

EUR Research Information Portal

Deep learning predicted perceived age is a reliable approach for analysis of facial ageing

Published in:

Journal of the European Academy of Dermatology and Venereology

Publication status and date:

E-pub ahead of print: 03/10/2024

DOI (link to publisher):

[10.1111/jdv.20365](https://doi.org/10.1111/jdv.20365)

Document Version

Publisher's PDF, also known as Version of record

Document License/Available under:

CC BY-NC

Citation for the published version (APA):

Turner, C., Pardo, L. M., Gunn, D. A., Zillmer, R., Mekic, S., Liu, F., Ikram, M. A., Klaver, C. C. W., Croll, P. H., Goedegebure, A., Trajanoska, K., Rivadeneira, F., Kavousi, M., Brusselle, G. G. O., Kayser, M., Nijsten, T., & Bacardit, J. (2024). Deep learning predicted perceived age is a reliable approach for analysis of facial ageing: A proof of principle study. *Journal of the European Academy of Dermatology and Venereology*, 38(12), 2295-2302. Advance online publication. <https://doi.org/10.1111/jdv.20365>

[Link to publication on the EUR Research Information Portal](#)

Terms and Conditions of Use

Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.



In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy

If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

ORIGINAL ARTICLE

Deep learning predicted perceived age is a reliable approach for analysis of facial ageing: A proof of principle study

Conor Turner¹ | Luba M. Pardo²  | David A. Gunn³ | Ruediger Zillmer⁴ | Selma Mekic⁵ | Fan Liu⁶ | M. Arfan Ikram⁷ | Caroline C. W. Klaver^{7,8} | Pauline H. Croll^{7,9,10} | André Goedegebure¹⁰ | Katerina Trajanoska^{7,11}  | Fernando Rivadeneira¹¹ | Maryam Kavousi⁷ | Guy G. O. Brusselle^{7,12,13} | Manfred Kayser⁶ | Tamar Nijsten² | Jaume Bacardit¹

¹Interdisciplinary Computing and Complex BioSystems (ICOS) Research Group, School of Computing, Newcastle University, Newcastle upon Tyne, UK

²Department of Dermatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

³Unilever King's Biosciences Innovation Hub, King's College London, London, UK

⁴Unilever Research and Development, Port Sunlight, UK

⁵Department of Dermatology, University Medical Center Utrecht, Utrecht, The Netherlands

⁶Department of Genetic Identification, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

⁷Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

⁸Department of Ophthalmology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

⁹Department of Radiology and Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

¹⁰Department of Otorhinolaryngology, Head and Neck Surgery, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

¹¹Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

¹²Department of Respiratory Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands

¹³Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium

Correspondence

Jaume Bacardit, Interdisciplinary Computing and Complex BioSystems (ICOS) research group, School of Computing, Newcastle University, 1 Science Square, Newcastle upon Tyne NE3 2HT, UK.

Email: jaume.bacardit@newcastle.ac.uk,

Luba Milena Pardo, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands.

Email: l.pardocortes@erasmusmc.nl

Abstract

Background: Perceived age (PA) has been associated with mortality, genetic variants linked to ageing and several age-related morbidities. However, estimating PA in large datasets is laborious and costly to generate, limiting its practical applicability.

Objectives: To determine if estimating PA using deep learning-based algorithms results in the same associations with morbidities and genetic variants as human-estimated perceived age.

Methods: Self-supervised learning (SSL) and deep feature transfer (DFT) deep learning (DL) approaches were trained and tested on human-estimated PAs and their corresponding frontal face images of middle-aged to elderly Dutch participants ($n = 2679$) from a population-based study in the Netherlands. We compared the DL-estimated PAs with morbidities previously associated with human-estimated PA as well as genetic variants in the gene MC1R; we additionally tested the PA associations with MC1R in a new validation cohort ($n = 1158$).

Results: The DL approaches predicted PA in this population with a mean absolute error of 2.84 years (DFT) and 2.39 years (SSL). In the training–test dataset, we found the same significant ($p < 0.05$) associations for DL PA with osteoporosis, ARHL, cognition, COPD and cataracts and MC1R, as with human PA. We also found a

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Journal of the European Academy of Dermatology and Venereology* published by John Wiley & Sons Ltd on behalf of European Academy of Dermatology and Venereology.

similar but less significant association for SSL and DFT PAs (0.69 and 0.71 years per allele, $p=0.008$ and 0.011 , respectively) with MC1R variants in the validation dataset as that found with human, SSL and DFT PAs in the training–test dataset (0.79, 0.78 and 0.71 years per allele respectively; all $p < 0.0001$).

Conclusions: Deep learning methods can automatically estimate PA from facial images with enough accuracy to replicate known links between human-estimated perceived age and several age-related morbidities. Furthermore, DL predicted perceived age associated with MC1R gene variants in a validation cohort. Hence, such DL PA techniques may be used instead of human estimations in perceived age studies thereby reducing time and costs.

INTRODUCTION

Perceived facial age (PA) (how old people look in facial photographs) has been studied as a biomarker of facial ageing. It been used to investigate factors driving facial ageing (e.g. genetic and lifestyle factors).¹ It has been also linked to skin senescence² and to how visible signpost of ageing also reflect ageing within the body. Indeed, mortality,^{3–5} osteoporosis,^{6,7} CVD,^{7–9} cataracts⁷ and worse cognition^{3,7} have all been linked to those who look old for their age across multiple studies. In addition, such findings are more than cross-sectional findings, with links to mortality⁵ proven in longitudinal studies. Even in young adults, perceived age is linked to the rate of systemic ageing as determined by changes to blood biomarkers of health over a 4-year period.¹⁰ Such consistent links between facial ageing and health are suggestive that it could be used not only for facial ageing research but to help determine risks in a population to certain diseases, although clinical validation of its utility beyond current risk measures would be required. In addition, genetic variation associated with PA was also investigated pointing to the *MC1R* gene involvement in explaining PA variation.¹¹ The *MC1R* gene encodes for a G protein-coupled receptor that binds melanocortin and has main effects in pigmentation (red hair), skin photoaging and skin cancer.¹² *MC1R* has been also shown to be involved in photoprotection and **antioxidation**.¹³ Liu et al. showed that homozygous for the composite genotype looked 2 years older than non-homozygous.¹¹

Human assessment of age is not always quick or cheap to implement; for example, in one study, 288 assessors were used over more than a year to generate 75,096 assessments of age.¹¹ Such a protocol, even if one of the more intensive of such studies reported to date, make perceived facial age a difficult phenotype to study and implement into large cohorts. In addition, the exact features driving any associations are also difficult to determine due to the subjective nature of perceived age human assessments.

Deep learning is the term applied in recent years to the reinvention of the concept of artificial neural networks, a computational paradigm for learning from data inspired by the structure of the brain. An artificial neural network is typically formed by a large number of computed ‘neurons’ organized in layers. An input layer takes pixel data from a photographic image, transforms it and passes it to an intermediate layer which in turn generates some continuous

Key points

Why was the study undertaken?

We aim to assess whether deep learning methods can be used to automate the estimation of perceived age (PA) using facial photographs. PA has been shown to be associated with age-related outcomes in previous studies.

What does this study add?

We used state-of-the-art machine learning approaches to calculate PA and tested for associations with outcomes that were previously associated with human annotated PA as a proof of principle. We found that DL approaches can estimate PA accurately, with a mean average error (MAE) of predicted PA between 2.4 and 2.8 years of accuracy, which was lower than the best current published DL model estimations of PA.

What are the implications of this study for disease understanding and/or clinical care?

This represents a major step forward in the use of PA as a marker for intrinsic ageing for clinical and epidemiological studies as it will reduce both costs and bias in the assessment of PA without loss in the accuracy of predictions.

number/vector in the output layer. This process is continuously repeated on the output layers to extract increasingly higher order concepts from the raw data, from pixels to edges, from edges to basic shapes (e.g. wrinkles) and from basic shapes to complex shapes (e.g. a nose). A supervised neural network is then trained on the data from these layers to produce an output resembling as much as possible a pre-defined target measure (e.g. age).

The use of deep learning to analysis photographic images has led to considerable progress in estimating the age of subjects

in photographs, achieving estimates close to 1.96 years¹⁴ of actual/chronological age (CA). For perceived age, a 2015 competition to judge the perceived age of 4699 subjects found that this could be achieved to within 3.2 years¹⁵ of human PA. In a study of how reproducible human-estimated ages were between groups of age assessors, the mean ages taken from groups of five assessors demonstrated that the estimated ages tended to be within 4 years of each other and that decreased to around 2 years for groups of 20 assessors.¹⁶ Hence, deep learning techniques can achieve similar age estimations from the averages of at least five human assessors. Further improvements would enable studies similar to many current human perceived age studies to be performed with speed and accuracy due the automated nature of running deep learning models.

To determine whether deep learning could identify similar biological associations as human perceived age, we tested two deep learning perceived age models, one that we substantially optimized on facial images and trained on human perceived age data¹¹ using a population cohort study and one that repurposes features learned using a cutting edge face recognition method.¹⁷ In addition, we tested whether the DL-estimated PAs associated with the same conditions as previously reported⁷ for which human PA estimates were available. Moreover, as a proof of principle of their performance on images unseen in any training dataset, we tested for association between the new DL PA outcomes and a combined genetic score of four *MC1R* non-synonymous single-nucleotide polymorphisms (SNPs) that has previously been proven to associate with human PA.¹¹

METHODS

Study population

This is a cross-sectional study from an ongoing population-based study from Rotterdam, The Netherlands (The Rotterdam Study, RS) that started in 1990. The RS has previously been approved by the institutional review board (medical ethics committee) of the Erasmus Medical Centre and by the review board of the Netherlands Ministry of Health, Welfare and Sports. Extensive details and objectives have been described elsewhere.¹⁸ Since 2010, the participants have been invited to participate in dermatological examinations that include taking 3D facial photographs (camera (3dMD, Atlanta, GA, USA)) under standard conditions (see ref. [19] for details of the procedure). In brief, high-resolution standardized full-face photographs of participants (asked not to wear any make-up) were obtained with a Premier 3dMD face3-plus UHD camera taken in a room without daylight. Participants focused on a standardized viewpoint. Three two-dimensional (2D) photographs were taken simultaneously from three prefixed angles (one upper frontal and two 45° lateral photos) and combined into a 3D image of the whole face using 3dMD software (www.3dmd.com). The machine was calibrated daily to control for camera position and environmental light intensity.

For this study we defined two main PA datasets for testing PA associations, namely the training–test dataset and validation dataset (Figure 1a). The training–test dataset ($n=2679$) was previously used in our publications on human-estimated PA associations.^{7,11} We used this dataset to train and test our deep learning methods against the human-estimated PA using a cross-validation method. We then used a separate validation image dataset consisting of 1158 participants with facial images and genome-wide SNP data, but no human-estimated PAs or morbidity data (Figure 1a). We used this validation cohort to determine if the previously reported associations with DNA sequence variants (namely single-nucleotide polymorphisms, SNPs) in the *MC1R* gene still hold in this dataset unseen during the cross-validation of the DL models.

Human perceived age estimation in the Rotterdam study

Participants were asked to avoid wearing make-up or cosmetic products during the photoshoot, as well as maintaining a neutral facial expression. Facial photographs were captured by the 3dMD photography system and a 3D facial image was created from the photographs for each participant followed by frontal and side image rendering and export using Blender²⁰ to match poses and control for lighting variation across the study. The front and side images were cropped around the face to remove scalp hair and clothing cues in the images, then presented side by side for each subject to age assessors ($n=27$ age assessments on average) without the knowledge of chronological age and as previously described.¹⁶ Human PAs were calculated as the best linear unbiased predictor of the mean estimated age of the person in the image from a Proc-mixed model (SAS 9.3; SAS, Cary, NC, USA) with the viewing order as a fixed effect and assessor and image as random effects.

Deep learning algorithms

Images were first pre-processed before the implementation of DL (Appendix S1). We selected two DL approaches,^{21,22} namely SSL and DFT, designed to extract age information from relatively small image datasets. Both approaches are initially trained for facial recognition before being optimized for perceived age estimation (Figure 1b). This face recognition ‘pre-training’ makes use of the filtered version of an extremely large image dataset, presented in,²³ which contains around 5 million facial images, allowing the neural networks to learn how to extract facial features across a very large population. The accuracy of the resulting facial recognition was tested in a separate dataset.²⁴ Both DL approaches used the RS participant images with human-estimated PA for training and testing in a cross-validation approach; see Appendix S1 for a more detailed description of the training processes associated with each DL approach.

Following the pre-training, a semi-supervised learning (SSL) approach was used to train a chronological age

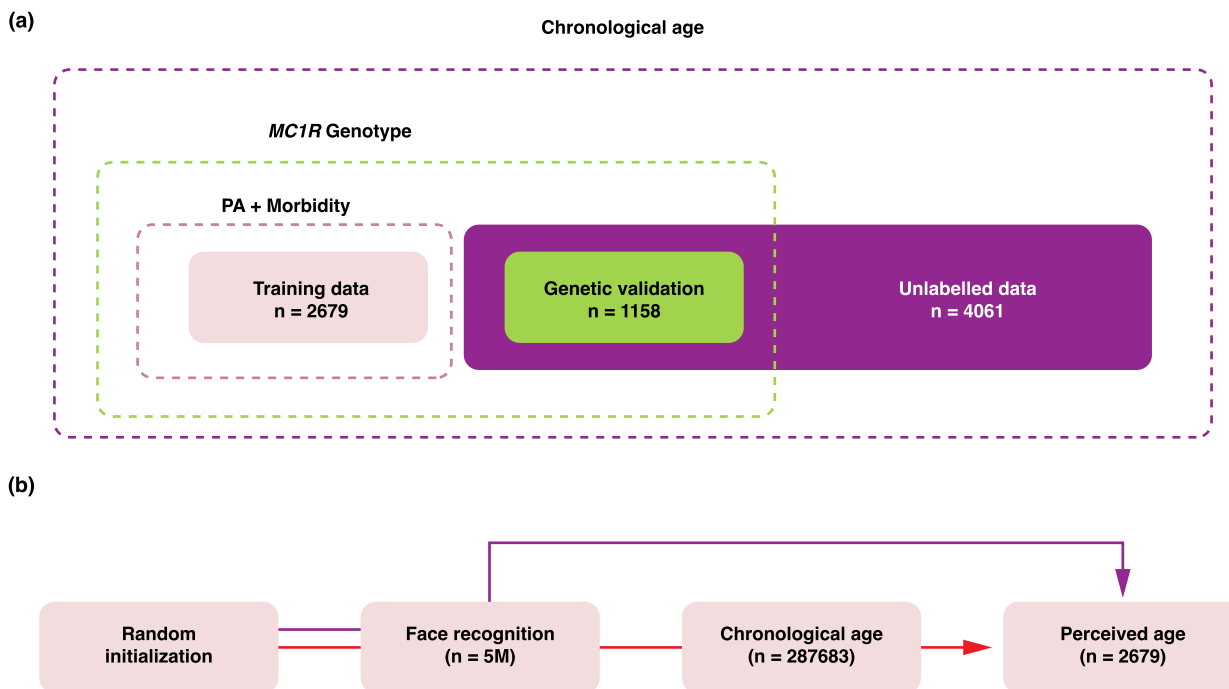


FIGURE 1 Dataset from the Rotterdam Study for the training and validation of DL algorithms. (a) Datasets used for training and testing the two DL approaches and an additional dataset for genetic validation associations with SNPs in the *MC1R* gene. Figure (b) presents represents the dataset used for the two DL methods with the Rrdom initiation, face recognition and chronological age being external datasets.

estimator using the largest publicly available chronological age (CA) dataset²⁵ which was filtered down (to $n = 287,683$) by removing images with noisy age labels.²⁶ This model was further trained on 'spare' RS images, and their corresponding chronological ages as these images were not available for any test or validation procedures due to their lack of human PA, genetic or morbidity data. The resulting SSL CA model was already able to predict perceived age with a mean absolute error of 3.9 years without having been trained on perceived age annotations, most likely due to the correlation between CA and PA. Finally, this model is trained on PA labels from RS images, as detailed below.

The second approach (deep feature transfer, DFT) does not modify the neural network at all during the perceived age training stage, but instead fits a Bayesian regression model to the continuous value output from each artificial neuron. The DFT was tested on both higher definition (640 by 640 pixels) and low-definition images (112 by 112 pixels) from the same subjects to test the importance of finer facial features in biological associations (Appendix S1 for details on DFT).

Statistical analysis

To validate the new DL predicted PAs, we calculated associations between previously associated human PA-associated morbidities and DL PA using the same methods as in Mekic et al.⁷ in the training-test dataset. In brief,

logistic regressions were used for associations between osteoporosis, chronic pulmonary disease (COPD), cataracts, cardiovascular disease (CVD), age-related macular disease (AMD) and human PA. The associations were adjusted for sex and age (Model 1) and additionally by BMI, UV exposure, smoking status and number of pack-years smoked (Model 2). For continuous outcomes (age-related hearing loss (ARHL) and cognition), multivariable linear regressions were used using the same two models as above.

To validate how accurately the DL PAs predict outcomes in a new dataset (the validation dataset) without human-estimated PA, we carried out an association analysis between the DL PAs and genetic variants from the *MC1R* gene, which was previously associated with human-estimated PA in the training-test dataset.¹¹ Briefly, the compound marker was a combination of four missense SNPs from the *MC1R* gene, namely rs1805005, rs1805007, rs1805008 and rs1805009 analysed in the same manner as previous.¹¹ We also removed first- and second-degree relatives ($n = 43$) to reduce bias in familial relatedness within this cohort.

RESULTS

Population characteristics

Our study population had a median age of 65.8 years, skewed to higher ages and consisted of a slightly higher percentage of women (54.1%) than men (45.9%) (Table S1). The median

chronological age of the participants was lower in the validation set, the proportion of females was higher, while the proportion of current smokers was lower (12% for validation and 31% for training–test dataset) (Table S1).

Morbidity associations

Figure 2 presents the average faces of seven women that looked older than their chronological age (Figure 2a; chronological ages: 60–65 years old) and seven women that looked younger than their chronological age (Figure 2b; similar chronological age). The most noticeable feature of the average faces is that women that looked younger had less wrinkles than the women that looked older of similar chronological age. Figure 2 also shows both mean DL estimates were similar to this from the human annotation. In fact, after PA training, the mean average error (MAE) of predicted PA for the test images in to the cross-validation model was 2.84 years for the DFT approach and 2.39 years for the SSL approach on average (Table S2).

We compared the test DL-derived PAs with morbidities that had previously been found to associate with human PA in the same cohort (Table 1). Consistent with the human annotation of PA, we found significant ($p < 0.05$) associations between both DL PAs and osteoporosis, COPD, cataracts, hearing loss and cognition in statistical model 1, although for human PA estimations, cataracts were outside significance in this set of subjects ($p = 0.09$). With further statistical adjustment for known risk factors (Model 2), both DL-estimated PAs were still significantly associated with COPD, cataracts, hearing loss and cognition, while osteoporosis lost significance ($p = 0.24$ for SSL PAs and $p = 0.20$ for DFT PAs). However, when the DFT PAs were estimated

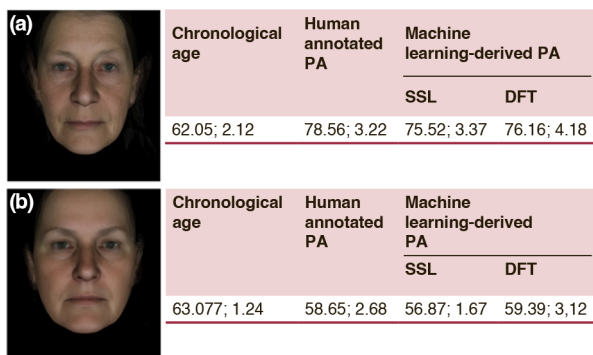


FIGURE 2 Average 2D facial photos of women looking older (a) and younger (b) than their chronological age. The figure presents the 2D average face of seven women that looked older than the chronological age (years) and seven women that looked younger than their chronological age. Below, the mean age and standard deviation of the ages are presented. With columns showing the mean chronological age (and standard deviation), as well as their mean perceive age estimated our previous study and as estimated using DL as described in this manuscript. SSL refers to self-supervised learning and DFT refers to deep feature transfer. The average faces were derived using publicly available software (details presented in the Appendix S1).

from higher quality images, a more significant association in Model 2 with osteoporosis ($p = 0.05$) resulted.

Genetic analysis

On testing the association of the *MC1R* genetic compound marker with the two different DL generated PAs in the training–test dataset, we found very similar significance and effect sizes of association as with human-estimated PAs, Table S3; for every minor allele of the compound *MC1R* marker, subjects looked 0.71 (DFT) and 0.78 (SSL) years older compared to 0.79 years older for human PAs. In the validation dataset, the direction of association and effect sizes were similar (0.69 years older per minor allele for DFT PAs and 0.71 years older per minor allele for SSL PAs) to the training–test results and both DL-estimated PAs were significantly ($p < 0.05$) associated.

DISCUSSION

In this work we showed that DL estimations of human PA give very similar associations with several age-related morbidities that have previously been shown to associate with human-estimated perceived age.⁷ Of the two methods analysed, we found that the DFT approach had slightly greater significant associations with morbidity than the SSL approach. In addition, both approaches generated PAs that associated with *MC1R* gene variants similarly to human PA association. Finally, in a dataset unseen during the training, both sets of DL-estimated PAs retained significant associations with the *MCR1* gene variants.

DL predictions of age offer a new and readily available way to generate PA from facial images. Initially, training costs associated with the deep neural network can be considerable, requiring a modest Linux server with GPU computing capabilities. Once the system is trained however, prediction of perceived ages can be run on very lightweight devices such as a basic office desktop or modern smartphone for little cost. Hence, the cheapness of this approach, once trained, should facilitate further research on perceived age that have previously been seen as cost prohibitive. For example, morbidity analysis and GWAS studies such as in Liu et al.¹¹ can be executed in exponentially larger populations, without the requirement for additional ratings of human PA. Such a system could also be trained to predict PA from smartphone or selfie images, unlocking the potential for entirely non-contact population studies. In addition, using explainable artificial intelligence (AI) approaches²⁷ to identify the facial features predominately influencing PA estimations should help determine which facial features and underlying ageing phenomena (e.g. cell senescence²⁸) are driving links between facial ageing and morbidity. Further examination of whether DL PA is able to predict risk of morbidities rather than presence, such as that found with human PA and mortality,⁵ would help determine the utility of such DL PA approaches to healthcare.

TABLE 1 Morbidity associations for human and DL PAs in the training–test dataset. Table presents the association results between different outcomes and PA as estimated using different methods, namely human annotation (a), deep feature transfer (DFT) (b) and semi-supervised learning (SSL) (c).

(A) Human					
Outcome	N	Model 1		Model 2	
		Odds ratio	p-Value	Odds ratio	p-Value
Osteoporosis	2197	0.651 (0.527 to 0.803)	0.00006	0.77 (0.622 to 0.96)	0.02252
COPD	2282	0.671 (0.603 to 0.746)	0.00000	0.76 (0.678 to 0.847)	0.00000
Cataracts	2185	0.888 (0.775 to 1.018)	0.08903	0.88 (0.768 to 1.019)	0.08923
CVD	2271	0.933 (0.798 to 1.092)	0.38851	0.93 (0.788 to 1.089)	0.35157
AMD	2282	0.925 (0.851 to 1.005)	0.06648	0.94 (0.8634 to 1.027)	0.17183
		B coefficient	p-Value	B coefficient	p-Value
ARHL	1945	−0.917 (−1.521 to −0.311)	0.00300	−0.87 (−1.495 to −0.237)	0.00699
Cognition	1864	0.07 (0.04 to 0.1)	0.00001	0.06 (0.03 to 0.097)	0.00011
(B) DFT					
Outcome	N	Model 1		Model 2	
		Odds ratio	p-Value	Odds ratio	p-Value
Osteoporosis	2197	0.638 (0.495 to 0.823)	0.00052	0.762 (0.582 to 0.998)	0.04788
COPD	2282	0.638 (0.560 to 0.727)	0.00000	0.729 (0.637 to 0.837)	0.00001
Cataracts	2185	0.844 (0.716 to 0.994)	0.04230	0.837 (0.7064 to 0.991)	0.03884
CVD	2271	1.019 (0.874 to 1.188)	0.80571	1.0167 (0.869 to 1.189)	0.83614
AMD	2282	1.016 (0.926 to 1.114)	0.73978	1.040 (0.944 to 1.145)	0.42758
		B coefficient	p-Value	B coefficient	p-Value
ARHL	1945	−0.972 (−1.693 to −0.251)	0.00824	−0.936 (−1.678 to −0.194)	0.01345
Cognition	1864	0.083 (0.046 to 0.121)	0.00001	0.0754 (0.037 to 0.114)	0.00012
(C) SSL					
Model	N	Model 1		Model 2	
		Odds ratio	p-Value	Odds ratio	p-Value
Osteoporosis	2197	0.709 (0.558 to 0.901)	0.00481	0.861 (0.671 to 1.106)	0.24121
COPD	2282	0.658 (0.583 to 0.741)	0.00000	0.748 (0.660 to 0.849)	0.00001
Cataracts	2185	0.832 (0.713 to 0.972)	0.02054	0.825 (0.703 to 0.967)	0.01793
CVD	2271	0.992 (0.829 to 1.187)	0.92914	0.988 (0.821 to 1.189)	0.89836
AMD	2282	0.970 (0.883 to 1.067)	0.53299	0.994 (0.901 to 1.096)	0.90455
		B coefficient	p-Value	B coefficient	p-Value
ARHL	1945	−1.020 (−1.701 to −0.339)	0.00334	−0.98809 (−1.695 to −0.282)	0.00612
Cognition	1864	0.0791 (0.043 to 0.115)	0.00001	0.068 (0.031 to 0.101)	0.00028

Note: Odds ratios and betas are presented per 5 years looking younger. Model 1: chronological age and sex. Model 2: ultraviolet exposure, BMI, smoking status and no pack years. For details on how the variables were defined please refer to Mekic et al. Sample size of our datasets may differ slightly with thus from Mekic et al. because not all the features from the photos could be automatically be retrieved in our study.

Abbreviations: AMD, age-related macular degeneration; ARHL, age-related hearing loss; COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease.

To support the transition to DL estimations of PA for facial ageing research purposes, we found very similar associations between the two DL approaches and morbidities as well as with *MC1R* gene variants as for human-estimated PAs associations. As the associations found here are independent (residual confounding aside²⁹) of already known morbidity risk factors such as chronological age, sex, BMI and smoking, the findings highlight that DL PAs give

additional information beyond these already utilized and readily available risk factors. Of the two DL approaches, the DFT approach can process relatively high-resolution images, ranging up to 640 pixels square when compared to the mere 112 pixels square accepted by the SSL model. At a resolution as small as 112 pixels, any training will miss granular features such as pigmentation, fine wrinkles and minor sagging. Interestingly, of all methods we evaluated, only the

high-resolution DFT approach was significantly associated with osteoporosis, indicating that this condition may be associated only with finer-grained facial features; further investigations into the exact features driving associations with morbidities, using explainable AI methods,³⁰ would help in fine-tuning such models (such as training on wrinkling data) and, perhaps, enable even stronger associations between facial ageing and morbidities could be discovered.

The results here are the most accurate for PA predictions of any model to date, and particularly fill a gap in older age-range predictions, which are under-represented in state-of-the-art DL PA models. Other reported datasets with PA labels do not have accompanying genetic or morbidity data, limiting our ability to validate our findings in other datasets and settings.¹⁵ Hence, the limitations to our findings include the lack of validation for PA associations with morbidities in other populations such as for different ethnicities. In addition, images shot with the 3DMD camera and pre-processed with semantic segmentation contain a unique set of computer vision features when compared to images shot with standard cameras (e.g. mobile phones). For this reason, the replication of our findings in other photography settings is required to determine the utility of the DL approaches beyond the setting described here, although further training in additional datasets should improve the generalizability of these approaches.

CONCLUSION

DL effectively predicts PA from facial images in a middle-aged and elderly cohort. Both DL PA-derived methods associated with COPD, ARHL, cataracts and osteoporosis replicate previous associations. These findings suggest that facial ageing features captured by both DL approaches may serve for researching facial ageing and morbidity and even determine whether they have any value for predicting individual health risks in a cost-effective manner.

AUTHOR CONTRIBUTIONS

Conceptualization: JB, CT, DG, LP and RZ; Data curation: CT, JB, SM, FL, MAI, CCWK, PHC, AG, KT, FR, MK, GGOB, MF and TN, Formal analysis: CT, JB and LP; Funding acquisition: DG and TN; Methodology: CT, JB, DG and RZ; Project administration: TN and LP; Resources: TN; Software: CT, JB and LP; Supervision: JB and DG; Writing: CT, DG, LP and JB; Review and editing: All authors.

ACKNOWLEDGEMENTS

We would like to thank the study participants of the Rotterdam Study, the staff and the participating general practitioners and pharmacists. We also would like to thank Xianjing Liu for his help in generating the average faces from [Figure 2](#).

FUNDING INFORMATION

This study is funded by Unilever. Authors J.F.B. and K.J.Z. were supported by Unilever. D.A.G. is an employee of Unilever. The

Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam.

CONFLICT OF INTEREST STATEMENT

CT, SM, FL, MAI, CCWK, PHC, AG, K.T, FR, MK, GGOB, MK, TN and JB declared no conflict of interest. DG and RZ are Unilever employers. No products produced by Unilever were tested in this study. However, it is possible that this study could be used to promote products and services in the future, leading to financial gain. L.M.P. has received consulting fees from Centogene.

DATA AVAILABILITY STATEMENT

Data can be obtained upon request. Requests should be directed towards the management team of the Rotterdam Study (datamanagement.ergo@erasmusmc.nl), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository.

ETHICAL APPROVAL


The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, licence number 1071272-159521-PG).

ETHICS STATEMENT

All participants provided written informed consent to participate in the study and have their information obtained from treating physicians.

ORCID

Luba M. Pardo  <https://orcid.org/0000-0003-0684-3175>

Katerina Trajanoska  <https://orcid.org/0000-0002-3792-4296>

REFERENCES

1. Gunn DA, Rexbye H, Griffiths CEM, Murray PG, Fereday A, Catt SD, et al. Why some women look young for their age. *PLoS One*. 2009;4:e8021.
2. Victorelli S, Lagnado A, Halim J, Moore W, Talbot D, Barrett K, et al. Senescent human melanocytes drive skin ageing via paracrine telomere dysfunction. *EMBO J*. 2019;38:e101982.
3. Christensen K, Thinggaard M, McGue M, Rexbye H, JvB H, Aviv A, et al. Perceived age as clinically useful biomarker of ageing: cohort study. *BMJ*. 2009;339:b5262.
4. Dykiert D, Bates TC, Gow AJ, Penke L, Starr JM, Deary IJ. Predicting mortality from human faces. *Psychosom Med*. 2012;74:560–6.
5. Gunn DA, Larsen LA, Lall JS, Rexbye H, Christensen K. Mortality is written on the face. *J Gerontol A Biol Sci Med Sci*. 2016;71:72–7.
6. Nielsen BR, Linneberg A, Christensen K, Schwarz P. Perceived age is associated with bone status in women aged 25–93 years. *Age (Dordr)*. 2015;37:106.

7. Mekić S, Pardo LM, Gunn DA, Jacobs LC, Hamer MA, Ikram MA, et al. Younger facial looks are associated with a lower likelihood of several age-related morbidities in the middle-aged to elderly. *Br J Dermatol*. 2023;188:390–5.
8. Gunn DA, de Craen AJM, Dick JL, Tomlin CC, van Heemst D, Catt SD, et al. Facial appearance reflects human familial longevity and cardiovascular disease risk in healthy individuals. *J Gerontol Ser A*. 2013;68:145–52.
9. Kido M, Kohara K, Miyawaki S, Tabara Y, Igase M, Miki T. Perceived age of facial features is a significant diagnosis criterion for age-related carotid atherosclerosis in Japanese subjects: J-SHIPP study. *Geriatr Gerontol Int*. 2012;12:733–40.
10. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. *Proc Natl Acad Sci USA*. 2015;112:E4104–E4110.
11. Liu F, Hamer Merel A, Deelen J, Lall Japal S, Jacobs L, van Heemst D, et al. The MC1R gene and youthful looks. *Curr Biol*. 2016;26:1213–20.
12. Guida S, Guida G, Goding CR. MC1R functions, expression, and implications for targeted therapy. *J Invest Dermatol*. 2022;142:293–302. e291.
13. Abdel-Malek ZA, Swope VB, Starner RJ, Koikov L, Cassidy P, Leachman S. Melanocortins and the melanocortin 1 receptor, moving translationally towards melanoma prevention. *Arch Biochem Biophys*. 2014;563:4–12.
14. Gao B-B, Liu X-X, Zhou H-Y, Wu J, Geng X. Learning expectation of label distribution for facial age and attractiveness estimation. *Pattern recognition*. 2021.
15. Rothe R, Timofte R, Van Gool L. Deep Expectation of real and apparent age from a single image without facial landmarks. *Int J Comp Vision*. 2018;126:144–57.
16. Gunn DA, Murray PG, Tomlin CC, Rexbye H, Christensen K, Mayes AE. Perceived age as a biomarker of ageing: a clinical methodology. *Biogerontology*. 2008;9:357–64.
17. Deng J, Guo J, Yang J, Xue N, Kotsia I, Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition. *IEEE Trans Pattern Anal Mach Intell*. 2021;44:1.
18. Ikram MA, Kieboom BCT, Brouwer WP, Brusselle G, Chaker L, Ghanbari M, et al. Design update and major findings between 2020 and 2024. *Eur J Epidemiol*. 2024;39:183–206.
19. Hamer MA, Jacobs LC, Lall JS, Wollstein A, Hollestein LM, Rae AR, et al. Validation of image analysis techniques to measure skin aging features from facial photographs. *Skin Res Technol*. 2015;21:392–402.
20. Foundation B. Blender – a 3D modelling and rendering package. In: Community BO, editor. 2018.
21. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. 2015.
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *arXiv:1512.03385 [cs]*. 2015.
23. Deng J, Guo J, Yang J, Xue N, Kotsia I, Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition. *IEEE Trans Pattern Anal Mach Intell*. 2022;44:5962–79.
24. Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. *Workshop on faces in 'Real-Life' Images: Detection, Alignment, and Recognition*. 2008.
25. Rothe R, Timofte R, Gool LV. DEX: deep Expectation of apparent age from a single image. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW). IEEE, 252–257. 2015.
26. Lin Y, Shen J, Wang Y, Pantic M. FP-Age: leveraging face parsing attention for facial age estimation in the wild. 2022.
27. Gupta LK, Koundal D, Mongia S. Explainable methods for image-based deep learning: a review. *Arch Comput Methods Eng*. 2023;30:2651–66.
28. Waaaijer ME, Gunn DA, Adams PD, Pawlikowski JS, Griffiths CE, van Heemst D, et al. P16INK4a positive cells in human skin are indicative of local elastic fiber morphology, facial wrinkling, and perceived age. *J Gerontol A Biol Sci Med Sci*. 2016;71:1022–8.
29. Danziger J, Zimolzak AJ. Residual confounding lurking in big data: a source of error. *Secondary Analysis of Electronic Health Records*. Cham: Springer; 2016. p. 71–8.
30. Holzinger A, Saranti A, Molnar C, Biecek P, Samek W. Explainable AI methods – a brief overview. In: Holzinger A, Goebel R, Fong R, Moon T, Müller K-R, Samek W, editors. *xxAI – beyond explainable AI: international workshop, held in conjunction with ICML 2020, July 18, 2020. Revised and Extended Papers, Vienna, Austria*. Cham: Springer International Publishing; 2022. p. 13–38.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Turner C, Pardo LM, Gunn DA, Zillmer R, Mekić S, Liu F, et al. Deep learning predicted perceived age is a reliable approach for analysis of facial ageing: A proof of principle study. *J Eur Acad Dermatol Venereol*. 2024;00:1–8. <https://doi.org/10.1111/jdv.20365>