Review

Machine learning for brain age prediction: Introduction to methods and clinical applications

Lea Baecker¹,*, Rafael Garcia-Dias¹, Sandra Vieira¹, Cristina Scarpazza¹,², Andrea Mechelli¹

¹ Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, UK
² Department of General Psychology, University of Padua, Italy

A R T I C L E   I N F O

Article History:
Received 26 May 2021
Revised 13 September 2021
Accepted 14 September 2021
Available online 4 October 2021

Keywords:
brain age
brain-age gap
machine learning
ageing

A B S T R A C T

The rise of machine learning has unlocked new ways of analysing structural neuroimaging data, including brain age prediction. In this state-of-the-art review, we provide an introduction to the methods and potential clinical applications of brain age prediction. Studies on brain age typically involve the creation of a regression machine learning model of age-related neuroanatomical changes in healthy people. This model is then applied to new subjects to predict their brain age. The difference between predicted brain age and chronological age in a given individual is known as ‘brain-age gap’. This value is thought to reflect neuroanatomical abnormalities and may be a marker of overall brain health. It may aid early detection of brain-based disorders and support differential diagnosis, prognosis, and treatment choices. These applications could lead to more timely and more targeted interventions in age-related disorders.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Ageing and its associated health conditions present a major challenge to individuals and societies worldwide. To address this challenge, increasing efforts are being made towards the early detection of age-related diseases with the ultimate aim of preventing or delaying their progression. The effects of ageing on the brain can be measured through an approach known as brain age prediction, which builds on the well-established relationship between age and neuroanatomy across the lifespan [1]. The past decade has seen an exponential increase in studies on brain age (Figure 1). The majority of these involve the application of machine learning methods to structural neuroimaging data. Machine learning models learn patterns from data and then use these patterns to make predictions about new data. A key advantage of these methods over traditional statistics is that it is possible to make inferences at individual level rather than at group level, thereby increasing the potential for clinical translation [2]. Brain age prediction studies commonly build a regression machine learning model using structural magnetic resonance imaging (MRI) data from healthy controls. This normative model is then applied to new subjects to assess to what extent their neuroanatomy deviates from the norm and estimate brain abnormalities, resulting in their predicted brain age.

The main outcome measure in brain age prediction is the difference between an individual’s predicted age and their chronological age, which is referred to as ‘brain-age gap’ in the present review. Studies of clinical groups typically estimate the mean brain-age gap across all patients and then either compare it to the mean brain-age gap of a control group or to zero, where predicted and chronological age are equal (Figure 2). A positive brain-age gap means that the individual’s predicted brain age was higher than their actual age, which is sometimes referred to as ‘accelerated’ or ‘premature’ ageing. A negative brain-age gap implies a lower predicted brain age, occasionally referred to as ‘delayed’ ageing. However, further research into the neurobiological mechanisms of brain ageing is needed to assess to what extent the terminology of ‘accelerated’ or ‘delayed’ ageing is warranted. In contrast, studies of healthy people typically assess the accuracy of a prediction model in terms of mean absolute error (MAE), which is the mean of the absolute brain-age gaps across subjects. Where applicable, therefore, this review will report a study’s results as mean brain-age gap (+/- X years) or MAE (MAE X years).

Research suggests that an individual’s brain-age gap can be understood as a marker of brain health [3]. The validity of brain age as an ageing biomarker is supported by the evidence that brain-age gap is significantly correlated with other measures of ageing, such as decline in cognitive function, weaker grip strength, and walking speed [4]. This indicates clear potential for clinical translation.

In this state-of-the-art review, we aim to introduce the reader to the field of brain age prediction and highlight its clinical potential. Our aim is not to present an exhaustive account of the literature but...
to explain the most common methodological approaches to brain age prediction and discuss five promising clinical applications and possible next steps, with reference to the most recent studies. As the vast majority of published studies on brain age prediction use structural MRI, we focus on findings from this modality.

2. Methodological basics of brain age prediction

2.1. Designing a brain age study

When designing a brain age study, the most important decisions a researcher will face concern the type of input data, machine learning model, and performance assessment. This section aims to provide a general overview on these three decisions in the context of a structural brain age prediction study.

As the first step, the researcher has to decide how to pre-process their MRI data. The most common approaches are region-based (e.g. FreeSurfer, http://surfer.nmr.mgh.harvard.edu/) [5–12] or voxel-based (e.g. Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm) [13–19] methods to obtain measures such as regional or tissue-specific volumes, cortical thickness, or surface area. Next, the researcher may choose to use measures from the whole brain [4,11], perform some kind of feature selection [9,20–22], or compare both of these approaches [1,23,24]. Dimensionality reduction through

Figure 1. Number of publications on brain age per year (2010-2020). The search was conducted on the PubMed database using the search term ‘brain age’. The number of publications per year was obtained using the ‘Results by Year’ function.

Figure 2. Overview on the machine learning method of a simplified brain age prediction study.

a. Training and cross-validation (CV): A brain age study often uses k-fold CV during training, which means that k models are trained using (k-1)/k of the main sample, while 1/k of the sample (different for each fold) is used as a hold-out set to test how well the model predicts the subjects’ ages. CV may be used to tune hyperparameters of the machine learning model, where a different parameter is tested in each fold. This figure illustrates a 10-fold CV approach.

b. Testing (optional): The trained model is applied to an independent dataset to test. Using an independent dataset allows a better estimation of model bias.

c. Calculation of brain-age gap: Brain-age gap is calculated for each subject as predicted age / chronological age.
automatic models like principal component analysis are commonly employed to reduce the high dimensionality of voxel-based data and remove redundant information, as this can reduce computational cost and increase accuracy [1,8,24–26].

As the second step, the researcher will choose and develop a machine learning model. The majority of publications on brain age prediction uses supervised machine learning methods, meaning that the models are first trained on labelled data (i.e. the subject’s MRI scan is associated with their chronological age) and then applied to a test dataset without labels to assess how well they predict the brain age of unseen subjects. The majority of these models make use of regression techniques, where structural brain features are the independent variables and chronological age is the dependent variable [1,4,11,12,24]. Overall, the available machine learning models for brain age prediction differ with regards to complexity, computational resources, and involvement by the researcher. Recent studies compared the performance of commonly used models like support vector regression and relevance vector regression to provide guidance on the most suitable model choices for brain age prediction (MAE 2.6–7.7 years [12]; MAE 3.7–4.7 years [24]); however, as demonstrated by Wolpert [27] in what is known as the ‘no free lunch’ theorem for machine learning, the performance of different models will depend on the characteristics of the datasets, so there is no single best model for a certain task. This means that a researcher may choose to train different types of models on their data before choosing the most suitable one.

Studies have begun to explore deep learning approaches for brain age prediction, which are potentially more complex and powerful than supervised methods [28–34]. Nevertheless, in direct comparison to the commonly used shallow machine learning approaches like relevance vector regression, deep learning approaches appear to be comparable (MAE 4–5 years) [28,29] or superior (MAE 7–8 years versus MAE 5–6 years) [33]. One of the main advantages of deep learning methods is that they can be applied to raw structural MRI data, which may make the prediction models less susceptible to bias from pre-processing decisions [28] and ultimately more translational.

As the third step, the researcher may choose to assess model performance through cross-validation (CV) in the same dataset used for training, and/or evaluate generalisation performance in an independent dataset (Figure 2). It is highly recommended that all models are trained and tested in distinct datasets, which provides a more reliable estimation of performance in unseen data from different scanner and acquisition protocols. In practice, however, the approach often depends on the amount of available data, with CV used in the context of smaller studies where it is not possible to have separate training and testing datasets.

2.2. Potential sources of bias

There are several sources of bias that may affect the performance of a brain age model. These include, among others, sex [4,12,16,35–39], body-mass index [26,34,40], physical exercise [4,35,41,42], substance use [20,26,35,43], and cognitive ability [4,34,37,41,44]. For clinical samples, studies commonly examine how medication [6,10,20,26,45], illness duration [6,14,15,20,43,45], and symptom severity [6,10,26,37,45] affect the brain-age gap. A particular challenge for clinical applications is that some of these factors, such as smoking and substance use, may be especially prevalent among certain clinical groups, so it is important to adjust for these to minimise model bias.

To date, little attention has been paid to the potential impact of ethnicity [46–49], socioeconomic status [50], and education [42] on brain age. For example, the vast majority of brain age studies has been conducted in Caucasian/Western subjects, although the association of ethnicity and/or culture with brain structure is recognised [46–49]. Hence, current brain age models might not allow for reliable predictions for people of other ethnicities, especially if disease effects are subtle; to address this limitation, it is crucial to take ethnicity into account to minimise model bias.

Chronological age is increasingly recognised as an important source of systematic bias [11,36,51–55]. Brain age models tend to be affected by regression to the mean, so the age of younger subjects is overestimated and the age of older subjects is underestimated. Various statistical approaches have been proposed to correct for this age bias [10,11,36,51–55]. Whether a study took age bias into account therefore is an important factor for their interpretation.

Other sources of bias and variance may stem from pre-processing decisions. For example, standardised pre-processing includes steps such as normalisation to a template, which may introduce bias when applied to brains that are considerably different from the template, especially in the presence of some kind of pathology [28]. Hence, further research is needed to minimise the required pre-processing decisions, for example using deep learning as discussed in section 2.1 [28].

3. Five promising clinical applications

Brain age has a range of potential applications for the clinical assessment of individual patients at various stages of health and disease, including the support of diagnosis, prognosis, and treatment decisions (Figure 3). Brain age studies are being conducted in a wide range of clinical populations including neurological conditions such as Alzheimer’s disease (AD) and mild cognitive impairment (MCI) [1,14,15,23,37,56–59], traumatic brain injury [60,61], epilepsy [18,19,62], multiple sclerosis [37,54,63], and stroke [64,65], as well as psychiatric disorders such as schizophrenia [10,12,69,20,26,37,38,45,66–68], including clinical high-risk for psychosis (CHR) and first-episode psychosis (FEP), bipolar disorder [37,38,68,69,70], major depressive disorder (MDD) [6,20,32,37,71,72], borderline personality disorder [20], autism spectrum disorder [8,37,73], and attention deficit hyperactivity disorder [37]. The results of the clinical studies are summarised in Tables 1 and 2.

3.1. Marker of general brain health

Predicted brain age could become part of regular clinical check-ups to assess general brain health. Here, a high brain-age gap may prompt the treating clinician to run further tests and/or suggest lifestyle changes. This clinical application is based on the evidence that brain age is correlated with a range of brain-related disorders, as described above (Tables 1 and 2), and is predictive of mortality risk [4].

Brain age as a marker of general brain health requires a clear understanding of what factors contribute to a positive or negative brain-age gap other than manifestations of disease, such as risky or protective lifestyle behaviours. For instance, higher brain-age gap has been found to be associated with a range of markers of poor health, such as smoking (+3.4 years relative to controls) and alcohol consumption (+4.1 years) [43] or high diastolic blood pressure (+6.6 years) [39]. Several studies - though not all [74] - have also reported an association between obesity/high body-mass-index and higher brain-age gap [26,39,40,69], with an increase of up to 10 years [40].

While the majority of studies illustrate the pathology of accelerated ageing patterns, a few studies have revealed protective effects of specific practices on brain age. For example, people who meditate regularly (-7.5 years) [16], practise amateur music (-4.5 years) [75], or have higher levels of education or physical activity (-1.5 years) [42] had lower brain-age gaps than controls.

A general barrier to any clinical implementation of brain age is the requirement for highly accurate and replicable performance across a variety of scanning environments (Section 4.1) and in the presence of
At different clinical stages, brain age may be used to...

**Healthy person**
- **... assess general brain health.**
  A clinician may suggest lifestyle changes to improve brain health (e.g., reduced substance use).

**Person with prodromal symptoms**
- **... detect disorders before symptom onset.**
  Early intervention may help prevent or delay illness onset.

**Person with a brain-based disorder**
- **... make prognostic predictions about transition risk, functional outcome, or relapse.**
  A clinician can allocate resources to those with a worse prognosis.
- **... support differential diagnosis.**
  Disorders with overlapping symptoms may have different patterns of brain ageing.
- **... aid treatment decisions.**
  Treatment nonresponse may be predicted at illness onset using brain age.

---

Figure 3. Potential clinical applications of brain age at different stages of the patient lifecycle. Brain age has a range of potential uses in health and disease of an individual person.

---

Table 1
Overview of brain age prediction studies on neurological disorders. Studies were included if they reported mean brain-age gaps from machine learning models trained on healthy controls and applied to clinical groups. Where the table lists more than one mean brain-age gap, the study evaluated multiple models.

| Authors          | Clinical group | n     | Age range | Mean brain-age gap |
|------------------|---------------|-------|-----------|-------------------|
| Mohajer et al. 2020 [58] | AD            | 48    | 56-91     | +9.10             |
| Ly et al. 2020 [14]       | AD            | 74    | 60-85     | +6.79             |
| Beheshti et al. 2018 [59] | AD            | 147   | n.s.      | +5.36             |
| Varikuti et al. 2018 [57] | AD (APOE carrier) | 101   | n.s.      | +5.76             |
| Lowe et al. 2016 [15]     | AD (APOE noncarrier) | 49    | n.s.      | +6.20             |
| Franke et al. 2012 [25]   | AD            | 150   | n.s.      | +6.67             |
| Franke et al. 2010 [1]    | AD            | 102   | 55-88     | +10.00            |
| Mohajer et al. 2020 [58]  | MCI           | 222   | 56-91     | +4.00             |
| Ly et al. 2020 [14]       | MCI (early stages) | 195   | 60-85     | +1.02             |
| Beheshti et al. 2018 [59] | MCI (stable)  | 102   | n.s.      | +2.38             |
| Varikuti et al. 2018 [57] | MCI           | 64    | 55-87     | +6.20(+5.40)      |
| Lowe et al. 2016 [15]     | MCI (stable, APOE carrier) | 14    | n.s.      | -0.88             |
| Gaser et al. 2013 [56]    | MCI (stable, APOE noncarrier) | 22    | n.s.      | +0.09             |
| Franke et al. 2012 [25]   | MCI (progressive, APOE carrier) | 78    | n.s.      | +5.83             |
| de Bezenac et al. 2021 [62] | MCI (progressive, APOE noncarrier) | 34    | n.s.      | +5.54             |
| Gaser et al. 2013 [56]    | MCI (progressive, early) | 58    | 55-86     | +8.73             |
| Gaser et al. 2013 [56]    | MCI (progressive, late) | 75    | 56-88     | +5.62             |
| Franke et al. 2012 [25]   | MCI (stable)  | 62    | 58-88     | +0.75             |
| de Bezenac et al. 2021 [62] | MCI (progressive) | 112   | n.s.      | +6.19             |
| Sone et al. 2021 [19]     | TLE (no psychosis) | 206   | n.s.      | +5.30             |
| Pardoe et al. 2017 [18]   | TLE (with psychosis) | 21    | n.s.      | +10.90            |
| Cole et al. 2020 [54]     | Focal epilepsy (refractory) | 94    | n.s.      | +4.50*             |
| Høgestøl et al. 2019 [63] | Multiple sclerosis | 1354  | 15-68     | +10.30            |
| Egorova et al. 2019 [65]  | Multiple sclerosis | 76    | 21-49     | +4.40*             |
| Savjani et al. 2017 [61]  | Stroke | 135   | >18       | +3.87*             |
| Cole et al. 2015 [60]     | Traumatic brain injury | 92    | 22-57     | +5.4(+3.6)+9.8    |

Abbreviations: AD, Alzheimer’s disease; MCI, mild cognitive impairment; n.s., not specified; TLE, temporal lobe epilepsy

* Mean brain-age gaps marked with an asterisk reported the brain-age gap difference to healthy controls.
Table 2

Overview of brain age prediction studies on psychiatric disorders. Studies were included if they reported mean brain-age gaps from machine learning models trained on healthy controls and applied to clinical groups. Where the table lists more than one mean brain-age gap, the study evaluated multiple models.

| Authors               | Clinical group                  | N   | Age range | Mean brain-age gap |
|-----------------------|---------------------------------|-----|-----------|--------------------|
| Lee et al. 2021 [12]  | Schizophrenia                   | 90  | n.s.      | +3.80 to +5.20     |
|                       | Schizophrenia                   | 75  | n.s.      | +4.53 to +11.72    |
| Nenadić et al. 2017 [38] | Schizophrenia                   | 45  | 21-64     | +2.56              |
| Schnack et al. 2016 [45] | Schizophrenia                   | 341 | 16-76     | +3.36              |
| Koutsouleris et al. 2014 [20] | Schizophrenia (total)         | 141 | n.s.      | +5.50              |
|                       | Schizophrenia (recent onset)   | 61  | n.s.      | +4.20*             |
|                       | Schizophrenia (recurring)      | 80  | n.s.      | +6.40*             |
| Van Gestel et al. 2019 [70] | Bipolar disorder (lithium Tx) | 41  | 20-72     | +0.48 (nonsignificant) |
|                       | Bipolar disorder (no lithium Tx) | 43 | 26-74     | +4.28              |
| Nenadić et al. 2017 [38] | Bipolar disorder               | 22  | 23-57     | -1.25 (nonsignificant) |
| McWhinney et al. 2021 [26] | First-episode psychosis         | 183 | 18-35     | +3.39              |
| Hajek et al. 2019 [67] | First-episode schizophrenia     | 43  | 15-35     | +2.64              |
| Chung et al. 2018 [10] | First-episode psychosis         | 14  | n.s.      | +1.17*             |
| Kolenuc et al. 2018 [69] | First-episode psychosis         | 120 | 18-35     | +2.64              |
| Hajek et al. 2019 [67] | Genetic risk of bipolar disorder | 96  | 15-35     | nonsignificant*    |
| Chung et al. 2018 [10] | CHR (total)                     | 275 | 12-21     | +0.64*             |
|                       | CHR (converted)                 | 17  | 12-17     | +1.58*             |
|                       | CHR (not converted)            | 125 | 12-17     | nonsignificant*    |
|                       | CHR (converted)                 | 22  | 17-21     | nonsignificant*    |
|                       | CHR (not converted)            | 120 | 17-21     | nonsignificant*    |
| Koutsouleris et al. 2014 [20] | CHR (total)                    | 89  | n.s.      | +1.70*             |
|                       | CHR (early onset)               | 21  | n.s.      | approx.-3*         |
|                       | CHR (late onset)                | 68  | n.s.      | +2.70*             |
| Koutsouleris et al. 2014 [20] | Borderline personality disorder | 57  | n.s.      | +3.10*             |
| Han et al. 2021 [72]  | MDD                             | 195 | 11-37     | +0.57              |
| Han et al. 2020 [6]   | MDD                             | 2675| 18-75     | +1.08*             |
| Christman et al. 2020 [32] | MDD (adult)                    | 76  | 20-50     | nonsignificant*    |
|                       | MDD (geriatric)                | 118 | 40-90     | approx.-4.5*       |
| Bestheer et al. 2019 [71] | MDD                            | 38  | 19-66     | nonsignificant*    |
| Koutsouleris et al. 2014 [20] | MDD                           | 104 | 18-65     | +4.00*             |

Abbreviations: CHR, clinical high risk for psychosis; MDD, major depressive disorder; n.s., not specified; Tx, treatment

* Mean brain-age gaps marked with an asterisk reported the brain-age gap difference to healthy controls.

a range of confounding factors (Section 2.2). Over the past years, brain age models have generally become more accurate, as more data sharing initiatives and advanced machine learning methods become available, and it is likely that this will continue getting better. However, a specific limitation for brain age as a marker of general brain health remains relevant: it might not add sufficient new information to a clinical assessment that would justify the costs of an MRI scan. For instance, a person’s weight, blood pressure, smoking and alcohol consumption and their potential negative effects on health are generally already known to the clinician. Nevertheless, brain age could become a standard output from every MRI scan that is already conducted for other reasons, especially if it can be provided in real-time at minimal extra cost [28]. It might still be a useful health marker, because, as pointed out by Cole and colleagues [76], the concept of an older-appearing brain could be easier for patients to understand than conventional clinical measures.

3.2. Early detection of brain-based disorders

Using brain age as a screening tool could facilitate early detection of disorders or even their preclinical stages, which, in turn, would allow early intervention. Early intervention in age-related disorders is an important clinical focus, because it tends to be associated with better functional outcome in the long run, e.g. for psychosis [77]. Studies suggest that for schizophrenia, its preclinical stage CHR (up to +2.7 years) [10,20] and early stage FEP (up to +3.4 years) [67] already appear to be associated with higher brain-age gaps (Table 2). Similarly, while the higher brain-age gap in subjects with AD is well known (up to +10.0 years) [1,15,25,37,56], its preclinical stage MCI also displays neuroanatomical changes that make it distinguishable from healthy controls (up to +6.2 years) (Table 1) [14,15,25,37,56]. Brain age may therefore present a helpful screening tool in these cases, especially in combination with the assessment of general brain health (Section 3.1). Of note, Beheshi et al. [59] showed associations between brain-age gap and traditional survey-based screening tools for AD, such as the Mini-Mental State Examination, which are prone to methodological bias and confounding factors. Adding a biological dimension to the screening process through brain age could therefore make diagnosis more reliable.

Tables 1 and 2 illustrate that the extent of the brain-age gap across brain disorders tends to be similar, indicating a lack of specificity. While this can be an issue in the diagnostic use (Section 3.4), brain-age gap may be useful as a transdiagnostic marker for early detection of brain disorders. For instance, a high brain-age gap in an individual without obvious clinical symptoms could lead to further tests, which might reveal pre-symptomatic disease. Future studies need to investigate whether a higher brain-age gap is present in the early stages of disorders other than schizophrenia and AD.

3.3. Prognosis of brain-based disorders

Once a preclinical stage of a disorder is identified, being able to establish a person’s (1) risk of transition to the full-blown disorder, (2) future functional outcome, or (3) risk of relapse can aid targeted intervention and thus save resources. For example, only about a third of CHR patients develop full-blown psychosis within three years [78]. If those patients at greatest risk of transition could be identified through measures like brain age (up to +2.7 years) [10,20], clinical resources could be targeted at them instead of the whole CHR population. Similarly, the higher brain-age gap of MCI subjects was associated with higher risk of converting to AD [15,25,56]. This early evidence suggests the potential of brain age to estimate risk of transitioning to full-blown illness.
Brain age is also associated with various functional markers, which may enable the identification of subjects that are likely to experience worse symptoms and would thus benefit most from clinical intervention. For instance, the observation that the brain-age gap appears to be linked with cognition [4,25,39,56,60,79] and scores on clinical scales [25,26,37,59] suggests that it might be possible to use it as marker of future cognitive decline and disease progression, e.g. in AD [25]. Longitudinal studies are needed to explore this further.

In disorders that are characterized by recurring episodes, such as psychosis or multiple sclerosis, it is highly beneficial to be able to predict which patients are more vulnerable to future relapses in order to intervene and potentially prevent them. To our knowledge, relapse prediction has not yet been studied with brain age.

Overall, prognostic applications of brain age could be helpful transdiagnostic as well as disorder-specific markers to highlight those individuals that require more clinical attention. Indeed, prognosis could be one of the most useful clinical applications, because it contributes information that the clinician would not have had otherwise. So far, research in this field has focused on schizophrenia and AD, so future research should address to what extent risk of transition or functional outcome can also be predicted in other disorders.

3.4. Differential diagnosis of brain-based disorders

The presence of some brain disorder or abnormality can usually be detected using standard clinical measurements or scales, but the challenge lies in making an accurate diagnosis. Differential diagnosis is a particular challenge in psychiatric disorders, where overlapping symptoms between diagnoses are a known issue, along with the prevalence of comorbidities. For example, misdiagnosis of the psychotic disorders schizophrenia and bipolar disorder is common and difficult to diagnose. Accurate diagnosis can take several years [77]. Initial studies suggest that the two psychiatric disorders may differ in brain age. An increased brain-age gap has consistently been found in subjects with schizophrenia (up to +11.7 years) [12,20,37,38,45], while the effect of bipolar disorder on brain-age gap is much less consistent [37,38,67,70] (Table 2). Although further studies are required, this initial evidence suggests that if an individual shows early symptoms of psychosis, an MRI scan may help identify if they are more likely to develop schizophrenia (increased brain-age gap relative to healthy controls) or bipolar disorder (more likely to have normal brain-age gap). However, a study looking at the effect of lithium in bipolar disorder found that those participants treated with lithium had normal brain-age gap while those not treated had a higher brain-age gap [70]. This highlights the importance of looking at medication use as a possible confounding factor (Section 2.2).

It is important to note that an abnormal brain age cannot be a stand-alone measure of diagnosis, as it lacks specificity, especially in light of the inherently large neuroanatomical heterogeneity in the general population. For example, a recent large-scale study from the ENIGMA-MDD working group found large within-group variance for both clinical and control samples with small (albeit significant) between-group difference (+1.1 years) [6]. The brain-age gap therefore has implications on the group level but has limited clinical meaning for individual MDD patients. Although the ENIGMA-MDD group did not find a significant association of brain-age gap and clinical factors such as symptom severity or remission status in a large sample of MDD [6], others reported that increased brain-age gap correlated with age of onset and symptom severity more than with the specific diagnosis in a sample of patients with schizophrenia, MDD and borderline personality disorder [20]. As discussed by Cole et al. [76], it is plausible that brain aging may be a “global phenomenon”, where different initial brain abnormalities manifest as similar changes downstream. As more large-scale longitudinal studies are being conducted, we gain greater understanding of dynamic ageing patterns [15,18,25,26,45,54,56,63,65], but further investigations are needed to examine brain age during the course of various diseases. These will help establish when abnormal brain ages start being noticeable and how they develop over time (Section 4.3), also on the regional level (Section 4.2).

3.5. Treatment outcome

Treatment nonresponse in disorders such as psychosis or epilepsy is common [80,81]. Longitudinal studies may be used to investigate if future treatment response can be predicted using brain age at baseline. To our knowledge, only three studies have examined treatment response with regards to brain age [18,64,70]. One longitudinal study assessed response to cognitive training interventions in stroke patients, but global brain-age gap was not sensitive enough to predict treatment outcome [64]. The other two studies used cross-sectional designs that did not allow for predictions about future outcome, but one of these found that subjects with treatment-resistant focal epilepsy had advanced brain age (+4.5 years) while this effect was small and nonsignificant in those with recently-diagnosed focal epilepsy [18]. The authors speculate that this nonsignificant effect might be due to the presence of subgroups who will and will not develop treatment-resistant epilepsy in the future, suggesting that brain age could be used to identify those more likely to be treatment resistant. Future studies will need to establish whether this speculation is the case not only for epilepsy but also for other brain disorders.

4. Outstanding questions and next steps

A number of outstanding questions need to be addressed before brain-age gap could be considered a reliable and specific clinical marker. These questions and possible next steps towards clinical implementation are discussed in this section.

4.1. Account for inter-scanner heterogeneity

How can we develop brain age models that are robust against inter-scanner heterogeneity? The impact of the scanner and the scanning protocol on the quality of the images is an important challenge in the field of neuroimaging in general, affecting the results of multi-site studies and the generalisation performance of machine learning models to new scanners. This is a particularly important consideration for applications to clinical practice, as scans obtained in the clinical setting tend to be of considerably lower resolution and higher slice thickness than in the research setting; the impact of using scans obtained in real-world clinical settings is unknown. Although some multi-site studies in children and detection of AD have found brain age to be relatively robust to scanner differences [1,25,37,39], others found scanner-dependent performance differences [25]. When a sufficiently large dataset from each scanner is available, it is possible to make corrections to multi-site data and reduce scanner bias, either by regressing out the scanner differences [25] or by applying harmonisation tools [82]. Future studies should examine the robustness of these corrections using a greater range of scanners, including acquisition protocols commonly used in clinical settings. It is also possible that brain age models using deep learning may be more robust to inter-scanner heterogeneity, so additional strategies for correcting scanner bias would not be required. However, as commented before, deep learning approaches have not consistently achieved better accuracies than shallow learning applications [28,29]. Hence, the advantages of deep learning for mitigating the impact of inter-scanner heterogeneity in brain age prediction look promising but require further research.

4.2. Increase granularity of brain age

To what extent do different brain regions follow different ageing patterns? At present, predicted brain age is typically studied as a
single whole-brain measure. This means the same brain-age gap in two different subjects may arise from very different neuroanatomical signatures. Looking at brain ageing patterns for specific brain regions could reveal distinct patterns of ageing between disorders. Initial studies have estimated regional differences either by (1) examining the weight of specific features providing information on different regions [10,12,45,74,79,83], (2) resampling [34,42], or (3) comparing models trained on individual regions (or a subset of regions) to those trained on the whole brain [37,63]. For example, using the latter approach, Kaufmann et al. [37] found that while most regional brain-age gaps corresponded to whole-brain models, larger brain-age gaps were reported in specific regions for disorders such as dementia and multiple sclerosis (cerebellum and subcortical regions), schizophrenia (frontal lobe), and MDD (temporal lobe). These findings suggest that region-level brain ages could support differential diagnosis.

Although this state-of-the-art review is focused on structural MRI as a single modality, it is important to note that other types of neuroimaging may hold complementary information about brain ageing. It has been shown that multimodal approaches lead to performance improvements, for example when combining structural with functional MRI [36,84,85] or diffusion MRI [66,85]. Overall, the integration of different types of data is not only likely to improve prediction, but it may also provide considerably greater granularity for a person’s brain-age gap because of potential tissue- or modality-specific brain ageing patterns.

4.3. Dynamic changes of brain age

How does the brain-age gap of a person change across their lifespan? The majority of studies on brain age are cross-sectional, so it is not yet clear how it develops over time. Longitudinal studies are needed to investigate the potential dynamic changes of brain age in health and disease, which may aid the early detection and the differential diagnosis of disorders with overlapping symptom profiles. For example, two disorders may be characterized by the same extent of brain age deviation, but longitudinal studies could reveal that one disorder displays a one-off insult to neuroanatomy while the other one is progressive. In longitudinal studies of schizophrenia, illness duration was associated with larger brain-age gap [15,25,45] but the acceleration rate may be faster at earlier than later stages [45].

Longitudinal studies could also shed light on potential reversal of advanced brain ageing through treatments, be it clinical intervention or lifestyle changes. Initial evidence suggests that clinical interventions could reduce brain age. For instance, in subjects with refractory epilepsy, neurosurgery reduced the brain-age gap compared to healthy controls from 7.9 years to 2.8 years [62]. In another study, receiving ibuprofen decreased the brain-age gap in healthy controls by 1.2 years after 45 min [74]. To our knowledge, the potential effect of lifestyle interventions on brain age has not been studied in a longitudinal setup yet. However, the past 20 years have increasingly seen evidence on learning-dependent structural neuroplasticity [86], so it seems logical that such increases in grey matter would also affect brain age. As obesity has repeatedly been linked to be associated with increased brain age, a small study found that exercise-dependent weight loss may induce plasticity [87]. Therefore, there is reason to speculate about reversal of accelerated brain ageing, but the types of beneficial interventions and the permanence of the effects remain an area of further investigation.

5. Conclusion

The increasing recognition of the clinical potential of brain age prediction has led to an exponential increase in the number of patient studies, but it has not been translated to clinical practice yet. Its clinical implementation will require greater evidence of clinical utility and cost-effectiveness, as well as translation of current machine learning models into practical and acceptable tools that can be used by clinicians without specialised methodological expertise [88]. To this end, the brain-age gap of an individual patient could be integrated within personalised reports of online clinical tools [89]. Among the five clinical applications discussed in this review, we suggest that the most promising one is the detection of disorders prior to symptom onset. Once a preclinical stage is detected, brain age may also be used to predict risk of transition to full-blown illness, so intervention efforts can target those individuals who are most likely to benefit from them.

6. Search strategy and selection criteria

Data for this state-of-the-art review were identified through searches of PubMed using the keyword “brain age”, along with reference tracking from relevant articles. For the individual clinical applications, we conducted additional searches of (“brain age” AND “clinical”), (“brain age” AND “diagnosis”), (“brain age” AND “prognosis”), and (“brain age” AND “treatment”). The searches were conducted in June 2021 and study selection prioritised publications from the past three years. The primary focus of the search was on structural MRI studies from adult subjects, specifically T1-weighted MRI. As a state-of-the-art review, articles were selected for inclusion based on the authors’ judgment of their relevance and suitability.

Contributors

L.B. - Writing – original draft
R.G.D. - Writing – review & editing
S.V. - Writing – review & editing
C.S. - Writing – review & editing
A.M. – Funding acquisition, Conceptualization, Supervision, Writing – review & editing
All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

This research has been supported by a Wellcome Trust’s Psychiatry Flagship Innovations (220402/Z/20/Z) to AM. The present work was carried out within the scope of the research programme Dipartimenti di Eccellenza (art.1, commi 314-337 legge 232/2016), which was supported by a grant from MIUR to the Department of General Psychology, University of Padua. Funders had no role in data collection, analysis, interpretation, writing, or the decision to submit this manuscript for publication.

References

[1] Franke K, Ziegler G, Klöppel S, Gaser C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. NeuroImage 2010;50(3):883–92.
[2] Vieira S, Pinaya W, Mechelli A. Introduction to machine learning. In: Mechelli A, Vieira S, editors. Machine learning: Methods and applications to brain disorders. Academic Press; 2020. p. 1–20.
[3] Cole BH, Franke K. Predicting age using neuroimaging: Innovative brain ageing biomarkers. Trends Neurosci 2017;40(12):681–90.
[4] Cole BH, Ritchie SJ, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle N, et al. Brain age predicts mortality. Mol Psychiatry 2018;23:1385–92.
[5] Lombardi A, Monaco A, Donvito G, Amoroso N, Bellotti R, Tanganaro S. Brain Age Prediction With Morphological Features Using Deep Neural Networks: Results From Predictive Analytic Competition 2019. Front Psychiatry 2021;11(January):1–15.
[6] Han LKM, Dinga R, Hahn T, Ching CRK, Eyler LT, Aftanas L, et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. Mol Psychiatry 2020.
Franke K, Gaser C. Longitudinal changes in individual BrainAGE in healthy aging.

McWhinney S, Kolenic M, Franke K, Fialova M, Knytl P, Matejka M, et al. Obesity and brain age prediction using cortical thickness data: A large scale cohort study. Front Aging Neurosci 2018;10:317;1–11.

Lee WH, Antoniades M, Schnack HG, Kahn RS, Frangou S. Predicting brain age with deep learning from raw imaging data results in a reliable estimate of brain age in mild cognitive impairment, and Alzheimer Disease. PLoS One 2018;13(1):e019063.

Stein MS, Shock J, Stein DJ, du Plessis S, Seedat S, Tsvetanov KA. Correlation Constraints for Regression Models: Controlling Bias in Brain Age Prediction. Front Aging Neurosci 2019;11(348):4.

Richard G, Kolska I, Dietzek M, Langbein K, Sauer H. BrainAGE score indicates physical ability. Neuroimage 2020;25:102159.

Varchione M, Rupavitch S. Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. JAMA Psychiatry 2017;74(5):459–66.

Christiansen S, Bermudez C, Hao L, Landman BA, Boyd B, Albert K, et al. Accelerated brain aging predicts improved cognitive performance and greater disability in geriatric but not mild adult depression. Trans Psychiatr Res 2020;34(1):15–21.

Jiang H, Lu N, Chen K, Yao L, Li K, Zhang J, et al. Brain structure differences between Chinese and Caucasian cohorts: A comprehensive morphometry study. Neuroimage 2019;200:104–14.

Mohajer B, Abbasi N, Mohammadi E, Khazaie H, Osorio RS, Rosenzweig I, et al. The association between “Brain-Age Score” (BAS) and traditional neuropsychological screening tools in Alzheimer’s disease. Brain Behav 2018;8(e01020):1–14.

Lee WH, Antoniades M, Schnack HG, Kahn RS, Frangou S. Predicting brain age with deep learning from raw imaging data results in a reliable estimate of brain age in mild cognitive impairment, and Alzheimer Disease. PLoS One 2018;13(1):e019063.

Varchione M, Rupavitch S. Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. JAMA Psychiatry 2017;74(5):459–66.
