BrainAGE revisited: Keeping up the pace with deep learning methods

Author names and affiliations:

Polona Kalc[1], Robert Dahnke[1,2], Felix Hoffstaedter[3,4], Christian Gaser[1,2]

[1]Structural Brain Mapping Group, Department of Neurology, Jena University Hospital

[2]Department of Psychiatry and Psychotherapy, Jena University Hospital

[3]Forschungszentrum Jülich, Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7)

[4]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf

Corresponding author: Christian Gaser

Address: Am Klinikum 1, 07747 Jena - DE

Abstract

Since the introduction of the BrainAGE method (Franke et al., 2010), novel machine learning methods of brain age prediction have continued to emerge. The idea of estimating the chronological age from (different modalities of) magnetic resonance images proved to be an interesting field of research due to the simplicity of its interpretation and its potential use as a biomarker of brain health. With the advancement of deep learning methods that are starting to dominate the field, it may seem like standard machine learning methods are not competitive anymore.
Here we present our revised BrainAGE framework that combines multiple weighted BrainAGE models. We employed the analyses on a sample of 36840 T1-weighted images from the UK Biobank. The most accurate prediction with a mean absolute error (MAE) of 2.02 years was obtained by relevance vector regression weighted model combination. Our results are comparable to those of deep learning methods. We additionally tested the relation of different BrainAGE models and white matter hyperintensities as markers of ageing brain and found no differences in association to biomedical data between models of differing accuracies.

Highlights

- Standard machine learning (BrainAGE) algorithms can outperform deep learning methods.

- Combining machine learning models can help increase accuracy and robustness of predictions.
- Higher accuracy of BrainAGE prediction does not necessarily imply higher/lower association with biomedical data.

Key words

brain age, UK Biobank, pre-processing, machine learning, mean absolute error, structural MRI

Introduction

Our original BrainAGE framework (Franke et al., 2010) is a supervised machine learning method that employs relevance vector regression in order to predict the chronological age of a subject based on a single T1-weighted magnetic resonance (MR) scan of its brain. The difference between the apparent age of the subject's brain as predicted by the algorithm and its chronological age is the so-called brain age gap estimation (BrainAGE). A positive gap implies more prominent structural brain changes that commonly occur with ageing progression, e.g. decrease of grey matter volume (Bethlehem et al., 2022), whereas a negative difference signifies a slower ageing in comparison to other subjects within a sample. The BrainAGE method was introduced more than a decade ago and was successfully applied to various subject groups, such as adolescents (Franke et al., 2012), elderly people with mild cognitive impairment and Alzheimer's disease (AD; Franke et al., 2010; Franke & Gaser, 2012), musicians (Rogenmoser et al., 2018), meditators (Luders et al., 2016), and even other animal species, such as rodents and non-human primates (Franke et al., 2017). The method proved to be a reliable biomarker of AD (Franke & Gaser, 2012), Parkinson disease (Eickhoff et al. 2021), and is associated with brain alterations in schizophrenia (Nenadić et al., 2017) and type 2 diabetes (Franke et al., 2013). A complete review can be found in Franke & Gaser (2019) or Cole & Franke (2017).

Since the introduction of the concept of brain age as a biomarker of ageing brains, numerous machine learning and deep learning models of estimating the age of the brain based on different MR imaging modalities have emerged (e.g. Beheshti et al., 2020; Cole et al., 2018; Smith et al., 2019; Varzandian et al., 2021). With the advancement of deep learning methods in the field of brain age prediction (e.g. Abrol et al., 2021; Gong et al., 2020; Lombardi et al., 2021; Peng et al., 2021), their promise of utilising representation learning (Abrol et al., 2021) and transcending the limitations of preprocessing (Vieira et al., 2017), it may seem that the standard machine learning methods are no longer competitive. Novel deep learning models trained on samples from the UK Biobank (UKB; https://www.ukbiobank.ac.uk/), the largest multimodal imaging data collection in the world (Littlejohns et al., 2020), have managed to decrease mean absolute errors (MAE; i.e., a metric commonly used to compare the accuracy of prediction) to as low as 2.14 years (Peng et al., 2021). However, the sole purpose of MAE reduction has recently been questioned since "tightly-fitting" models may

overlook important biological features/brain pathologies while attempting to make the most accurate prediction of age possible (Baysham et al., 2021).

Current practices in machine learning competitions utilise model ensembling/combination to prevent overfitting and improve prediction accuracy without the intensive tuning of the parameters (van Veen et al., 2015). Inspired by these procedures, we enhanced our classical single- and concatenated tissue class BrainAGE estimation procedures (Franke et al., 2010; Franke & Gaser, 2012) by two new multiparameter models, while retaining the parsimonious and straightforward nature of voxel-based morphometry (VBM). The first model ensemble consists of different tissue segments with multiple parametrization of different smoothing kernels and spatial resolution as input to obtain a brain age prediction. The second combined single independent models by weighing them into a general BrainAGE score. We employed these BrainAGE algorithms to a large-cohort sample of the UK Biobank (UKB) – a biomedical database and research resource that contains genetic, lifestyle and health information from half a million UK Biobank participants – in order to obtain easily comparable results to those of other groups and to see whether a standard supervised machine learning algorithm can render results comparable to state-of-the-art deep learning models.

Method

Data

This research has been conducted using data from UK Biobank under the application number 41655. The UK Biobank has ethical approval from the North West Multi-Centre Research Ethics Committee (MREC) and is in possession of informed consents from the study cohort participants.

Participants with pre-existing neurological or psychiatric diagnoses were excluded from the analyses. Overall, we analysed 36840 T1-weighted MR scans (52% women; age range: 45–82; M_{Age} = 64.09, SD_{Age} = 7.53 years). To reduce the computational load in the model ensembling, we created a subsample of 7241 scans (52% women; age range: 45–82; M_{Age} = 64.08, SD_{Age} = 7.54 years), including only subject IDs starting with 1* to avoid the introduction of selection bias. For the cross-site prediction, we randomly sampled a fifth of the full sample (n = 7368) to obtain two size-matched subsamples from the scanning site 1 (Cheadle) and the scanning site 3 (Newcastle).

Data acquisition

All the information regarding data acquisition is available in the UK Biobank Imaging documentation (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf). In short, T1-weighted MR images were acquired on a Siemens Skyra 3T scanner with a 32-channel RF receive head coil. The MPRAGE sequence was used with 1-mm isotropic resolution,

inversion/repetition time = 880/2000 ms, acquisition time: 5 minutes, FOV: 208x256x256 matrix, in-plane acceleration factor = 2.

Data preprocessing

T1-weighted images were segmented into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) and affinely normalised by using the default preprocessing of the CAT12.7 Toolbox (Gaser et al., 2022) as standalone version compiled under Matlab 2019b and SPM12 (Wellcome Centre for Human Neuroimaging, https://www.fil.ion.ucl.ac.uk/spm/) massively parallelized on the JURECA High Performance Computing system (JSC). In order to keep the process as parsimonious as possible, only an affine registration was used (unless otherwise specified). The preprocessing utilises the unified segmentation (Ashburner & Friston, 2005) to remove B0-inhomogeneities and create an initial segmentation that is used for (local) intensity scaling and adaptive non-local means denoising (Manjon et al., 2010). An adaptive maximum a posteriori (AMAP; Rajapakse et al., 1997) segmentation with a hidden Markov random field (Cuadra et al., 2005) and partial volume effect model (Tohka et al., 2004) is utilised to create the final segmentation. For the optional non-linear registration, the Shooting method (Ashburner & Friston, 2011) with modulation was used.

BrainAGE models

We compared 4 classes of BrainAGE models defined by the complexity of their input and combination: (a) single tissue class models using only one linear/non-linear registered tissue class, (b) tissue concatenation models of multiple tissue classes, (c) ensembled models of multiple tissues and varying parameterisations of spatial resolution (4 mm / 8 mm) and Gaussian smoothing (FWHM: 4 mm / 8 mm), and finally (d) a novel combination of single independent models joined in a general BrainAGE estimation (see Figure 1).

In order to obtain a general BrainAGE score, four different combinations were tested, namely (i) selecting the single model with the lowest MAE, (ii) median of all the estimated models, (iii) weighted mean obtained by running an additional relevance vector regression to minimise the MAE, and (iv) weighted mean calculated by employing a general linear model that minimises MAE.

All models were defined and processed in the same way. Singular value decomposition was applied to the training data only and the resulting transformation was applied to the test sample. We employed a relevance vector regression (RVR) with a radial basis function kernel (sigma = 5), as it provided the most stable results in the subsample (see the Supplement). Ten-fold cross-validation was performed on the subsample ($n_{training}$= 6517 scans) and a 5-fold cross-validation procedure was employed on the full data-set ($n_{training}$ = 29472). The estimated brain age was age-bias adjusted by a quadratic term as described in Smith et al. (2019). The estimations were run in Matlab 2021a (Mathworks Inc., Natick, MA, USA).

Finally, we tested the cross-site prediction by training the algorithm on a sample of 6631 T1-weighted MRI scans (again using 90% of the sample to keep the size of training sample constant) from the scanning site 1 (Cheadle) and we tested its performance on a dataset from the scanning site 3 (Newcastle), and vice versa.

Association of BrainAGE and biomedical data

In addition, we were interested in whether minimising the MAE influences the explained variance of biomedical data connected to (brain) ageing and noncommunicable diseases. To test this, we evaluated the correlation of BrainAGE scores and three UKB features. We have chosen global volume of white matter hyperintensities (from T1 and T2 FLAIR images) (UKB data field #25781-2.0, normalised for GM and WM volume), which had previously been shown to increase with brain ageing (de Leeuw et al., 2001; Garnier-Croussard et al., 2020). The log-transformation was applied to render the data more normally distributed. Additionally, associations to diastolic- and systolic blood pressure (UKB data fields #4079-2.0 and #4080-2.0, respectively) were tested as these measures have been previously correlated to brain age estimates (Cole, 2020). The analyses were conducted in R Statistical Software (v4.1.1, R Core Team, 2022).

Results

Model estimation and combination

The results of estimated models and their combination are presented in Figure 1. We report the MAEs and the MAEs normalised for the age-range (MAE/age range) as proposed by Cole et al. (2019).

First, we investigated the performance of the models trained on various single brain tissue segments and different combinations thereof on a smaller test sample (n = 7241). Single tissue models provide the least accurate prediction. Interestingly, the estimation obtained on affinely registered grey matter is comparable to the estimation from the non-linearly registered grey matter model, with the first slightly outperforming the latter. Model ensembling provides a further improvement of BrainAGE estimation accuracy, with ensembles coupling all tissue classes achieving the most accurate estimations. The most striking improvement, however, presents model combination by RVR-obtained weights. The MAE of BrainAGE estimations of the full sample is as low as 2.02 years.

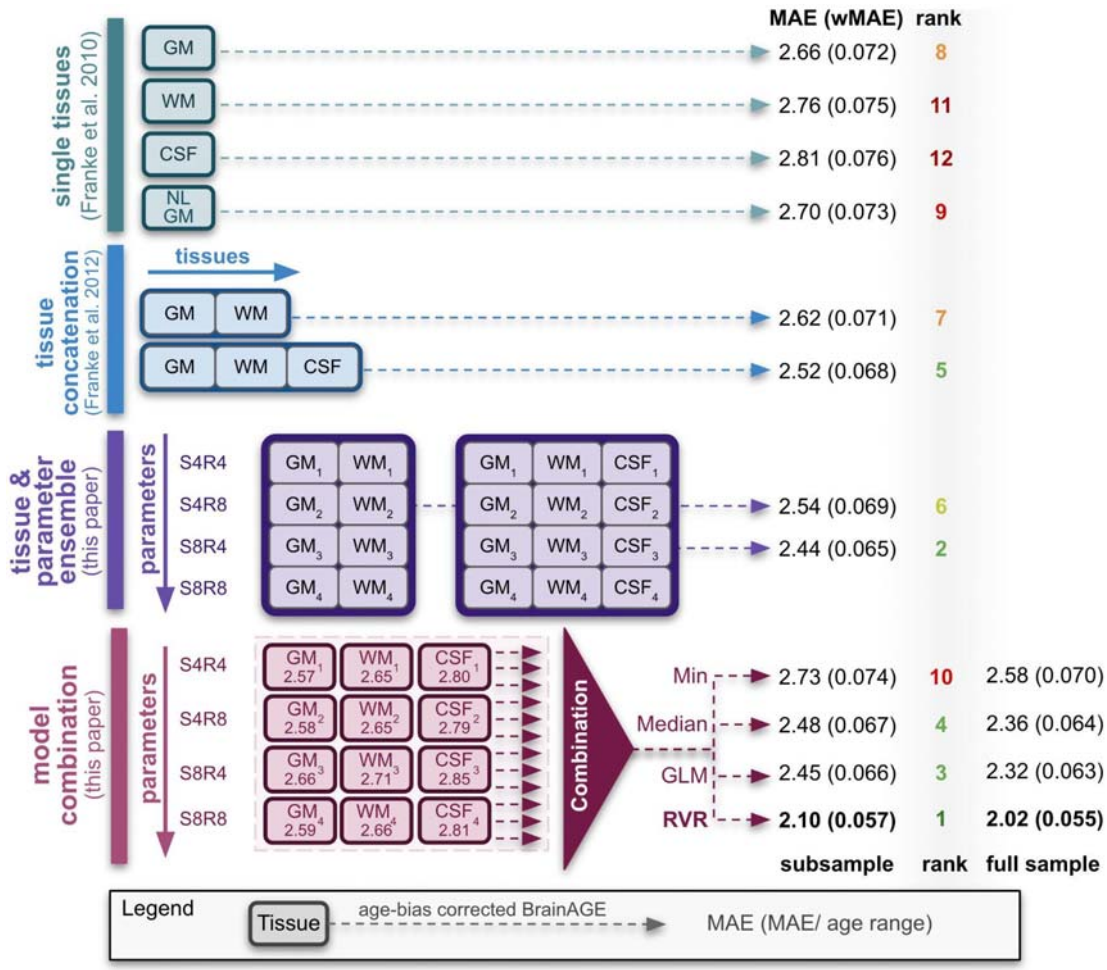

Figure 1

Different tissue segments (GM, WM, CSF, non-linearly modulated GM) as well as the concatenation of their vectors and their ensembles were tested against each other. In the combination step, twelve models differing in spatial resolution (R: 4 mm / 8 mm), smoothing kernel (S: 4 mm / 8 mm), and tissue segment (GM, WM, CSF) were estimated and combined using weights obtained by either GLM or RVR. The results of model estimations are presented as MAE and normalised MAE (in brackets) for analyses run on a subsample and a full sample. (Abbreviations: NL GM: non-linearly modulated GM, S: smoothing kernel, R: resolution, GLM: general linear model, RVR: relevance vector regression).

Relation between BrainAGE and biomedical data

The associations of the (weighted) BrainAGE models estimated on a full sample with three features related to brain ageing and disease, namely global volume of white matter hyperintensities (from T1 and T2 FLAIR images, r_M= 0.22), systolic- (r_M = 0.05), and diastolic blood pressure (r_M = 0.06) are presented in Figure 2. The lowest correlation coefficients were observed for a single model with the highest MAE (i.e., the CSF model, smoothing kernel: 8

mm, resolution: 4 mm), however the best performing single model (GM, smoothing: 4 mm, resolution: 4 mm) did not differ much from the combined models. Differences in correlation coefficients to biomedical data between different combined BrainAGE models were negligible.

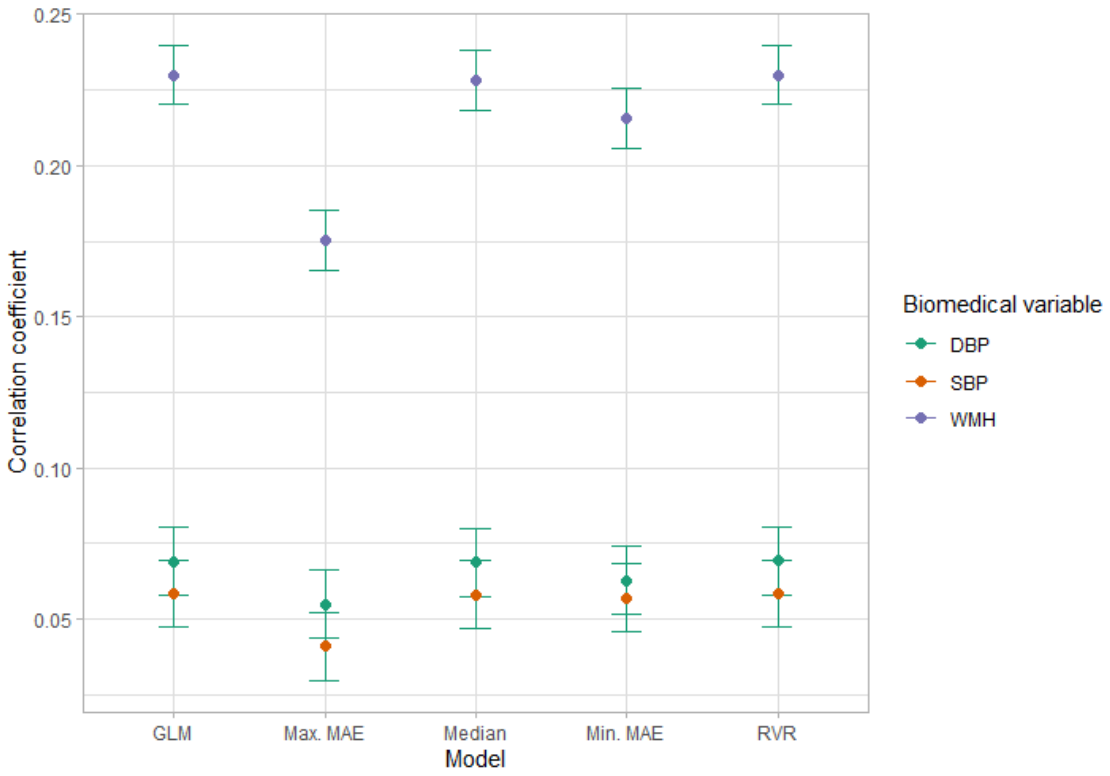

Figure 2

Associations (Pearson's r and their 95% CI) between BrainAGE scores obtained by different models and biomedical data (DBP: diastolic blood pressure, SBP: systolic blood pressure, WMH: white matter hyperintensities).

Cross-site prediction

The reliability of BrainAGE estimation was tested with a cross-site prediction. We randomly sampled a fifth of the full sample (*n* = 7368) to obtain two size-matched subsamples from the scanning site 1 (Cheadle) and the scanning site 3 (Newcastle). The algorithm was trained on a sample of 6631 T1-weighted MRI scans from the scanning site 1 and we tested its performance on a dataset from the scanning site 3, and vice versa. When predicting from the scanning site 1 to site 3, we obtained a MAE of 2.47 years, with a similar accuracy of MAE = 2.52 years when doing the reverse estimation.

Discussion

In our present work, we revised the BrainAGE model that was originally introduced more than a decade ago (Franke et al., 2010). We first explored the contribution of various brain tissue segments to BrainAGE prediction accuracy, followed by different combinations of these models in order to improve prediction accuracy. By employing the algorithm on a large population based cohort sample of the UK Biobank, we showed that a standard supervised machine learning algorithm does not only keep up with state-of-the-art deep learning algorithms but can surpass them in accurately estimating brain age without compromising its relation to relevant biomedical variables. We discuss the advantages and limitations of our new combined BrainAGE framework compared to other approaches in the following sections.

Deep Learning vs. Machine Learning / original vs. preprocessed data

Although it is advised against directly comparing the performance measure of MAE between different studies, as the metric depends on the age range and other sample-specific characteristics (de Lange et al., 2021), we presume the results of our study can be placed in juxtaposition with other brain age studies that employed their algorithms on the UK Biobank sample. Our revised BrainAGE algorithm achieved a MAE of 2.10 years on the subsample ($n_{training}$ = 6517) and 2.02 years on the full sample ($n_{training}$ = 29472), hence surpassing the performance of the majority of deep learning algorithms.

The latter were originally introduced to the brain imaging field with a promise of bypassing the limitations of preprocessing and making use of detailed information hidden in the raw data (Vieira et al., 2017; Abrol et al., 2021). However, we argue that well-selected preprocessing steps can help stabilise brain age prediction by reducing protocol-based imaging properties (e.g. resolution, contrast, noise, inhomogeneities etc.), while keeping important structural aspects. In fact, despite deep learning algorithms' promise to omit the preprocessing steps, basic processing such as image registration, denoising, bias-correction, brain masking or intensity normalisation were used in the majority of reported deep learning brain age estimations (Abrol et al., 2021; Gong et al., 2020; 2021; Lombardi et al., 2021; Peng et al., 2021). Moreover, prediction accuracy of the CNN algorithm that Cole et al. (2017) employed on the raw data is worse compared to that of the preprocessed data, and Abrol et al. (2020; 2021) report that priorly smoothing the data at the individual level improves the performance of the deep learning algorithm.

Our results, together with the reports of researchers utilising some basic preprocessing steps prior to employing deep learning algorithms, indicate that these methods are not superior in brain age prediction to standard machine learning models. As we discuss in the next section, utilising approaches of successful machine learning competitors can improve prediction accuracy of standard machine learning methods.

Combined BrainAGE models

Based on our analyses, a combination of models trained on affine registered GM, WM and CSF segments jointly contribute to the best prediction. Although the GM segments seem to contribute the most relevant information to the algorithm – it is the GM volume that declines

most prominently with age after all (Bethlehem et al., 2022) – models with multiple tissue segments perform better. Interestingly, the use of non-linearly registered GM data – typically used in VBM studies – did not lead to a better performance in comparison to the model that utilised affinely registered GM. We speculate that linear registration provides more spatial information of the individual brain anatomy and its atrophy pattern, being more analogous to raw brain anatomy used for deep learning methods, whereas the segmentation (with its denoising and bias-correction steps) helps to remove protocol-based image properties (Ashburner & Friston, 2000). This supports our intention of keeping the BrainAGE model parsimonious.

An important factor that strongly improved the prediction accuracy of BrainAGE estimation was combining various single BrainAGE models. Model ensembling is an effective way of improving the brain age estimation as it serves as a regularisation step to prevent overfitting and thus improves prediction accuracy (van Veen et al., 2015). Even though we ensembled models that were overall quite similar, we believe this is not a strong drawback since even similar models still provide additional variation (van Veen et al., 2015) and our results show benefits without significantly increased computational costs. As can be seen from our results, employing an additional relevance vector regression to obtain the loadings to combine different models together substantially decreased the MAE.

Additional improvements of age prediction may be possible by ensembling more independent brain age models, for example by using different image modalities, morphometric metrics (e.g. thickness, sulcal depth), or other brain-age frameworks (e.g. Cole et al., 2018; Smith et al., 2019). Although improvements with adding modalities were shown before (e.g., Cole, 2020; Niu et al., 2020), we focused only on broadly available T1 image data and a standard VBM paradigm in order to keep our framework parsimonious. Operating within this paradigm renders BrainAGE less of a black box type of measure in comparison to deep learning models, potentially enabling explainable artificial intelligence approaches for BrainAGE measures (Boscolo Galazzo et al., 2022). Furthermore, omitting working with other imaging modalities or surface based measures, despite their probable superior performance that comes at the cost of higher complexity and computational efforts, enables the BrainAGE algorithm to be widely applicable (potentially) also in the clinical usage.

The optimisation of MAE within BrainAGE framework and the relation to biomedical variables

Even though the current state of affairs in training brain-age models is the continuation of decreasing the MAE, the purposefulness of such actions is under debate (e.g. Bashyam et al., 2021; Hahn et al., 2021). As Bashyam et al. (2020) pointed out, models with the lowest MAE or the so-called “tightly-fitting models” may overlook meaningful biological variability and potentially covert pathologies in (ageing) brain while attempting to make the most accurate prediction of age possible. They suggest moderately-fitting models (MAE of approx. 6 years as opposed to tightly-fitting 4 years) provide better solutions at least for clinical purposes. Although not directly tested in this study, when comparing performance of different model combinations and their associations with biomedical data, our results show negligible differences in correlation coefficient values. These are indicating that better prediction accuracy of the model does not necessarily lead to a greater/smaller proportion of

explained variance of biomedical data. Nevertheless, all our models could be regarded as more or less tightly-fitting models, as we did not intentionally induce a lesser-fit during training, and were employed on healthy subjects only, therefore, further studies and metrics are needed to explore this topic.

Cross-site prediction

Although our results show a comparable performance of the algorithm based on training and testing on samples from two different scanner sites, both scanning sites in the UK Biobank underwent a thorough calibration and quality control, following strict imaging protocols (Alfaro-Almagro et al. 2018), therefore limiting our insight into the generalizability of the BrainAGE measure to a closely controlled scanner setup. Nevertheless, BrainAGE's generalizability had already been proven in the initial study of Franke et al. (2010) where a similar performance of the algorithm trained and tested on two different datasets (namely ADNI and IXI) was demonstrated. Future studies could benefit from employing brain age algorithms on harmonised data from multiple imaging studies (Pomponio et al. 2020), thus allowing the application of brain age prediction to more diverse data. As Pomponio and colleagues (2020) have demonstrated, harmonisation allowed to further reduce the prediction error in comparison to MAE obtained on unharmonised data.

Conclusions

In the present study, we demonstrated that BrainAGE is a meaningful biomarker of ageing brain as confirmed by its association to white matter hyperintensities. Regardless of the prediction accuracy of the models, their relation to relevant biomedical variables was similar. We provide evidence for the utilisation of standard machine learning methods and model ensembling helps to increase the prediction accuracy of brain age estimation. Although the present study was not conducted (solely) to compete with deep learning algorithms, the results we obtained demonstrate that standard machine learning techniques can still keep up the pace and provide novel contributions to the research area.

Acknowledgements

This research has been conducted using data from UK Biobank under the application number 41655.

The work was supported by Carl Zeiss Stiftung as a part of the IMPULS project (IMPULS P2019-01-006), the Federal Ministry of Science and Education (BMBF) under the frame of ERA PerMed (Pattern-Cog ERAPERMED2021-127) and the Marie Skłodowska-Curie Innovative Training Network (SmartAge 859890 H2020-MSCA-ITN2019).

The authors gratefully acknowledge the computing time granted by the JARA Vergabegremium and provided on the JARA Partition part of the supercomputer JURECA at Forschungszentrum Jülich.

Data availability statement

The data used for this work was obtained from the UK Biobank Resource (Project Number 41655). Due to the nature of the data sharing agreement, we are not allowed to publish the data.

Conflict of interest

The authors declare no conflict of interest.

References

Abrol, A., Bhattarai, M., Fedorov, A., Du, Y., Plis, S., & Calhoun, V. (2020). Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease. *Journal of Neuroscience Methods*, *339*, 108701. 10.1016/j.jneumeth.2020.108701

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., & Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature Communications*, *12*(1), 353. https://doi.org/10.1038/s41467-020-20655-6

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernández-Fernández, M., Vallée, E., Vidaurre, D., Webster, M. A., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., Miller, K. L. & Smith, S. M. (2018). Image processing and Quality Control for the first 10, 000 brain imaging datasets from UK Biobank. *NeuroImage*, 166, 400-424.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. *NeuroImage, 11*, 805–821. doi:10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Ashburner, J., & Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage*, *55*(3), 954–967. https://doi.org/10.1016/j.neuroimage.2010.12.049

Bashyam, V. M., Erus, G., Doshi, J., Habes, M., Nasrallah, I. M., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Satterthwaite, T. D., … Davatzikos, C. (2020). MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain*, *143*(7), 2312–2324. https://doi.org/10.1093/brain/awaa160

Bashyam, V., Shou, H., & Davatzikos, C. (2021). Reply: From 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. *Brain*, *144*(3), e32–e32. https://doi.org/10.1093/brain/awaa455

Beheshti, I., Mishra, S., Sone, D., Khanna, P., & Matsuda, H. (2020). T1-weighted MRI-driven Brain Age Estimation in Alzheimer's Disease and Parkinson's Disease. *Aging and Disease*, *11*(3), 618. https://doi.org/10.14336/AD.2019.0617

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), 525–533. https://doi.org/10.1038/s41586-022-04554-y

Boscolo Galazzo, I., Cruciani, F., Brusini, L., Salih, A., Radeva, P., Storti, S. F., & Menegaz, G. (2022). Explainable Artificial Intelligence for Magnetic Resonance Imaging Aging Brainprints: Grounds and challenges. *IEEE Signal Processing Magazine*, *39*(2), 99–116. https://doi.org/10.1109/MSP.2021.3126573

Cole, J. H. (2020). Multimodality neuroimaging brain-age in UK Biobank: Relationship to biomedical, lifestyle, and cognitive factors. *Neurobiology of Aging*, *92*, 34–42. https://doi.org/10.1016/j.neurobiolaging.2020.03.014

Cole, J. H., & Franke, K. (2017). Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences*, *40*(12), 681–690. https://doi.org/10.1016/j.tins.2017.10.001

Cole, J. H., Franke, K., & Cherbuin, N. (2019). Quantification of the Biological Age of the Brain Using Neuroimaging. In A. Moskalev (Ed.), *Biomarkers of Human Aging* (Vol. 10, pp. 293–328). Springer International Publishing. https://doi.org/10.1007/978-3-030-24970-0_19

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124. https://doi.org/10.1016/j.neuroimage.2017.07.059

Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S. E., Zhang, Q., Wray, N. R., Redmond, P., Marioni, R. E., Starr, J. M., Cox, S. R., Wardlaw, J. M., Sharp, D. J., & Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry*, *23*(5), 1385–1392. https://doi.org/10.1038/mp.2017.62

Cuadra, M. B., Cammoun, L., Butz, T., Cuisenaire, O., & Thiran, J.-P. (2005). Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Transactions on Medical Imaging*, *24*(12), 1548–1565. https://doi.org/10.1109/TMI.2005.857652

de Lange, A.-M. G., Anatürk, M., Rokicki, J., Han, L. K. M., Franke, K., Alnæs, D., Ebmeier, K. P., Draganski, B., Kaufmann, T., Westlye, L. T., Hahn, T., & Cole, J. H. (2021). *Mind the gap: Performance metric evaluation in brain-age prediction* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.05.16.444349

de Leeuw, F.-E. (2001). Prevalence of cerebral white matter lesions in elderly people: A population based magnetic resonance imaging study. The Rotterdam Scan Study. *Journal of Neurology, Neurosurgery & Psychiatry*, *70*(1), 9–14. https://doi.org/10.1136/jnnp.70.1.9

Eickhoff, C. R., Hoffstaedter, F., Caspers, J., Reetz, K., Mathys, C., Dogan, I., Amunts, K., Schnitzler, A., Eickhoff, S. B. (2021). Advanced brain ageing in Parkinson's disease is related to disease duration and individual impairment. *Brain Communications, 3*(3), 1–12. https://doi.org/10.1093/braincomms/fcab191

Franke, K., Clarke, G. D., Dahnke, R., Gaser, C., Kuo, A. H., Li, C., Schwab, M., & Nathanielsz, P. W. (2017). Premature Brain Aging in Baboons Resulting from Moderate Fetal Undernutrition. *Frontiers in Aging Neuroscience*, *9*. https://doi.org/10.3389/fnagi.2017.00092

Franke, K., & Gaser, C. (2012). Longitudinal Changes in Individual *BrainAGE* in Healthy Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *GeroPsych*, *25*(4), 235–245. https://doi.org/10.1024/1662-9647/a000074

Franke, K., & Gaser, C. (2019). Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Frontiers in Neurology*, *10*, 789. https://doi.org/10.3389/fneur.2019.00789

Franke, K., Gaser, C., Manor, B., & Novak, V. (2013). Advanced BrainAGE in older adults with type 2 diabetes mellitus. *Frontiers in Aging Neuroscience*, *5*. https://doi.org/10.3389/fnagi.2013.00090

Franke, K., Luders, E., May, A., Wilke, M., & Gaser, C. (2012). Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *NeuroImage*, *63*(3), 1305–1312. https://doi.org/10.1016/j.neuroimage.2012.08.001

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, *50*(3), 883–892. https://doi.org/10.1016/j.neuroimage.2010.01.005

Garnier-Crussard, A., Bougacha, S., Wirth, M., André, C., Delarue, M., Landeau, B., Mézenge, F., Kuhn, E., Gonneaud, J., Chocat, A., Quillard, A., Ferrand-Devouge, E., de La Sayette, V., Vivien, D., Krolak-Salmon, P., & Chételat, G. (2020). White matter hyperintensities across the adult lifespan: Relation to age, Aβ load, and cognition. *Alzheimer's Research & Therapy*, *12*(1), 127. https://doi.org/10.1186/s13195-020-00669-4

Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., & Luders, E. (2022). *CAT - A Computational Anatomy Toolbox for the Analysis of Structural MRI Data* [Preprint]. biorXiv. https://doi.org/10.1101/2022.06.11.495736

Gong, W., Beckmann, C. F., Vedaldi, A., Smith, S. M., & Peng, H. (2020). *Optimising a Simple Fully Convolutional Network (SFCN) for accurate brain age prediction in the PAC 2019 challenge* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.11.10.376970

Hahn, T., Fisch, L., Ernsting, J., Winter, N. R., Leenings, R., Sarink, K., Emden, D., Kircher, T., Berger, K., & Dannlowski, U. (2021). From 'loose fitting' to high-performance, uncertainty-aware brain-age modelling. *Brain*, *144*(3), e31–e31. https://doi.org/10.1093/brain/awaa454

Jülich Supercomputing Centre. (2018). JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre Journal of large-scale research facilities, 7, A182. http://dx.doi.org/10.17815/jlsrf-7-182

Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boultwood, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., … Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nature Communications*, *11*(1), 2624. https://doi.org/10.1038/s41467-020-15948-9

Lombardi, A., Monaco, A., Donvito, G., Amoroso, N., Bellotti, R., & Tangaro, S. (2021). Brain Age Prediction With Morphological Features Using Deep Neural Networks: Results From Predictive Analytic Competition 2019. *Frontiers in Psychiatry*, *11*, 619629. https://doi.org/10.3389/fpsyt.2020.619629

Luders, E., Cherbuin, N., & Gaser, C. (2016). Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. *NeuroImage*, *134*, 508–513. https://doi.org/10.1016/j.neuroimage.2016.04.007

Manjon, J., Coupe, P., Marti-Bonmati, L., Collins, D.L., & Robles, M. (2010). Adaptive Non-Local Means Denoising of MR Images With Spatially Varying Noise Levels. *Journal of Magnetic Resonance Imaging, 31*, 192–203. 10.1002/jmri.22003

Nenadić, I., Dietzek, M., Langbein, K., Sauer, H., & Gaser, C. (2017). BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Research: Neuroimaging*, *266*, 86–89. https://doi.org/10.1016/j.pscychresns.2017.05.006

Niu, X., Zhang, F., Kounios, J., & Liang, H. (2020). Improved prediction of brain age using multimodal neuroimaging data. *Human Brain Mapping*, *41*(6), 1626–1643. https://doi.org/10.1002/hbm.24899

Pardoe, H. R., Kucharsky Hiess, R., & Kuzniecky, R. (2016). Motion and morphometry in clinical and nonclinical populations. *NeuroImage*, *135*, 177–185. https://doi.org/10.1016/j.neuroimage.2016.05.005

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., & Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, *68*, 101871. https://doi.org/10.1016/j.media.2020.101871

Pomponio, R., Erus, G., Habes, M., Doshi, J., Srinivasan, D., Mamourian, E., Bashyam, V., Nasrallah, I. M., Satterthwaite, T. D., Fan, Y., Launer, L. J., Masters, C. L., Maruff, P., Zhuo, C., Völzke, H., Johnson, S. C., Fripp, J., Koutsouleris, N., Wolf, D. H., … Davatzikos, C. (2020). Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, *208*, 116450. https://doi.org/10.1016/j.neuroimage.2019.116450

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, *16*(2), 176–186. https://doi.org/10.1109/42.563663

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure and Function*, *223*(1), 297–305. https://doi.org/10.1007/s00429-017-1491-2

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E., & Miller, K. L. (2019). Estimation of brain age delta from brain imaging. *NeuroImage*, *200*, 528–539. https://doi.org/10.1016/j.neuroimage.2019.06.017

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage*, *23*(1), 84–97. https://doi.org/10.1016/j.neuroimage.2004.05.007

van Veen, H. J., The Dat, L. N., Segnini, A. (2015). Kaggle Ensembling Guide. https://mlwave.com/kaggle-ensembling-guide/

Varzandian, A., Razo, M. A. S., Sanders, M. R., Atmakuru, A., & Di Fatta, G. (2021). Classification-Biased Apparent Brain Age for the Prediction of Alzheimer's Disease. Frontiers in Neuroscience, 15, 673120. https://doi.org/10.3389/fnins.2021.673120

Vieira, S., Pinaya, W. H. L., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews*, *74*, 58–75. https://doi.org/10.1016/j.neubiorev.2017.01.002

Supplementary material

Table 1

Performance of different kernels used in RVR on a subsample

Kernel	MAE [years]
polynomial (degree=1)	2.5438
linear	2.5423
rbf=1	2.5348
rbf=2	2.5268
rbf=5	2.5217
rbf=10	2.5243

Table 2

Mean absolute errors for different age-bias corrections of the BrainAGE estimated on a subsample

		MAE
trend correction	omitted	2.9953
	mean correction	2.9954
	linear	2.5502
	quadratic	2.5438
	cubic	2.5338