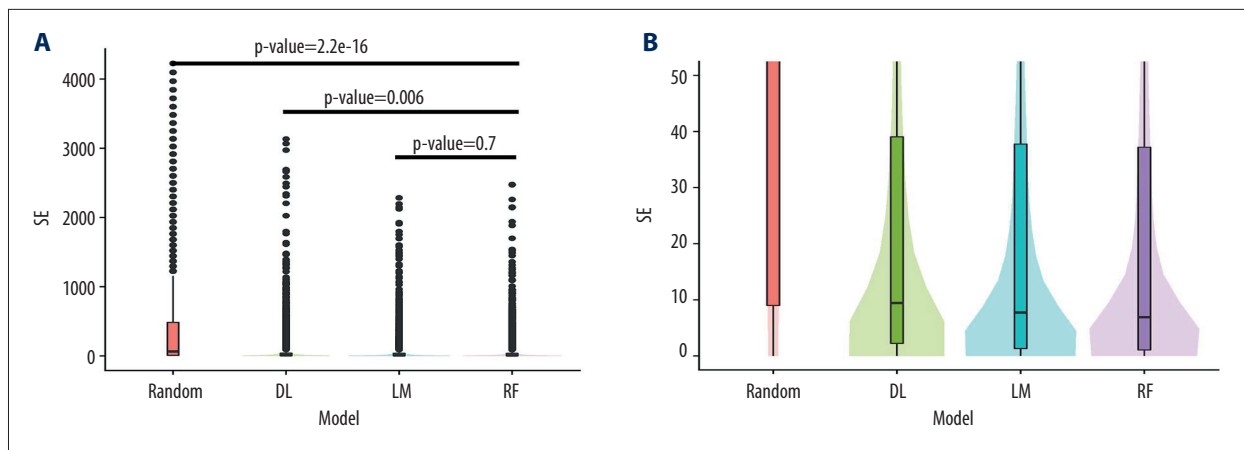
### Sociodemographic factors and health metrics

The health metric varied from  $65.95 \pm 13.56$  in men to  $62.83 \pm 5.01$  in women ( $p < 0.001$ ),  $68.92 \pm 11.31$  in people  $< 65$  years,  $65.75 \pm 12.79$  in people between 65–75 years, and  $54.73 \pm 17.24$  in people  $> 79$  years ( $p < 0.001$ ). The health metrics were  $56.36 \pm 17.05$ ,  $61.41 \pm 15.22$ ,  $64.97 \pm 13.02$ ,  $66.84 \pm 11.98$ , and  $69.60 \pm 9.89$ , respectively in the 1<sup>st</sup> to 5<sup>th</sup> quintiles of household wealth, ( $p < 0.001$ );  $66.24 \pm 13.06$  in educated people, to  $59.05 \pm 16.37$  in uneducated people ( $p < 0.001$ );  $66.68 \pm 12.65$  in married people,  $63.50 \pm 14.91$  in single people, and  $59.30 \pm 16.46$  in previously married people ( $p < 0.001$ ). The health metrics were  $65.40 \pm 14.10$  in non-smokers,  $63.72 \pm 14.56$  in former smokers, and  $62.23 \pm 15.13$  in current smokers ( $p < 0.001$ ). The health metrics were  $55.91 \pm 18.73$  in non-alcohol drinkers,  $66.18 \pm 11.93$  in people who drank twice or less each week, and  $68.86 \pm 9.56$  in those who drank more than twice a week of a regular glass of wine or the equivalent of 12% alcohol ( $p < 0.001$ ). The health metrics were  $72.21 \pm 5.52$  in those who were employed and  $62.56$  in the unemployed ( $p < 0.001$ ). The health metrics were

$40 \pm 17.34$  in the non-active,  $59.21 \pm 14.95$  in the moderately active, and  $68 \pm 10.01$  in the active population ( $p < 0.001$ ). The social health metrics were  $62.41 \pm 15.53$  for people with a small social network to  $66.77 \pm 11.69$  for people with a large social network ( $p < 0.001$ ). The health metrics were  $65.70 \pm 13.68$  for people without a history of falls and  $59.84 \pm 15.79$  with a history of falls ( $p < 0.001$ ).

### Sociodemographics, personal-fitted, and health trend as health predictors

Before comparing different prediction models, validation that the newly extracted variables, personal-fitted, and health trend, added information to the model. The scatter plots presented in Figure 3A and 3B show the correlation between personal-fitted, health trend, and health metric, respectively. The health trend variable was more closely correlated (0.81) with the health metric than the personal-fitted variable (0.62). Figure 3C shows the square error (SE) of the model with sociodemographic and extracted variables. Having more variables changed the performance in most cases, which also showed improvement in performance (more values in the



**Figure 4.** Performance comparison between three prediction models and random prediction. (A) Box plots show the distribution of squared errors of the random prediction model, deep learning (DL), the linear model (LM), and the random forest (RF) model. (B) A magnified version of (A) shows the difference between the DL, LM, and RF. Student's t-test was used to calculate the P-values.

upper part of the diagonal line) (Figure 3D). From the distribution of SEs, the model with the new variables was significantly lower than the model with only sociodemographic characteristics ( $p < 0.001$ ) [19].

#### Comparison of the machine learning prediction methods

To determine whether one model was better than another, the mean square error (MSE) was calculated [29]. The higher the MSE the model had, the worse was the performance of the model. Also random prediction was done as a baseline. The random prediction was done by label permutation to maintain the health metric distribution. From 10-fold cross-validation, RF, LM and DL were much better than random prediction (Figure 4A, 4B). The best model was RF, with MSE of 51.11, while LM, DL and random prediction had a higher MSE of 52.07, 59.08 and 418.40, respectively. Figure 4 shows the performance of each model by the distribution of their SEs. The SEs of RF were significantly lower when compared with the random predictions ( $P < 0.001$ ), and also significantly lower when compared with DL ( $P = 0.006$ ), but were comparable with LM ( $P = 0.7$ ).

#### Predictor importance assessment

To understand the model, variable importance from RF and the standardized coefficient from LM were required. The health trend, physical activity, and personal-fitted variables were the main predictors of health metrics with  $\%incMSE$  of 85.76%, 63.40%, and 46.71%, respectively. Age, employment status, alcohol consumption, and household income were also the main characteristics that help determine the health metric with  $\%incMSE$  of 20.40%, 20.10%, 16.94%, and 13.61%, respectively. Table 1 shows all the  $\%incMSE$  and the standardized coefficients of the predictors. The overall ranking of predictors

by  $\%incMSE$  and the standardized coefficient were aligned, except for social network size characteristics which were less important than age, employment status, alcohol consumption, and household income.

The results of the predictor importance assessment confirmed the hypothesis that having health trend and personal-fitted variables would help improve the performance of the model. Also, the effects of most of the sociodemographic characteristics found in a previous study [21], were confirmed by these findings.

#### Discussion

This study aimed to build a predictive model to accurately estimate health status based on sociodemographic characteristics in an aging population using data from the English Longitudinal Study of Ageing (ELSA), which included socioeconomic and sociodemographic characteristics and history of falls. Because the dataset analyzed was quite large, consisting of more than 6,000 samples, this allowed the analysis of many potential predictive factors to obtain the best health metric prediction model. However, although the sample size was quite large, the number of time points was very limited. As a result, the model was unable to recognize the changes in health patterns over time, and also many time series data analytical approaches could not be applied.

This study included three machine learning methods, random forest (RF), deep learning (DL) and the linear model (LM), with calculation of the percentage increase in mean square error ( $\%incMSE$ ) as a measure of the importance of a given predictive variable, when the variable was removed from the model. An advantage of using RF was that it was transparent and the

**Table 1.** Summarized %incMSE and coefficients by predictors.

Predictors	%incMSE	Standardized coefficients
Health trend (health metric estimated by 4 previous health metric)	85.76	8.13
Physical activity (active vs. moderate vs. inactive)	63.40	3.30
Personal-fitted variable (health metric estimated by 11 socio-demographics and fall history of previous 4 waves)	46.71	1.54
Age groups (<65 vs. 65–79 vs. >79)	20.40	-0.61
Employment (in work vs. not in work)	20.10	0.51
Alcohol consumption	16.94	0.74
Quantiles Household Wealth (Q1–Q5)	13.61	0.46
Social network size (<5 vs. 5–9 vs. >9 people)	7.16	1.22
Falls (have fall history vs. no fall history)	5.30	-0.60
Marital status (married vs. never married vs. others)	5.29	-0.03
Smoke (never vs. former smoker vs. current smoker)	4.54	-0.19
Sex (males vs. females)	4.35	0.32
Education (no qualification vs. some formal education)	2.86	0.05

importance of the variables could be assessed. DL has previously been reported to outperform state-of-the-art algorithms, but this was not the case for this dataset, which might have been due to the simplicity of its implementation, as the optimization was performed as recommended by Keras [28]. The choice of optimizers is also very important, and there are different optimizers that may help to improve the networks, such as stochastic gradient descent (SGD), AdaGrad, and RMSProp, so further studies are needed to identify the best optimizer to adjust network structure and parameters [30]. A further reason for the poor performance of DL might be due to the limited number of predictors used, as DL is a powerful learning method but that requires a significant amount of data [20]. A disadvantage of the use of DL is its complexity, which can make it difficult to interpret. LM and RF are more transparent and may be preferred. LM is commonly used in this field of study, and its performance is comparable to RF, while its standardized coefficient is more informative. The standardized coefficient not only can indicate how important the predictor is but also can indicate the direction of the relationship between the predictor and the outcome variable.

Adding newly extracted variables improved the performance of the model. Using health trend as a predictor helped to estimate the health metric for each subject, provided that the health metric of the subject was constant. However, if the current health metric significantly deviated from the previously, predicting the health metric based only on the health trend variable would be difficult. Using a personal-fitted variable that captured the

relationship between the changes of the sociodemographic and health metrics might also help to reduce the error rate in cases where the current health metric suddenly changed. However, the change in the current health metric needs to be the product of a change in some of the sociodemographic characteristics to be able to be captured by the personal-fitted variable. If there is a change in the health metric from anything other than factors in the captured sociodemographic characteristics, the health trend variable might be helpful.

However, in this study, because the number of time points was small, four time points for the training set were used, while at least 50 time points have been recommended for the autoregressive integrated moving average (ARIMA) statistical analysis model to reliably recognize patterns [31]. As the results of this study showed, increasing the number of time points might have helped to reduce the error in this study. Therefore, in future, a more efficient way to capture the sudden change in current health metrics needs to be applied. Either more data points for each subject need to be collected so that a health trend can be recognized, or additional features that can better capture the changes in the health metric should be acquired.

## Conclusions

This study investigated the application of machine learning algorithms to accurately predict a health metric related to health status using sociodemographic in an aging population.

Data from the English Longitudinal Study of Ageing (ELSA) were used and the study was part of the Ageing Trajectories of Health: Longitudinal Opportunities and Synergies (ATHLOS) project. Personal-fitted and health trend variables were incorporated in the model, which were shown to be beneficial. Three prediction methods, random forest (RF), deep learning (DL) and the linear model (LM), with the calculation of the percentage increase in mean square error (%IncMSE) were applied, and the best results were achieved and tested from RF. DL may be superior in other studies, but a significant amount of data and expertise are required. Different parameter optimization techniques can be applied and more predictors can be added to improve the current DL model. The recommended setting can be applied to other datasets in the ATHLOS project, in case those datasets have similar characteristics as the ELSA dataset. For the non-similar datasets, the testing procedure used in this study can be performed to find the most suitable setting for the dataset.

## References:

1. Population structure and ageing – Statistics Explained. Available from: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Population\\_structure\\_and\\_ageing#Further\\_Eurostat\\_information](http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing#Further_Eurostat_information)
2. Mather M: Fact Sheet: Aging in the United States (2015). URL:<http://www.prb.org/Publications/Media-Guides/2016/aging-unitedstates-fact-sheet.aspx>
3. Asian Development Bank, Population and Aging in Asia: The Growing Elderly Population. Article, 18 Jan 2017. URL:<https://www.adb.org/features/asia-s-growing-elderly-population-adb-s-take>
4. World Health Organization (WHO). Men, ageing and health: Achieving health across the life span, 2008. URL:<http://www.who.int/ageing/publications/men/en/>
5. United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Ageing 2015 (ST/ESA/SER.A/390). URL:[http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015\\_Report.pdf](http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf)
6. World Health Organization (WHO). Global status report on noncommunicable diseases (2010). URL:[http://apps.who.int/iris/bitstream/10665/44579/1/9789240686458\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/44579/1/9789240686458_eng.pdf)
7. The Institute for Health Metrics and Evaluation (IHME). Washington, D.C. USA. Global Health Data. URL:<http://www.healthdata.org/>
8. Salminen A, Hyttinen JM, Kaarniranta K: AMP-activated protein kinase inhibits NF- $\kappa$ B signaling and inflammation: Impact on healthspan and lifespan. *J Mol Med (Berl)*, 2011; 89(7): 667–76
9. Seeman TE, Crimmins E, Huang MH et al: Cumulative biological risk and socioeconomic differences in mortality: MacArthur studies of successful aging. *Soc Sci Med*, 2004; 58(10): 1985–97
10. Wu MS, Lan TH, Chen CM et al: Sociodemographic and health-related factors associated with cognitive impairment in the elderly in Taiwan. *BMC Public Health*, 2011; 11: 22
11. Ravussin E, Redman LM, Rochon J et al: A 2-year randomized controlled trial of human caloric restriction: feasibility and effects on predictors of health span and longevity. *J Gerontol A Biol Sci Med Sci*, 2015; 70(9): 1097–104
12. McGeer PL, Schulzer M, McGeer EG: Arthritis and anti-inflammatory agents as possible protective factors for Alzheimer's disease: A review of 17 epidemiologic studies. *Neurology*, 1996; 47(2): 425–32
13. Holzinger A: Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Inform*, 2016; 3(2): 119–31
14. Lee KS, Lee BS, Semmani S et al: Curcumin extends life span, improves health span, and modulates the expression of age-associated aging genes in *Drosophila melanogaster*. *Rejuvenation Res*, 2010; 13(5): 561–70
15. Mathias JS, Agrawal A, Feinglass J et al: Development of a 5-year life expectancy index in older adults using predictive mining of electronic health record data. *J Am Med Inform Assoc*, 2013; 20(e1): e118–24
16. Pogorelc B, Bosnić Z, Gams M: Automatic recognition of gait-related health problems in the elderly using machine learning. *Multimed Tools Appl*, 2012; 58: 333
17. Song X, Mitnitski A, Cox J, Rockwood K: Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Stud Health Technol Inform*, 2004; 107(Pt 1): 736–40
18. Smith JP, Kington R: Demographic and economic correlates of health in old age. *Demography*, 1997; 34(1): 159–70
19. Kotsiantis SB: Supervised machine learning: A review of classification techniques. *Informatica*, 2007; 31: 249–68
20. LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature*, 2015; 521(7553): 436–44
21. Caballero FF, Soulis G, Engchuan W et al: Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: The ATHLOS project. *Sci Rep*, 2017; 7: 43955
22. Steptoe A, Breeze E, Banks J, Nazroo J: Cohort profile: The English longitudinal study of ageing. *Int J Epidemiol*, 2013; 42(6): 1640–48
23. de la Fuente J, Caballero FF, Sánchez-Niubó A et al: Determinants of health trajectories in England and the US: An approach to identify different patterns of healthy aging. *J Gerontol A Biol Sci Med Sci*, 2018; 73(11): 1512–18
24. von Elm E, Altman DG, Egger M et al; STROBE Initiative: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Int J Surg*, 2017; 12(12): 1495–99
25. Yu H-F, Lo H-Y, Hsieh J-K et al: Feature engineering and classifier ensemble for KDD Cup 2010. URL:<http://pslcdatashop.org/KDDCup/workshop/papers/kdd2010ntu.pdf>
26. Liaw A, Wiener M: Classification and regression by randomForest. *R News*, 2002; 2/3: 18–22. URL:[https://www.r-project.org/doc/Rnews/Rnews\\_2002-3.pdf](https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf)
27. The R Foundation. The R Project for Statistical Computing. URL:<https://www.r-project.org>
28. Keras FC: Deep learning library for the ano and tensorflow 2015. URL:<https://keras.io>
29. Draper NR, Smith H: Applied regression analysis. John Wiley & Sons, 2014
30. Wilson AC, Roelofs R, Stern M et al: The marginal value of adaptive gradient methods in machine learning. The 31<sup>st</sup> Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. MIT Press. 2017. URL:<https://papers.nips.cc/paper/7003-the-marginal-value-of-adaptive-gradient-methods-in-machine-learning.pdf>
31. Box GEP, Tiao GC: Intervention analysis with applications to economic and environmental problems. *J Am Stat Assoc*, 1975; 70(349): 70–79

## Acknowledgments

This study was conducted within the Ageing Trajectories of Health: Longitudinal Opportunities and Synergies (ATHLOS) project.

## Statement

The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of the World Health Organization.

## Conflict of interest

The authors declare that they have no conflict of interest.