Environmental Research 267 (2025) 120728

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Investigating associations between the physical living environment and hippocampus in adulthood and older age

Johanna Drewelies [a,b], Angela Fiedler [a], Timothy R. Brick [c], Simone Kühn [a,d,e,*]

[a] Center for Environmental Neuroscience, Max Planck Institute for Human Development, Berlin, Germany
[b] Humboldt Universität zu, Berlin, Germany
[c] Institute for Computational and Data Sciences, The Pennsylvania State University, State College, PA, USA
[d] University Medical Center Hamburg-Eppendorf, Clinic and Policlinic for Psychiatry and Psychotherapy, Hamburg, Germany
[e] Max Planck-UCL Center for Computational Psychiatry and Ageing Research, Germany

ARTICLE INFO

Keywords:
UK biobank
Environment
Brain volume
Hippocampus

ABSTRACT

It is by now well known that the physical living environment has a major impact on people's life, but the neural structures involved in this relationship remain to be explored. Most studies investigating this relationship only focus on single environmental predictors. In order to understand how the multitude of factors constituting the living environment relate to brain structure we used data from the UK Biobank (n = 21,094; age Mean = 63.35 years; SD = 7.46; range = 45–82) to examine how individuals' immediate characteristics around the home address (e.g., green space; air pollution in the neighborhood) are associated with hippocampal volume, a brain region known to be highly plastic. We accounted for common demographic factors that have been shown to be associated with brain structure and known factors such as sex, income, education, and age. We made use of an analytical paradigm based on the feature importance estimation and recursive feature elimination with decision tree ensembles as well as linear regression analysis. Results identified a subset of environmental measures (e.g., pollution, green space, noise) most strongly associated with hippocampal volume across adulthood. Findings highlight the importance of the environment for individuals' brain structure.

1. Introduction

It has been repeatedly shown that the environment one lives in has an effect on cognitive and physical health. For example, the negative effect of air pollution on health has been demonstrated in a number of studies (Peeples, 2020; Dockery et al., 1993; Anderson et al., 2012). However, the brain structures underlying the association of the environment and cognitive and physical health remain largely unexplored. The few previous studies that do exist predominantly focus on a single or a limited set of environmental features, a notable limitation given the multitude of different features that characterize the living environment of humans (Tost et al., 2019; Kühn et al., 2021; Dadvand et al., 2018; Nicole, 2018; Mascherek et al., 2022). The present study aims to explore the association between exposure to a multidimensional set of environmental factors and brain structure, specifically the hippocampus, across adulthood and old age, while acknowledging the complexities and limitations inherent in disentangling the contributions of highly intercorrelated predictors.

The hippocampus is a core region of brain plasticity and is linked to a range of cognitive, emotional and behavioral functions (Erickson et al., 2011a; Lagali et al., 2010; Tartt et al., 2022; Kharabian et al., 2020; Kobayashi and Matsuo, 2023). Multiple studies have repeatedly demonstrated its involvement in various aspects of human behavior, characterizing its functioning pattern. Specifically, the hippocampus plays a pivotal role in the formation of long-term memory, learning and consolidation processes (Lagali et al., 2010; Chun and Turk-Browne, 2007; Stella and Treves, 2011; Eichenbaum, 2000; Bonnici et al., 2013) and the deterioration, memory decline in late adulthood and pathological neurodegeneration (Carlesimo et al., 2010; Ewers et al., 2012; Dugger and Dickson, 2017), emotional processing (Thomas et al., 2007; MacQueen and Frodl, 2010a; Fossati, 2012; McEwen and Gianaros, 2011), and special spatial navigation (Rolls, 2020; Chersi and Burgess, 2015), but also psychiatric diseases such as schizophrenia or depression (MacQueen and Frodl, 2010b; Lieberman et al., 2018).

* Corresponding author: Max Planck Institute for Human Development, Lise Meitner Group for Environmental Neuroscience, Lentzeallee 94, 14195, Berlin, Germany.
E-mail address: kuehn@mpib-berlin.mpg.de (S. Kühn).

https://doi.org/10.1016/j.envres.2024.120728
Received 22 January 2024; Received in revised form 22 December 2024; Accepted 27 December 2024
Available online 28 December 2024

Numerous factors have been recognized as influencing changes in hippocampal volume during adulthood and as individuals progress through aging, including stress, physical exercise, and sleep behaviors, among others (Erickson et al., 2011b). A number of environmental factors have been shown to significantly impact the hippocampus across various domains and study designs, including those examining the immediate social environment. Given the hippocampus's central role in cognitive function and its sensitivity to environmental influences throughout adult development and aging, it is crucial to integrate these aspects to better understand the interactions between the physical environment and brain health.

However, individual development does not occur in isolation but is perpetually intertwined with a socio-environmental context, as highlighted by the World Health Organization in 2020. Considering the myriad environmental challenges humankind is facing, such as climate change, heightened pollution levels, and the destruction of natural habitats, there is an urgent need for a better understanding of how the physical environment impacts an individual's brain structure, and specifically the hippocampus. This understanding is particularly crucial in light of global urbanization trends, which involve substantial population shifts from rural to urban areas, resulting in a steady rise in the percentage of people residing in cities where environmental pollutants are more common and green spaces are more rare (McDonald et al., 2013). Currently, over 57% of the world's population is living in urban areas (Nations et al., 2018).

It is noteworthy that most studies investigating the association between the environment and brain structure tend to focus on specific environmental aspects. For instance, a number of studies point to the potential beneficial effects of green spaces on overall brain health (Tost et al., 2019; Kühn et al., 2021; Dadvand et al., 2018; Nicole, 2018), while conversely, research suggests that factors like noise and traffic may have detrimental effects and show negative associations with brain health (Vlahov and Galea, 2002; Xu et al., 2021; Pujol et al., 2016; Nuβbaum et al., 2020). However, it is essential to recognize that environment is a complex interplay of multiple factors, including but not limited to green spaces, noise, pollution, and water, ranging from macro to micro structures, many of which have intrinsic interactions (Mascherek et al., 2022). Given the complex nature of the environment, it becomes imperative to identify the important factors for brain structure, with specific attention to their link to the hippocampus. While the susceptibility of the hippocampus to various contextual factors has been well-documented (Erickson et al., 2011a, 2011b; Kobayashi and Matsuo, 2023; Ehsanifar et al., 2022), there is less understanding of how environmental context specifically affects this critical brain region. Highlighting this gap underscores the importance of focusing our research on the hippocampus in adults.

The present study aims to close the gap by examining the associations between a large set of environmental factors with hippocampus volume in a sample of adults. We associate characteristics around the home address (e.g., the green in the surrounding or air pollution) with data from brain imaging, while accounting for common sociodemographic factors that have been shown to be related to brain structure such as age, sex or income.

2. Methods

2.1. Participants

In the current study, we used cross-sectional, observational data from the UK Biobank (http://www.ukbiobank.ac.uk/), a population-based biomedical study developed to study the environmental, social, and genetic causes of chronic diseases (Sudlow et al., 2015; Collins, 2012). As the purpose of this study was to investigate the relationship between hippocampus volume and environmental variables, we restricted our sample to participants (i) with available data for all variables and excluding outliers (see below for the list of relevant measures and data cleaning procedure) (ii) who reported to live at the same home address throughout the assessment period. The latter point is important, as it ensures that there has been a long-term exposure to the environmental characteristics of the respective home locations, allowing for the possibility to impact hippocampus volume of the residents. The selection criteria implied that the final sample was reduced to n = 21,049 participants, who have lived at their home address for more than 25 years at the time of scanning on average. The final sample consists of 10,121 males and 10,928 females (age mean=63.35 years; SD=7.46; range=45-82). From the total sample of 40,851 participants, data for 19,802 participants were excluded because of missing data on any of the variables under study.

To ensure robustness, we conducted a sensitivity analysis as a follow-up with an even more restricted sample, following the approach of Nobis et al. (2019). Participants were excluded if they reported any conditions related to neurological disorders, psychiatric conditions, substance abuse, or head trauma, as recorded in UK Biobank's illness variable. Neurological conditions included epilepsy, Parkinson's disease, multiple sclerosis, migraine, and dementia. Psychiatric exclusions covered self-reported depression, schizophrenia, anxiety/panic attacks, and bipolar disorder. Participants with substance abuse or dependency issues, including alcohol and opioid dependency, and those reporting a history of head injury, were also excluded. This additional restriction allowed to minimize potential confounding from health conditions known to be associated with brain structure.

2.2. Procedure

The UK Biobank assessed a total of 500,000 participants who were recruited from across Great Britain (England, Scotland, and Wales) between 2006 and 2014 (Allen et al., 2012; Collins, 2012; Miller et al., 2016). All participants provided informed consent ("Resources tab" at https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=200). The majority of participants (approx. 86.2%) resided in urban areas. Data were collected through questionnaires on sociodemographic, medical and psychosocial factors. Verbal interviews were conducted by trained staff members. Additionally, data on participants' residential locations, which includes address history, were collected.

After an initial visit for the assessment of medical information, a subset of participants underwent MRI scanning of the brain. MRI data was collected at an average of around 4 years after the initial visit, and completed on an MRI scanner in Bristol, Newcastle, Reading, and Cheadle. Data on neighborhood characteristics were derived from the UK Biobank Urban Morphometric Platform (UKBUMP). The UKBUMP is a linked database of objectively measured urban morphological metrics quantifying environmental exposures within functional neighborhoods around UK Biobank participants' geocoded dwelling locations (for further details please see (Sarkar et al., 2018a; Sarkar et al., 2018b)). UK Biobank received ethical approval from the Research Ethics Committee (reference 11/NW/0382). The present study was conducted as part of the UK Biobank application (Application-ID: 64615). All participants provided informed consent to participate. Further information on the consent procedure can be found under: UK Biobank Field 200; http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id = 200.

2.3. Measures

An overview of all variables used in this study and their corresponding UK Biobank variable IDs can be found in Table A.4. in the Appendix.

Brain data *Hippocampus Volume* is an imaging-derived phenotype consisting of a raw volumetric measure. As suggested by Williams et al. (2021a) we control for intracranial volume (ICV), where *ICV* was calculated as the sum of the total gray matter volume, cerebellum and cerebral white matter volume from the UK Biobank aseg Freesurfer segmentations. Specifically, the FreeSurfer pipeline (Version 7.2) was

computed and quality control according to the suggestions of ENIGMA (https://github.com/ENIGMA-git) was performed. The results of the aseg segmentation Left-Hippocampus and Right-Hippocampus and the EstimatedTotalIntraCranialVol were computed. To eliminate outliers with implausibly high or low measures for *hippocampus volume* we applied a mean + -2.5 standard deviations (SD) criterion, where all values outside this range were excluded from further analysis. A threshold of ± 2.5 SD was chosen to balance the need for robust statistical analysis with the need to lose as little data as possible. By setting this threshold, the aim was to minimize the impact of atypical measurements arising due to technical errors or biological anomalies. Using the threshold, less than 5 % participants were removed from the analysis. These did not differ from the rest of the sample. In addition, we conducted a comprehensive quality control assessment using the MRIQC framework on the raw T1-weighted images.

We proceeded in the same way with *ICV*. Following Williams et al. (2021) we also control for *MRI Site* of scanning (Cheadle, Reading, Newcastle, Bristol) as a factor variable. As suggested by Alfaro-Almagro et al. (2021), we also take into account possible effects on the brain measures caused by different positions of the scanner-table, captured by the X (*MRI_x*) and Z (*MRI_z)* coordinates of the Centre of Gravity of the T1w brain mask and the Y (*MRI_y*) position of the most posterior part of the same brain mask.

Environmental characteristics. Environmental characteristics were determined using each participants' home address as a reference point. *Major Road (binary)* indicates whether a residential address was located within 50 m of a class 1 or 2 type road and/or within 100 m of a class 0 road, based upon the central road network. The central road network was taken from EuroStreets Version 3.1 digital road network (scale 1:10000), derived from the TeleAtlas MultiNet TM dataset for the year 2008. More details can be found elsewhere (de Hoogh et al., 2016; Vienneau et al., 2013). Motorways, main roads of major importance, other main roads, secondary roads and local connecting roads were classified as major roads (classes 0–4).

Nearest Road was calculated as the distance (m) between an individual residential address and the nearest major road based upon the local road network. The local road network was taken from the Ordnance Survey Meridian 2 road network (scale 1:50000, 1 m accuracy) (2009) (UK Biobank-Enviromental Exposures-Metadata; Li et al., 2023).

Air pollution exposure. Nitric Oxide (NO) and Nitrogen Dioxide (NO2), collectively known as Nitrogen oxides (NOx), are quantified in micrograms per cubic meter (μg/m^3). Similarly, Particulate Matter (PM) is assessed in micrograms per cubic meter (μg/m^3) and is categorized based on particle sizes. $PM_{2.5}$ denotes particles with a size of 2.5 micrometers or less, while $PM_{2.5\text{-}10}$ encompasses particles ranging from 2.5 to 10 micro-meters, commonly referred to as PM coarse. Additionally, $PM_{2.5abs}$ is a metric that captures the darkness of $PM_{2.5}$ filters and serves as an indicator for the predominant light-absorbing substance, elemental carbon (Eeftens et al., 2012). Data on air pollution referred to the year 2010.

Noise Pollution. A number of indicators for noise pollution was included in our analysis. First, *Daytime Noise pollution* were expressed in dB, indicating the average sound level pressure LAeq between the hours of 7:00 a.m. to 7:00 p.m. Noise estimates for the year 2009 were modeled using a version of the CNOSSOS-EU noise model (Morley et al., 2015). Second, *evening noise pollution* was expressed in dB, indicating the average sound level pressure LAeq between the hours of 7:00 to 11:00 p.m. Noise estimates for the year 2009 were modeled using a version of the CNOSSOS-EU noise model. Third, *Nighttime Noise pollution* was expressed in dB, indicating the average sound level pressure LAeq overnight 11:00 p.m. to 7:00 a.m. Noise estimates for the year 2009 were modeled using a version of the CNOSSOS-EU noise model.

Green Space. A number of variables were used as proxies for green space exposure. First, *Natural environment percentage (buffer 300m & 1000m)* indicates the percentage of the 300m/1000m buffer around the home location that is classified as 'Natural Environment' in the Land Cover Map (LCM) 2007. Therefore the 23 land cover classes were reclassified to a binary classification with the classes 'Natural Environment' and 'Built Environment' (Mamouei et al., 2022). Second, *Greenspace percentage* buffer 300m/1000m. The percentage of the 1000m buffer around the home location classed as 'Greenspace', as a proportion of all land use types was modeled using 2005 data from the Generalized Land Use Database for England (GLUD) or the 2001 output Areas in England. Data are only available for participants whose home location was in England. Third, *Domestic Garden percentage (buffer 300m & 1000m)* around the home location classed as 'Domestic Garden', as a proportion of all land use types was modeled using 2005 data from the Generalized Land Use Database for England (GLUD) for the 2001 output Areas in England. Data are only available for participants whose home location was in England.

Water was assessed using two indicators. To begin with, *water percentage (buffer 300m/1000m)* was examined as the percentage of 300m/1000m buffer around the home location classed as 'Water', as a proportion of all land use types that were modeled using 2005 data from the Generalized Land Use Database for England (GLUD) for the 2001 output Areas in England. Second, *Distance (Euclidean)* to coast was operationalized as the distance from the home location to the coast and expressed in km. The geographical data implemented here are the coastline at mean high water, derived from dissolved statistical area boundaries (2004 Lower-layer Super Output Areas) (Mamouei et al., 2022).

Housing. We used a number of indicators describing participants' housing situation. To begin with, *accommodation type* indicates a participant's self-reported type of accommodation. Categories include house or bungalow; flat, maisonette or apartment, mobile or temporary structure (i.e., caravan), sheltered accommodation, and care home. Second, *Rent* was assessed by asking participants whether they own or rent the accommodation they live in was assessed by asking a closed touchscreen question. We also included a dummy variable *gas fire* to indicate if subjects reported to use gas for cooking or were regularly using open fires in winter time.

Townsend deprivation was calculated using a Townsend deprivation index of the participants area calculated immediately prior to participants joining UK Biobank. The index includes the four variables unemployment, non-car ownership, non-home ownership, and household overcrowding. Data is based on the preceeding national census of each output area. Assignment to each area is based on the participant's postcode. Greater Townsend scores indicate greater degrees of material deprivation (Fry et al., 2017).

Individual characteristics. We gathered characteristics including, socio-demographic variables, comorbidity, and variables that have been previously shown to be relevant for brain volume and neighborhood contextual factors, such as age, sex or ICV (Ritchie et al., 2018; Kaczkurkin et al., 2019; Ruigrok et al., 2014; Fjell et al., 2013; Narvacan et al., 2017; Finlay et al., 2001). *Age* was calculated as the difference between the date of the interview and a participant's date of birth and scaled in years. *Sex* was a dichotomous variable (0 = female; 1 = male) and, following standard procedure, recoded to (0.5 = female; 0.5 = male). Centering the variable around zero simplifies the interpretation of the intercept in any regression models (Cohen et al., 2013). *Education.* Highest education was a six-factor variable (coded as other qualification; O levels (Ordinary Levels-part of the General Certificate of Education), General Certificate of Secondary Education (GCSE), or Certificate of Secondary Education; A levels or AS levels (Advanced Levels qualifications after completing their GCSEs or O levels); National Vocational Qualification, Higher National Diploma, Higher National Certificate, or equivalent; college or university degree). *Income.* Average total annual income before tax, measured at household-level was a four-level factor (<£18,000, £18,000–30 999, £31,000–51 999, ≥£52,000). *Stressors* were examined by asking participants whether or not they had experienced the following stressors in the last 2 years: illness, injury, bereavement, or stress. Due to the distribution, we dichotomized

the variable to 0 = no stressor and 1 = stressors. *Cardiovascular Conditions.* The cardiovascular variable captures self-reported diagnoses indicating the presence of cardiovascular conditions, including conditions such as hypertension, heart disease, and stroke. This variable was coded as a binary indicator classified as having a cardiovascular condition (1) or not (0). *Respiratory Conditions.* The respiratory variable includes self-reported diagnoses related to respiratory conditions, such as asthma, chronic obstructive pulmonary disease (COPD). This variable was also coded as a binary indicator, with participants classified as having a respiratory condition (1) or not (0). Intercorrelations can be found in the Appendix (Table A.3.1 and A.3.2). An overview of all variables can be found in Appendix (A.4.).

2.4. Data analysis

We conducted a number of analyses. Analysis was conducted using the statistical software R (R Core Team, 2016). First, we applied the Boruta feature selection algorithm to identify important features for hippocampal volume in adulthood and older age (Kursa and Rudnicki, 2010). The Boruta algorithm works as a wrapper around a Random Forest classification algorithm. This method aims to select important features by iteratively removing irrelevant attributes. It is thus useful for data with a large number of correlated variables, such as environmental variables. In the following, we will briefly describe the Boruta algorithm procedure, although a more detailed description can be found elsewhere (Kursa and Rudnicki, 2010; Rudnicki et al., 2015a). First, the dataset is extended by creating duplicates of all features included in the dataset. Second, by shuffling the values of these duplicated features, so-called shadow features are created with the goal of removing any meaningful correlations with the outcome variable. Third, a random forest algorithm is applied to the extended dataset, and Z-scores are computed. Fourth, the maximum Z-score among the shadow attributes (MZSA) is identified. Fifth, features are permanently removed from the dataset when importance is remarkably less than MZSA or kept in the dataset when importance is remarkably greater than MZSA. In other words, the importance of each feature is measured and compared against the Maximum Z-Score of Importance (MZSA) threshold. Features with a Z-score greater than this threshold are marked as important. A feature is considered "important" if its importance score is significantly higher than the importance scores of shadow features (randomly permuted versions of the original features). This significance is determined by comparing the feature's Z-score to a threshold. Sixth, shadow features are removed from the dataset. Seventh, this process is repeated until all unimportant features in the dataset are removed.

We proceeded systematically by applying the Boruta algorithm to identify key features associated with hippocampal volume. In the initial analysis, we examined both environmental factors and individual characteristics, such as age, intracranial volume (ICV), and sex, to determine their relative importance (for full list see Appendix Table A.4.). In a second step, following previous work (Williams et al., 2021b), we repeated the procedure with the residualized hippocampal volume by age, sex, ICV. This allowed us to take into account sex, age, and the allometric relationships between hippocampus measures. The code can be found in Appendix (A.5.)

3. Results

With data from 21,049 individuals, we aimed to assess the roles of neighborhood and environmental characteristics for hippocampal brain volume over and above individual difference characteristics, such as age, sex or ICV, by using data from the UK Biobank. As expected, age was negatively correlated with hippocampal volume (r = 0.31) and positively correlated with sex (r = 0.30), but showed no large correlations with environmental characteristics (see also Table A.3.1.). Indicating that there was no strong link between how old people were and where they lived. Environmental features were mostly correlated with one another - highlighting the need for methods accounting for multicollinearity. We found that environmental factors differed in how closely they were related. As expected, both traffic intensity of the nearest road was strongly correlated with PM2.5 (r = 0.79). Similarly, green space was highly correlated with natural environment (r = 0.96) (see also Tables A.3).

3.1. Investigating associations between the physical environment and brain structure

We applied the Boruta algorithm to examine which environmental and individual factors were most strongly associated with hippocampal volume. First, we assessed the relative importance of individual characteristics, including age, sex, intracranial volume (ICV), and cardiovascular and respiratory health, alongside environmental variables such as household income, education, stress, Townsend deprivation score, gas heating, proximity to nature, greenspace, garden access, water bodies, coastal distance, nitrogen dioxide air pollution (NO_2 & NO), particulate matter, distance to major and minor roads, traffic intensity, and daytime, evening, and nighttime noise levels (for full list see Table A.4.). This analysis identified 32 features as important, with ICV, age, and sex emerging as the most influential, as expected (Fig. 1, Panel A). Environmental factors, including nitrogen dioxide air pollution (NO_2), natural environment, green space, and noise indicators, also ranked among highly in importance. Thus, following previous work (Williams et al., 2021b), we then conducted an analysis with residualized hippocampal volume as an outcome—controlling for age, sex, ICV, and MRI site in an initial regression. In this residualized model, Boruta identified 22 features as important for revisualized hippocampal volume, with air pollution indicators emerging among the most important environmental factors (Fig. 1).

Specifically, Nature, PM air pollution and Nitrogen Dioxide air pollution (NO2) were among the most important features. Interestingly, these characteristics were identified as more important than closeness to water or ocean but also individual difference characteristics such as education. As can also be seen in Fig. 1 (Panel B), indicators of noise were identified as equally important. As can be seen in Fig. 2, there seems to be a systematic relationship between NO2 and hippocampal volume for subjects at the low end of the distribution. Comparing participants in the lowest quintile of hippocampal volume (bottom 20%) to the second lowest quintile (20%–40%), participants in the lowest quintile show, on average, higher levels of NO2. The difference is 26.04 vs. 25.63 and is statistically significant ($p < .02$). For NO the effect goes in the same directions, with higher NO levels (43.09) in the bottom quintile compared to NO levels (42.23) in the second lowest quintile, again significant at the 2% level, indicating that high levels of NO2 and NO is negatively associated with hippocampal volume. The same negative relationship is present for PM2.5, where the difference of 9.95 versus 9.89 is significant at the 1% level.

The same negative relationship is present for PM2.5, where the difference of 9.95 versus 9.89 is statistically significant ($p = .0036$). Expectedly, the effects for noise showed a similar pattern (see also Appendix Table A.1. Figure A.1.).

The effect of nature and green space is shown in Fig. 3. As can be seen, lower levels of coverage with natural environment around the home address were linked to hippocampal volume primarily for participants in the lowest quintile of hippocampal volume.

Specifically, the difference is 43.95 vs. 45.30 and is significant at the 2% level, indicating that lower levels of nature are linked to smaller hippocampus volume. A similar effect is evident for green where the difference is 47.28 vs. 48.51and are statistically significant ($p < .02$).

4. Discussion

This present data analysis aimed to investigate the roles of neighborhood and environmental characteristics for hippocampal brain

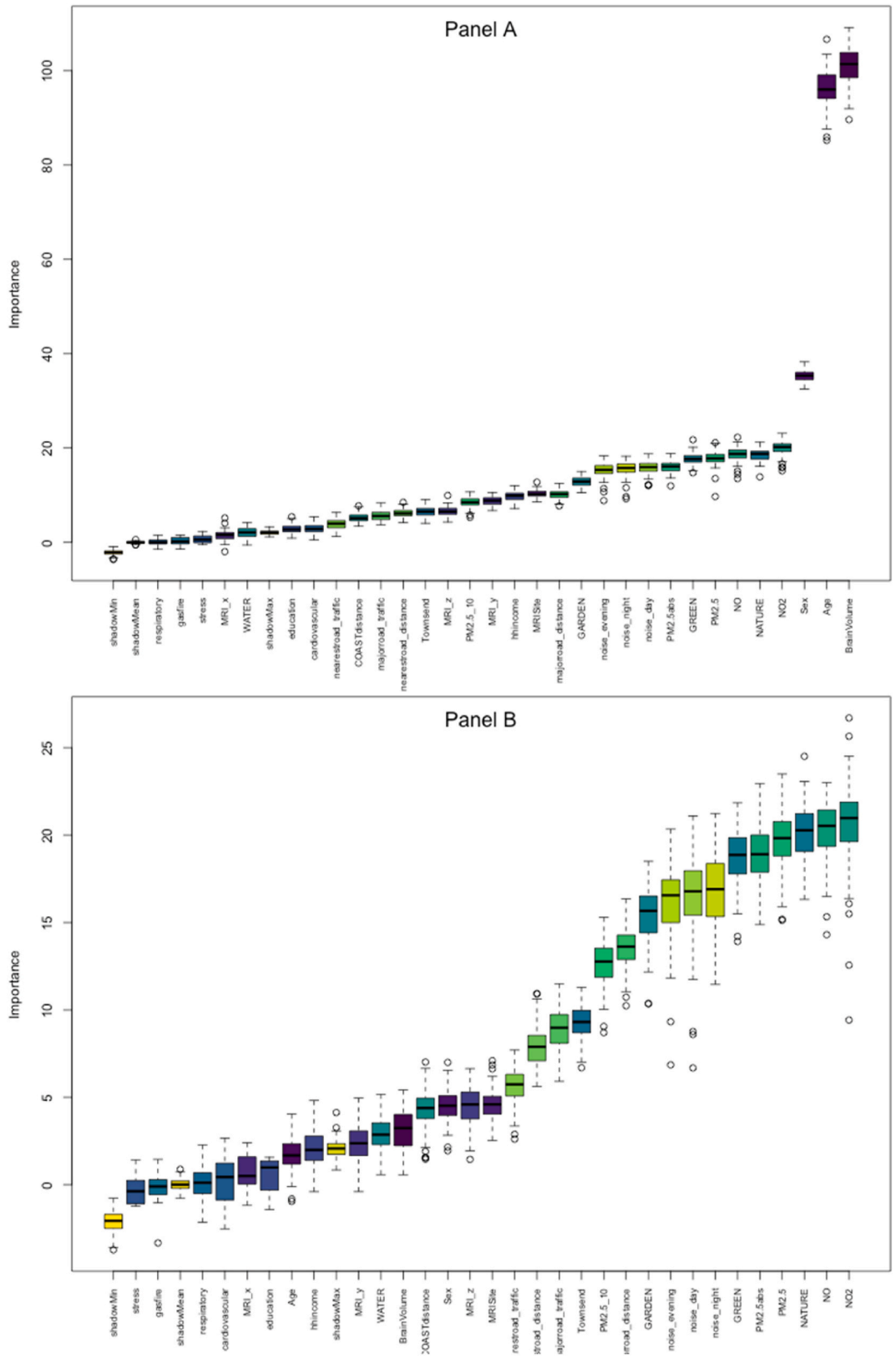
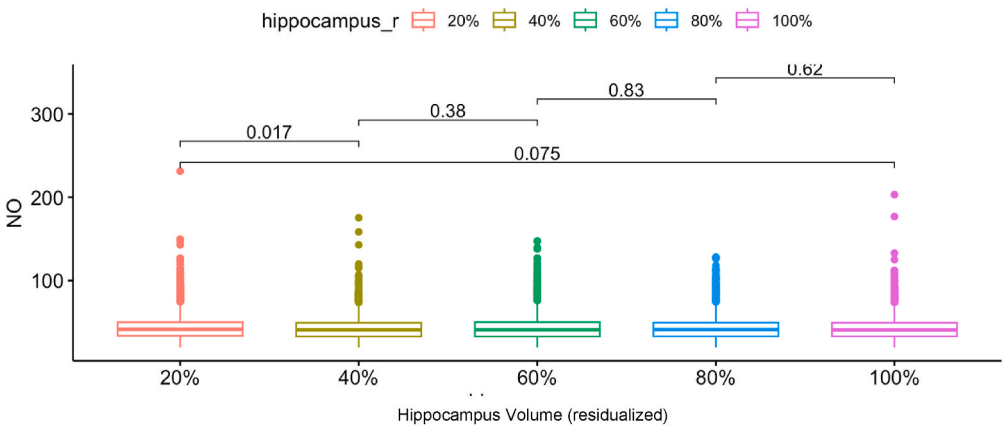
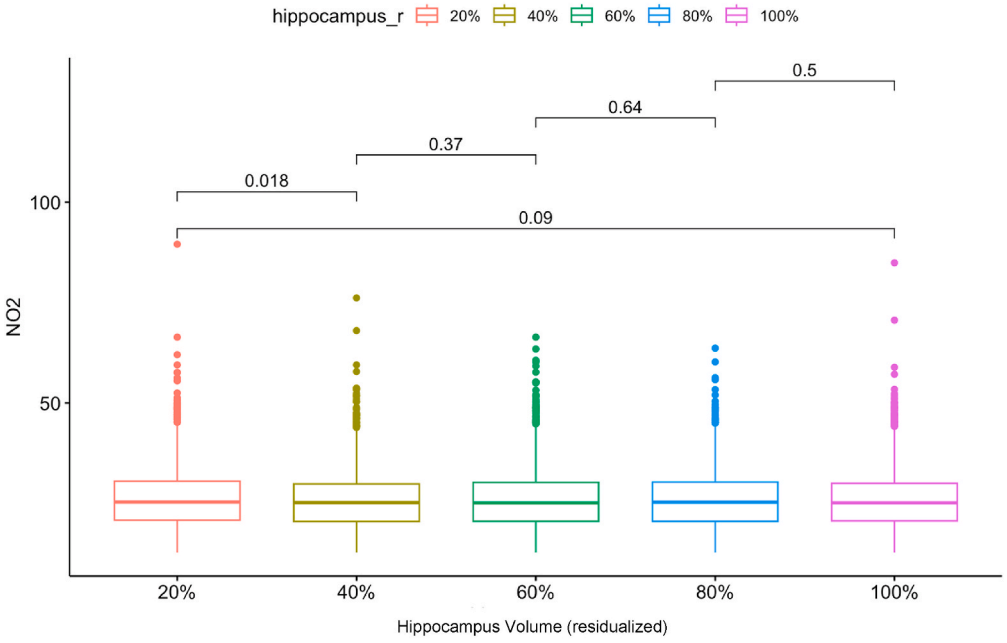
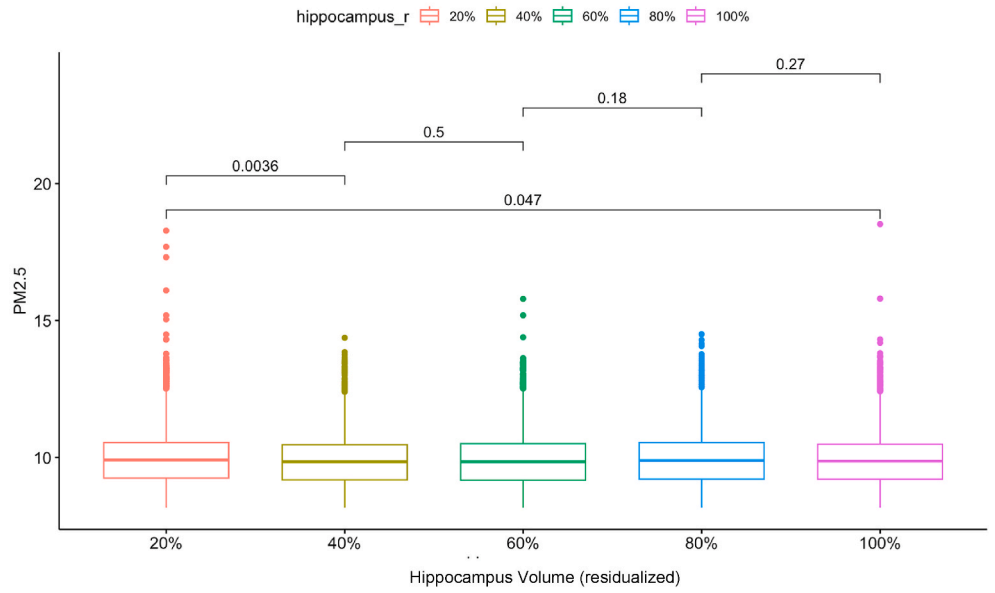


Fig. 1. Panel A - Associating Hippocampus Volume with Environmental and Individual Difference Characteristics. Panel B - Associating Residualized Hippocampus Volume (for sex, age, Brain Volume (ICV)), green indicates variables considered as important, red shows they are rejected and yellow color of box plot indicates they features are tentative. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Fig. 1. Panel A - Associating Hippocampus Volume with Environmental and Individual Difference Characteristics. Panel B - Associating Residualized Hippocampus Volume (for sex, age, Brain Volume (ICV)), green indicates variables considered as important, red shows they are rejected and yellow color of box plot indicates they features are tentative.







(caption on next page)

Fig. 2. Illustrates the relationship between air pollution variables NO2, NO, and PM2.5 with residual hippocampal volume. As can be seen, there is a systematic relationship between NO2 and hippocampal volume for subjects at the low end of the distribution.
Fig. 2. Illustrates the relationship between air pollution variables NO2, NO, and PM2.5 with residual hippocampal volume (adjusted for age, sex, and intracranial volume (ICV)). The residuals were categorized into quintiles to facilitate and enhance the clarity of our graphical representations. This categorization effectively grouped each observation into one of five distinct categories, based on its relative position within the overall distribution of residuals. As can be seen, there is a systematic relationship between NO2 and hippocampal volume for subjects at the low end of the distribution. Comparing the lowest quintile of the hippocampal volume (bottom 20%) to the second lowest quintile (20%–40%), subjects in the lowest quintile are exposed to higher levels of NO2 on average. The difference is 26.04 vs. 25.63 and are statistically significant ($p < .02$). For NO the effect goes in the same directions, with higher NO levels (43.09) in the bottom quintile compared to NO levels (42.23) in the second lowest quintile, again significant at the 2% level, indicating that high levels of NO2 and NO might have a negative impact on hippocampal volume. The same negative relationship is present for PM2.5, where the difference of 9.95 versus 9.89 is significant at the 1% level.

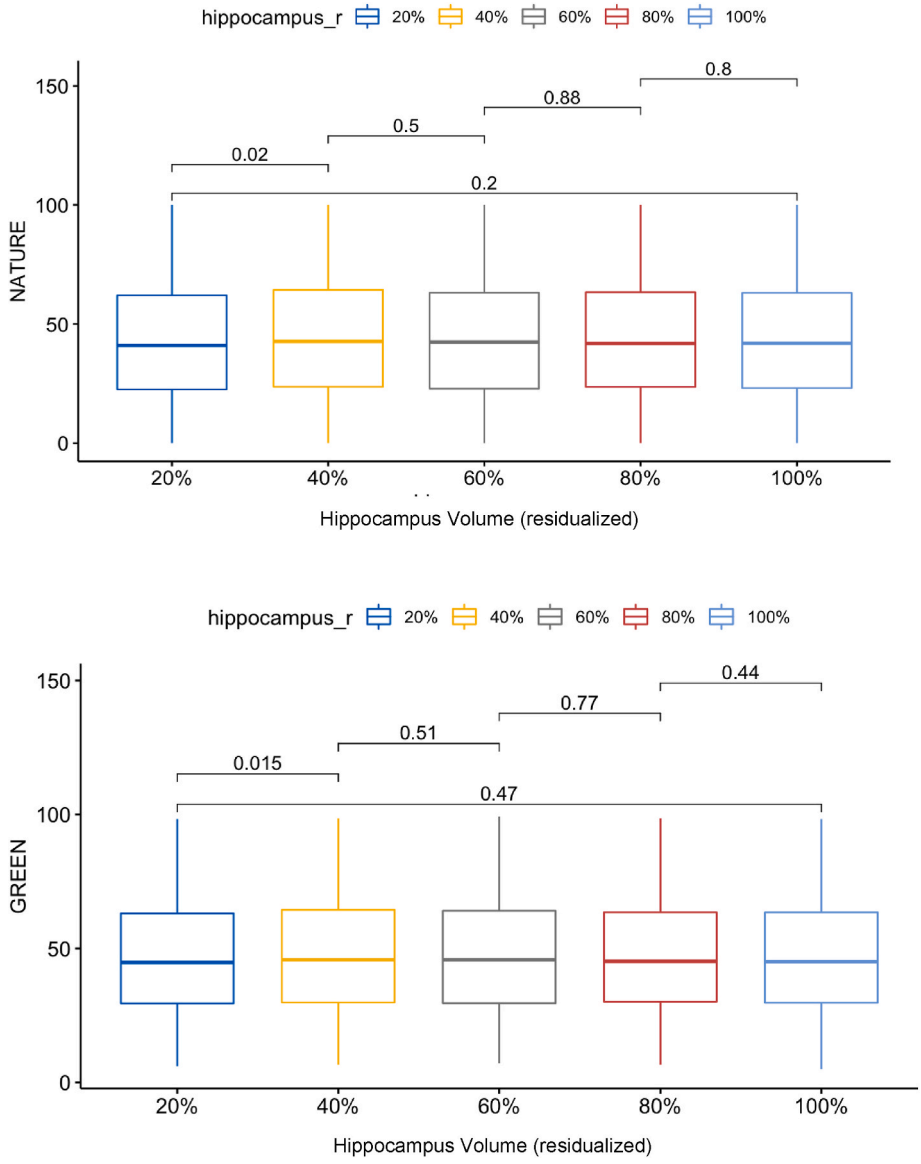


Fig. 3. Illustrates the association between natural environment, green spaces and hippocampal volume. Comparing the lowest quintile of the hippocampus volume (bottom 20%) to the second lowest quintile (20%–40%), subjects in the lowest quintile are exposed to significantly lower levels of NATURE, on average.). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
Fig. 3. Illustrates the association between natural environment, green spaces and residual hippocampal volume (adjusted for age, sex, and intracranial volume (ICV)). The residuals were categorized into quintiles to facilitate and enhance the clarity of our graphical representations. This categorization effectively grouped each observation into one of five distinct categories, based on its relative position within the overall distribution of residuals. Comparing the lowest quintile of the hippocampus volume (bottom 20%) to the second lowest quintile (20%–40%), subjects in the lowest quintile are exposed to significantly lower levels of NATURE, on average. The difference is 43.95 vs. 45.30 and is significant at the 2% level, indicating that lower levels of NATURE are negatively associated with hippocampus volume. A similar effect is evident for GREEN where the difference is 47.28 vs. 48.51and are statistically significant ($p < .02$).

volume, taking into account individual differences such as age, sex, and brain volume (ICV). Utilizing data from the UK Biobank encompassing a vast sample of 21,049 individuals, our analysis revealed several important insights into the interplay between the physical environment and brain structure.

Findings align with previous research, highlighting the relevance of individual differences on hippocampal volume. As anticipated, age demonstrated a negative correlation with hippocampal volume, consistent with the well-established notion of age-related atrophy in this brain region (Fjell et al., 2013; Narvacan et al., 2017; Finlay et al., 2001; Blinkouskaya et al., 2021). Conversely, sex exhibited a positive correlation, with women showing larger hippocampal volumes than men, consistent with previous neuroanatomical studies highlighting sex-related differences in hippocampal structure (Ritchie et al., 2018; Kaczkurkin et al., 2019). Importantly, no significant correlations emerged between age, sex, and environmental characteristics, underscoring the need to explore the distinct contributions of environmental factors to brain structure (Dadvand et al., 2018; Nicole, 2018; Xu et al., 2021; Balboni et al., 2022).

One notable observation was the significant intercorrelations among various environmental features (see also Appendix Figure A.3). This multicollinearity underscores the complex and interconnected nature of the physical environment (Kühn et al., 2017, 2021). Specifically, we found high correlations between daytime and nighttime noise pollution and the traffic intensity of the nearest road, as well as a strong association between green space and the natural environment. These interrelationships among environmental variables highlight the necessity for analytical approaches that can effectively account for multicollinearity when assessing their independent links to brain health (Kursa and Rudnicki, 2010; Kaneko, 2023).

Employing machine learning techniques, the study aimed to identify which environmental factors are associated with hippocampal volume, while accounting for individual differences. Findings demonstrated that indicators of air pollution, particularly PM air pollution and Nitrogen Dioxide air pollution (NO2), emerged as the most important factors for hippocampal volume. These results corroborate previous research and, for example, emphasize the detrimental impact of air pollution on brain health (Balboni et al., 2022). Importantly, air pollution appeared to outweigh the effects of proximity to water or individual difference characteristics, such as education level. Air pollution is a pervasive environmental problem affecting both urban and rural populations alike. Unlike the more localized benefits of living near water, air pollution impacts a broad spectrum of the population, making its overall health impact more substantial and pressing (Peeples, 2020; Dockery et al., 1993; Anderson et al., 2012; Ehsanifar et al., 2022). It is possible that the continuous exposure to air pollutants means that individuals are affected on a daily basis. Interestingly, green space also emerged as an important factor above and beyond these air pollution measures. This is in line with previous research highlighting the salutogenic effects of green spaces coexisting with the adverse effect of air pollution. This is worth noting that these factors do not seem to cancel each other out but rather appear to independently contribute to the overall effect (Bloemsma et al., 2019).

Our analysis also highlighted the significance of noise as an important factor for hippocampal volume. This observation is consistent with a growing body of research suggesting that noise pollution can have adverse effects on cognitive function and brain structure (Peeples, 2020; Dockery et al., 1993; Ehsanifar et al., 2022; Zundel et al., 2022; Attademo and Bernardini, 2017; Dominski et al., 2021). It is noteworthy that the noise was found to be as important as air pollution in their association with hippocampal volume, underscoring the multifaceted nature of environmental determinants.

Examining the relationships more closely, our results revealed systematic patterns. For instance, there appeared to be a negative association between NO2 and hippocampal volume, particularly among individuals in the lowest quintile of hippocampal volume.

Similarly, PM2.5 exhibited a significant negative association with hippocampal volume, aligning with previous work that highlights the adverse impact of fine particulate matter on brain health.

Moreover, the importance of natural environment coverage and green space coverage for hippocampal volume was more pronounced among individuals in the lowest quintile of hippocampal volume. This might imply that individuals in the lowest quartile may be particularly susceptible and at greater risk experiencing the potential negative effects of environmental conditions.

4.1. Limitations

We note several limitations of our study. First, our results found in data from the UK Biobank are based on a healthy, national UK sample of a specific age group, findings thus need to be corroborated, and replicated in other samples (Williams et al., 2021b). Future work should systematically test whether and how our results generalize to less positively selected and more diverse segments of the population, such as individuals with lower education levels, poorer health, or more disadvantaged backgrounds. This study, like other UK Biobank studies, may be limited by the strong selection bias toward healthier, more educated individuals, as only 5.5% of the initially invited participants enrolled. This bias could impact the representativeness of our findings, especially given that our results showed pronounced effects in the lowest quintile of hippocampal volume. To account for this we conducted follow-up analyses with a sample excluding participants with substance abuse disorder, psychiatric disorder and head trauma (n = 18,518). Results remained stable. Nevertheless, given the structured nature of the UK Biobank's data collection, which predominantly includes participants with stable residential addresses, it is unlikely that a significant number of homeless individuals are well represented in the dataset. This may further limit the generalizability of our findings to populations with stable residential addresses.

It is reasonable that progressive processes associated with approaching death (e.g., deteriorating health) are correlated with notable changes in brain structure among older adults (Blinkouskaya et al., 2021). While we have included old individuals in our sample, few people older than age 80 years with more severe functional or cognitive limitations were able to participate in the UK Biobank, leaving many questions open about the effect of the environment in very old age and towards the end of life. Finally, we acknowledge a number of measurement issues of the study design which might have implications for conceptual interpretation of our findings. Using a cross-sectional study design like the UK Biobank data used here, does not enable us to draw temporal, let alone causal, inferences about how environmental factors shape brain health in adulthood and old age. Relatedly, in large-scale studies relying on secondary data, such as the UK Biobank, perfect alignment of assessments is inherently challenging due to logistical and methodological constraints. Consequently, environmental measures and brain scans were not collected simultaneously, which limits our ability to capture the immediate structural changes in the brain that might correspond to specific environmental exposures. Additionally, the cross-sectional design restricts our ability to draw temporal or causal inferences about how these environmental factors might shape brain health across adulthood and into older age. In interpreting the results of feature selection methods applied to a large set of environmental factors affecting brain structure, we make the implicit assumption that accumulating environmental influences such as pollution do shape brain structure over time. However, we cannot rule out the possibility that differences in brain structure affect where and how people choose to live (selective migration) (Norman et al., 2005).

We note that this study uses Freesurfer for hippocampal segmentation, in contrast to the FSL algorithms employed by others (Nobis et al., 2019), which may lead to differences in volume measurements across studies. Differences in hippocampal volume measurements may arise from segmentation software variation.

Although Boruta feature selection is a powerful technique for identifying relevant variables for an outcome, it does come with limitations (Anand et al., 2021; Rudnicki et al., 2015b). While Boruta allows to determine which predictors are robustly associated with the outcome, it does not fully resolve the issue of shared variance or provide a definitive decomposition of effects across predictors.

Additionally, Boruta does not provide insights into the direction or magnitude of the relationships between selected features and the outcome variable, in our case hippocampal volume. We therefore conducted a follow up analysis with conventional linear regression modeling the statistical software R with the revisualized hippocampus volume as the criterion (see also Appendix) (Cohen et al., 2013; R Development Core Team, 2017). The baseline model (Table A.2.1) included foundational demographic variables, establishing a reference for understanding core associations with hippocampal volume. We then ran a second regression (Table A.2.2.) using residualized predictors to capture unique effects beyond the baseline, incorporating intermediate covariates like brain volume and demographic controls. Finally, the full model (Table A.2.3.) included a comprehensive set of residualized predictors, adding environmental factors. This multi-layered approach allowed us to build on Boruta's selection.

Regression analysis revealed that, while biological factors like intracranial volume (ICV), age, and sex remained the strongest predictors of hippocampal volume, key environmental variables—previously identified as important by the Boruta feature selection—also contributed to the model. Specifically, household income, noise exposure, and air pollution (e.g., NO_2, PM2.5) showed associations with hippocampal volume, albeit with smaller effect sizes than biological predictors. This suggests that environmental factors are indeed relevant for hippocampal structure, supporting the idea that exposure to socio-environmental conditions, even at modest levels, is associated with brain health when also accounting for biological and demographic factors. This finding highlights the potential associations between environmental exposures and hippocampal volume, suggesting they may play a role alongside biological determinants. However, we want to note that Boruta is designed to identify and confirm the most relevant features by iteratively testing the importance of each variable, while regression aims to model relationships between independent and dependent variables, in our case hippocampus volume and environmental variables. The primary goal of our study is feature selection using Boruta, which does not focus on relationship estimation. Due to the highly interdependent nature of these environmental variables, conventional regression analysis is limited, as it assumes independent contributions of each predictor, which does not reflect the overlapping and potentially synergistic effect of environmental exposures on the hippocampus. Instead, we utilized Boruta, a feature selection method based on recursive feature elimination, which is better suited for identifying the most relevant predictors in the presence of multicollinearity and non-linear interactions. This approach allowed us to rank the importance of each environmental factor, rather than forcing them into a linear model that might obscure their true impact. Our findings, therefore, emphasize the strengths of Boruta for uncovering meaningful environmental predictors of hippocampal volume, with regression analysis included only as a supplementary verification step.

We also cannot draw any conclusions about the process by which individuals interact with their living environment. For example, adults with larger hippocampal volume or of a certain age or socioeconomic status might intentionally select themselves into environments which affect their longer-term (brain) development. To better understand the underlying mechanisms of how environmental factors are associated with hippocampal volume, more mechanism-oriented and longitudinal research is needed.

5. Conclusion

In the present study, we examined the association between hippocampal brain volume and a large number of environmental characteristics to examine the effect of the living environment on the human brain. Using large scale data from the UK Biobank we applied the Boruta algorithm, a wrapper method built around the Random Forest classification algorithm, in order to identify all the important environmental features for individual differences in hippocampal volume, over and above the well-known effects of chronological age, sex, education, or income. These results emphasize the need for comprehensive strategies to mitigate environmental factors that may compromise brain health, especially in urban settings where exposure to pollutants is prevalent. Understanding these relationships is crucial in the context of the global urbanization trend, where an increasing proportion of the population resides in urban areas. By providing a better understanding of the relationship between the living environment and the brain while also accounting for individual difference factors, we hope to inform the designing of physical environments in ways that will optimize well-being and cognitive functioning as well as human mental and physical health.

CRediT authorship contribution statement

Johanna Drewelies: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Angela Fiedler:** Writing – review & editing, Validation, Methodology, Formal analysis, Data curation. **Timothy R. Brick:** Writing – review & editing, Validation, Methodology, Formal analysis. **Simone Kühn:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Table A.1
Nature and Green spaces, and air pollution variables (NO2 NO, PM2.5): relationship with residual hippocampus volume

Residual Hippocampus Volume	Nature	Green	PM2.5 abs	NO2	NO
20%	0.015*	0.021*	0.008*	0.020*	0.010*
40%	0.011*	0.019*	0.019*	0.000*	0.005*
60%	0.016	0.013	0.025	0.008	0.001
80%	0.004	0.011	0.004	0.002	0.008
100%	0.014	0.005	0.004	0.015	0.008

Note. *$p < .001$.

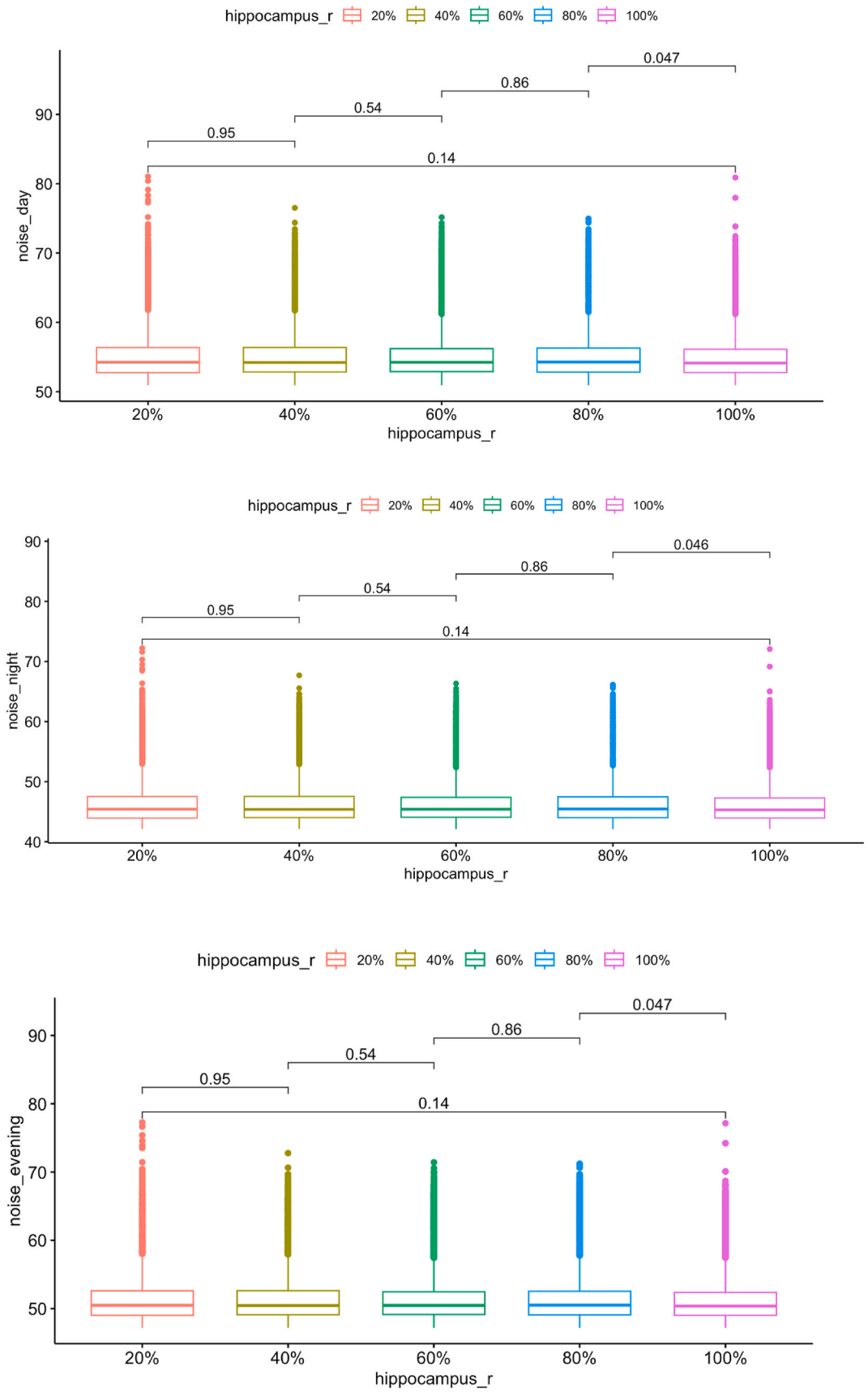


Fig. A.1. Illustrates the association between noise at day, evening, and night and hippocampal volume (residualized = hippocampus_r). Comparing the lowest quintile of the hippocampus volume (bottom 20%) to the highest quintile (100%), subjects in the lowest quintile are exposed to significantly higher levels of Noise, on average.

Table A.2.1
Regression results using Hippocampus(log) as the criterion; baseline regression

Predictor	b^a	*b*	sr^2	sr^2
		95% CI		95% CI
		[LL, UL]		[LL, UL]
(Intercept)	0.13**	[0.11, 0.15]		
ICV	0.49**	[0.47, 0.51]	0.09	[0.08, 0.09]
Age	0.34**	[-0.35, 0.33]	0.07	[0.07, 0.08]
Sex	0.12**	[0.08, 0.15]	0.00	[0.00, 0.00]
Age^2	0.10**	[-0.11, 0.09]	0.01	[0.01, 0.01]
MRI Site Reading	0.07**	[-0.10, 0.04]	0.00	[0.00, 0.00]
MRI Site Newcastle	0.01	[-0.04, 0.01]	0.00	[-0.00, 0.00]
ICV: Age	0.02**	[-0.03, 0.01]	0.00	[-0.00, 0.00]
ICV: Sex	0.05**	[-0.09, 0.02]	0.00	[-0.00, 0.00]
Age: Sex	0.05**	[-0.08, 0.02]	0.00	[-0.00, 0.00]
ICV: Age^2	0.01	[-0.02, 0.01]	0.00	[-0.00, 0.00]
Sex: Age^2	0.02	[-0.00, 0.05]	0.00	[-0.00, 0.00]
ICV: Age: Sex	0.05**	[0.02, 0.07]	0.00	[-0.00, 0.00]
ICV: Sex: Age^2	0.01	[-0.02, 0.03]	0.00	[-0.00, 0.00]
Fit	$R^2 = 0.379$**			
	95% CI[0.37,0.39]			

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
[a] = Adjustment for multiple comparisons. Bonferroni. *; **p = sig after adjustment.

Table A.2.2
Regression results using Hippocampus(log) as the criterion (residual 1)

Predictor	b[a]	b	sr2	sr2
		95% CI		95% CI
		[LL, UL]		[LL, UL]
(Intercept)	0.13**	[0.11, 0.15]		
ICV	0.48**	[0.47, 0.50]	0.16	[0.15, 0.16]
Age	0.34**	[-0.35, 0.32]	0.07	[0.07, 0.08]
Sex	0.14**	[0.11, 0.16]	0.00	[0.00, 0.00]
Age^2	0.10**	[-0.11, 0.09]	0.01	[0.01, 0.01]
MRI Site Reading	0.07**	[-0.10, 0.04]	0.00	[0.00, 0.00]
MRI Site Newcastle	0.01	[-0.04, 0.01]	0.00	[-0.00, 0.00]
ICV: Age	0.02**	[-0.03, 0.01]	0.00	[-0.00, 0.00]
ICV: Sex	0.05**	[-0.07, 0.02]	0.00	[-0.00, 0.00]
Age: Sex	0.06**	[-0.08, 0.03]	0.00	[0.00, 0.00]
ICV: Age: Sex	0.04**	[0.02, 0.07]	0.00	[-0.00, 0.00]
Fit	$R^2 = 0.379$**			
	95% CI[0.37,0.39]			

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. a = Adjustment for multiple comparisons. Bonferroni. *; **p = sig after adjustment.

Table A.2.3
Regression results using Hippocampus(log) as the criterion; full set of predictors (residual 2)

Predictor	b^a	*b*	sr^2	sr^2
		95% CI		95% CI
		[LL, UL]		[LL, UL]
(Intercept)	0.09**	[0.05, 0.14]		
ICV	0.48**	[0.47, 0.49]	0.15	[0.14, 0.16]
Age	0.34**	[-0.35, 0.32]	0.07	[0.06, 0.07]
Sex	0.13**	[0.11, 0.16]	0.00	[0.00, 0.00]
Age^2	0.10**	[-0.11, 0.09]	0.01	[0.01, 0.01]
MRI Site Reading	0.08**	[-0.12, 0.04]	0.00	[0.00, 0.00]
MRI Site Newcastle	0.03	[-0.06, 0.01]	0.00	[-0.00, 0.00]
household income (low)	0.04*	[0.00, 0.08]	0.00	[-0.00, 0.00]
household income (average)	0.05**	[0.01, 0.09]	0.00	[-0.00, 0.00]
household income (high)	0.04*	[0.00, 0.09]	0.00	[-0.00, 0.00]
household income (very high)	0.05	[-0.00, 0.11]	0.00	[-0.00, 0.00]
MRI X	0.00	[-0.01, 0.01]	0.00	[-0.00, 0.00]
MRI Y	0.00	[-0.01, 0.02]	0.00	[-0.00, 0.00]
MRI Z	0.02**	[-0.03, 0.01]	0.00	[-0.00, 0.00]
Townsend	0.01*	[-0.02, 0.00]	0.00	[-0.00, 0.00]
Education A levels	0.01	[-0.04, 0.03]	0.00	[-0.00, 0.00]

(*continued on next page*)

Table A.2.3 (*continued*)

Predictor	*b*[a]	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]
Education O Levels	0.02	[-0.05, 0.01]	0.00	[-0.00, 0.00]
Education CSE	0.02	[-0.07, 0.04]	0.00	[-0.00, 0.00]
Education NVQ/HND/HNC/equivalent	0.01	[-0.04, 0.05]	0.00	[-0.00, 0.00]
Education other qualification	0.03	[-0.08, 0.02]	0.00	[-0.00, 0.00]
Gas fire	0.01	[-0.01, 0.04]	0.00	[-0.00, 0.00]
Stressors	0.01	[-0.03, 0.01]	0.00	[-0.00, 0.00]
ICV: Age	0.02**	[-0.03, 0.01]	0.00	[-0.00, 0.00]
ICV: Sex	0.04**	[-0.07, 0.02]	0.00	[-0.00, 0.00]
Age: Sex	0.06**	[-0.09, 0.03]	0.00	[0.00, 0.00]
ICV: Age: Sex	0.04**	[0.02, 0.07]	0.00	[-0.00, 0.00]
Fit	$R^2 = 0.380$** 95% CI[0.37,0.39]			

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

[a] = Adjustment for multiple comparisons. Bonferroni. *; **p = sig after adjustment.

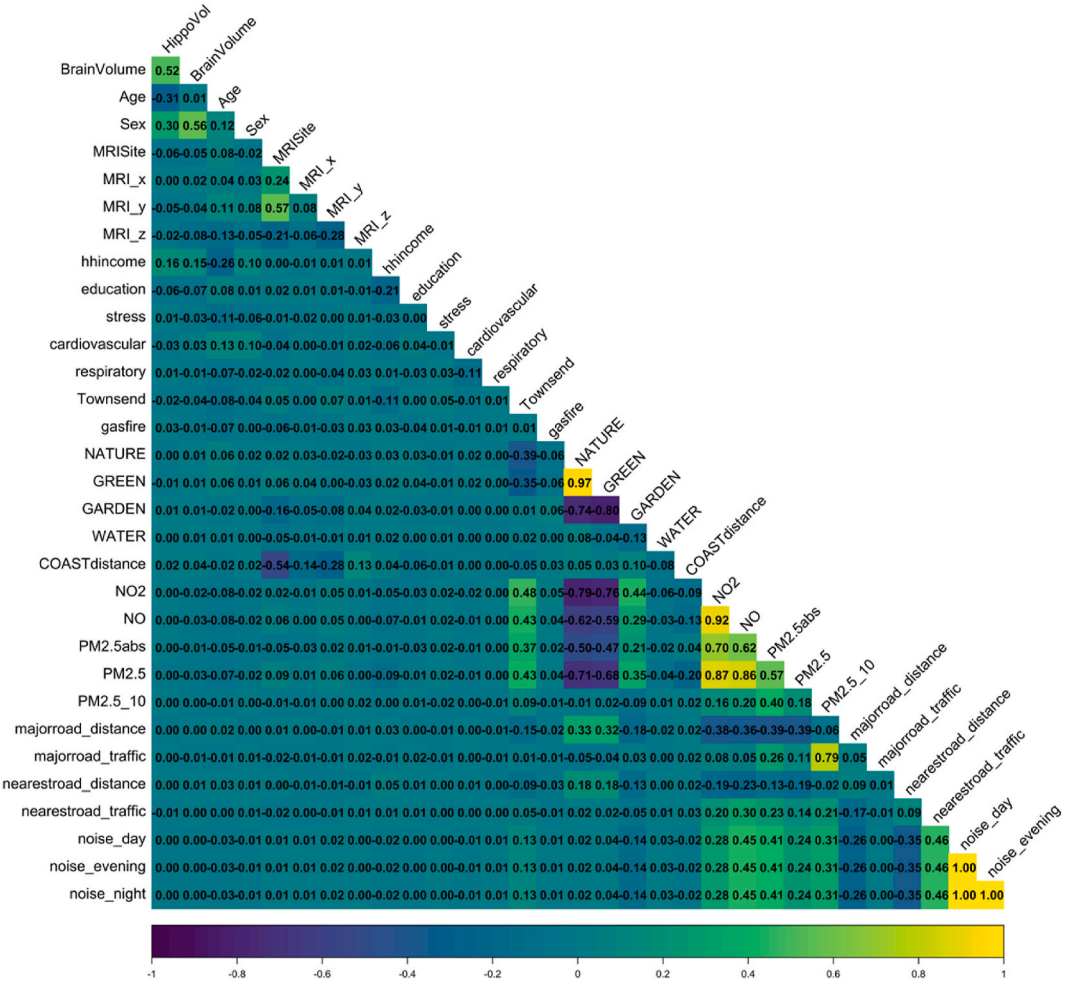


Fig. A 3.1. Figure displays the correlation between variables under study, with stronger correlations indicated by color intensity.

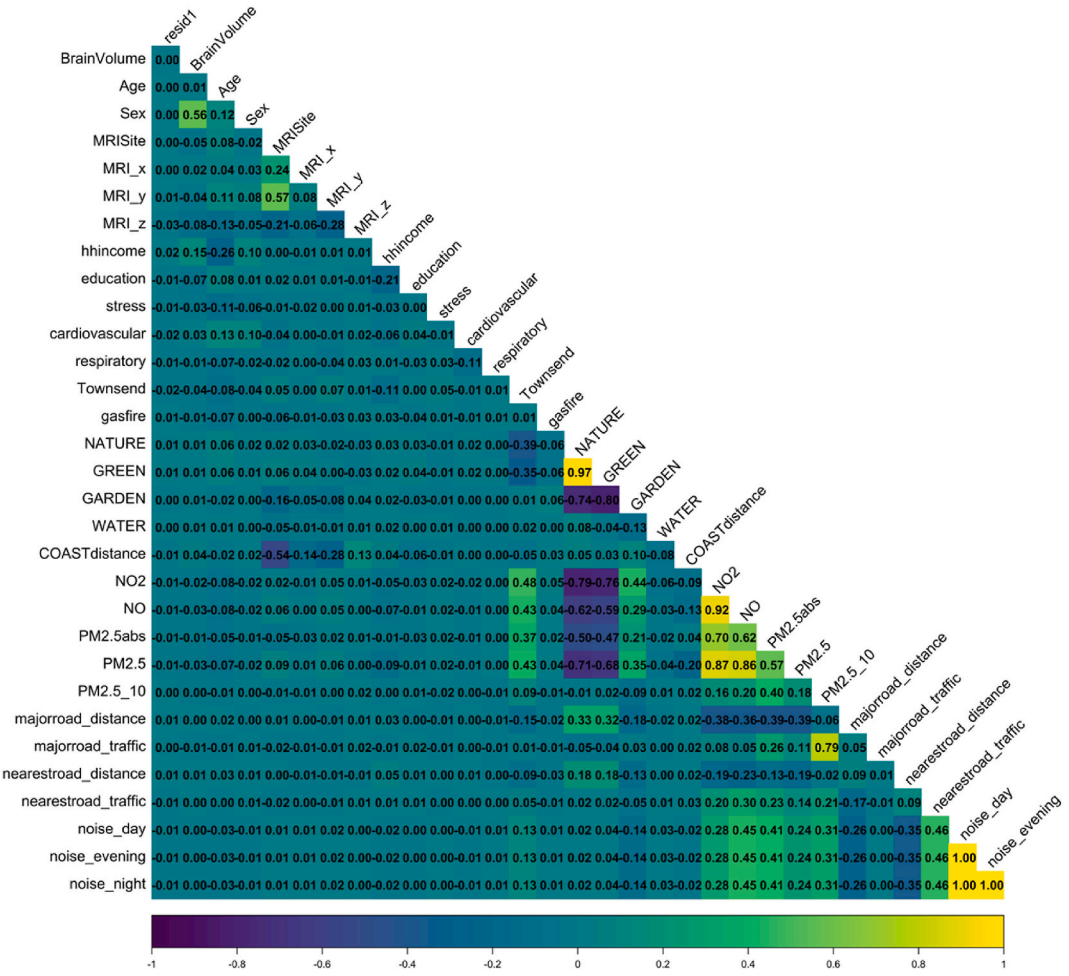


Fig. A 3.2. Figure displays the correlation between variables under study, and the residualized hippocampus variable.

A.4

Codebook for UK Biobank Variables under StudyThis codebook provides a comprehensive overview of variables from the UK Biobank dataset. Each variable is documented with its name, label, description, field ID, URL link, and any processing steps.

Variable Name	Variable Label	Description	UK Biobank Field ID	Link	Processing Step
age	age_21003_2	Participant's age at assessment	21003	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21003	
sex	sex_31_2	Participant's sex	31	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=31	Set to 0.5 (female) and 0.5 (male)
MRI Site	site_54_2	MRI assessment site location	54	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=54	
HippoVol	LeftHippocampus + RightHippocampus	Total hippocampal volume	25020/ 25019	https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25019 https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25020	Log transformed
logBrain	total_brain_volume	Log-transformed total brain volume	110	https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=110	Log transformed
MRI X Position	MRI_Xposition_25756_2	MRI X position coordinate	25756	http://biobank.ctsu.ox.ac.uk/cryst	

(*continued on next page*)

A.4 (*continued*)

Variable Name	Variable Label	Description	UK Biobank Field ID	Link	Processing Step
				al/field.cgi?id=25756	
MRI Y Position	MRI_Yposition_25757_2	MRI Y position coordinate	25757	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25757	
MRI Z Position	MRI_Zposition_25758_2	MRI Z position coordinate	25758	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=25758	
education	education_6138_2	Education level at time of MRI scan	6138	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6138	Converted to factor with labels
income	household_income_before_tax_738_2	Household income before tax	738	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=738	Converted to factor with income levels
stressors	stressors_2years_6145_2	Major stressors in the past 2 years	6145	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6145	Dichotomized: 1 if $\geq$ 1 stressor, 0 otherwise
gas_cooking	gas_fuel_cooking_heating_6139_2	Use of gas fuel for cooking or heating	6139	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6139	Dichotomized: 1 if any gas or solid fuel, 0 otherwise
GREEN	greenspace_1000m_24500	Amount of greenspace within 1000 m of residence	24500	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24500	
GARDEN	domesticgarden_1000m_24501	Amount of domestic garden space within 1000 m	24501	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24501	
NATURE	naturalenv_1000m_24506	Amount of natural environment within 1000 m	24506	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24506	
WATER	water_1000m_24502	Amount of water features within 1000 m	24502	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24502	
COASTdistance	distance_coast_24508	Distance from participant's residence to the coast	24508	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24508	Expected sign inversion for analysis
NO2	NO2_2010_24003	Nitrogen Dioxide levels in 2010	24003	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24003	
NO	NO_2010_24004	Nitric Oxide levels in 2010	24004	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24004	
PM2.5 abs	PM2.5_abs_2010_24007	PM2.5 absorbance levels in 2010	24007	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24007	
PM2.5	PM2.5_2010_24006	PM2.5 particulate matter levels in 2010	24006	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24006	
PM2.5_10	PM2.5_10_2010_24008	PM2.5-10 particulate matter levels in 2010	24008	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24008	
majorroad_distance	invdist_nearest_major_road_24012	Distance to the nearest major road (inverse transformed)	24012	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24012	Redefined as distance (inverse transformation)
majorroad_traffic	trafficintensity_nearest_major_road_24011	Traffic intensity of nearest major road	24011	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=24011	
cardiovascular	illness_20002_2	Self-reported cardiovascular diagnoses (e.g., hypertension, myocardial infarction, stroke)	20002	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002	Binary indicator: 1 if any cardiovascular diagnosis (codes 1065, 1075, 1081), 0 otherwise

(*continued on next page*)

A.4 (*continued*)

Variable Name	Variable Label	Description	UK Biobank Field ID	Link	Processing Step
respiratory	illness_20002_2	Self-reported respiratory diagnoses (e.g., asthma, chronic bronchitis)	20002	http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002	Binary indicator: 1 if any respiratory diagnosis (codes 1111, 1112), 0 otherwise

A.5. R Code for Feature Selection Using Boruta.

Set Working Directory setwd("setworkingdirectory")

Load Required Packages library(dplyr) # Data manipulation library(ggplot2) # Plotting library(corrplot) # Correlation plots install.packages ("Boruta") # Feature selection library(Boruta).

Load Data load("thedata.RData")

Define Dataset for Boruta Procedure vars < - c("HippoVol", "BrainVolume", "Age", "Sex", "MRISite", "MRI_x", "MRI_y", "MRI_z", "hhincome", "education", "stress", "cardiovascular", "respiratory", "Townsend", "gasfire", "NATURE", "GREEN", "GARDEN", "WATER", "COASTdistance", "NO2″, "NO", "PM2.5 abs", "PM2.5″, "PM2.5_10″, "majorroad_distance", "majorroad_traffic", "nearestroad_distance", "nearestroad_traffic", "noise_day", "noise_evening", "noise_night")

traindata < - data[vars]

Run Boruta Feature Selection set.seed(123) boruta.train < - Boruta(Hippocampus ~ ., data = traindata) print(boruta.train) ‴

Data availability

The authors do not have permission to share data, but the UK Biobank data can be requested by researchers (for further information see www.ukbiobank.ac.uk).

References

Alfaro-Almagro, F., McCarthy, P., Afyouni, S., et al., 2021. Confound modelling in UK Biobank brain imaging. Neuroimage 224. https://doi.org/10.1016/J.NEUROIMAGE.2020.117002.

Anand, N., Sehgal, R., Anand, S., Kaushik, A., 2021. Feature selection on educational data using Boruta algorithm. Int. J. Comput. Intell. Stud. 10 (1), 27. https://doi.org/10.1504/IJCISTUDIES.2021.113826.

Anderson, J.O., Thundiyil, J.G., Stolbach, A., 2012. Clearing the air: a review of the effects of particulate matter air pollution on human health. J. Med. Toxicol. 8 (2), 166–175. https://doi.org/10.1007/S13181-011-0203-1/TABLES/5.

Attademo, L., Bernardini, F., 2017. Air pollution and urbanicity: common risk factors for dementia and schizophrenia? Lancet Planet. Health 1 (3), e90–e91. https://doi.org/10.1016/S2542-5196(17)30042-6.

Balboni, E., Filippini, T., Crous-Bou, M., Guxens, M., Erickson, L.D., Vinceti, M., 2022. The association between air pollutants and hippocampal volume from magnetic resonance imaging: a systematic review and meta-analysis. Environ. Res. 204, 111976. https://doi.org/10.1016/J.ENVRES.2021.111976.

Blinkouskaya, Y., Caçoilo, A., Gollamudi, T., Jalalian, S., Weickenmeier, J., 2021. Brain aging mechanisms with mechanical manifestations. Mech. Ageing Dev. 200, 111575. https://doi.org/10.1016/J.MAD.2021.111575.

Bloemsma, L.D., Gehring, U., Klompmaker, J.O., et al., 2019. Green space, air pollution, traffic noise and cardiometabolic health in adolescents: the PIAMA birth cohort. Environ. Int. 131. https://doi.org/10.1016/J.ENVINT.2019.104991.

Bonnici, H.M., Chadwick, M.J., Maguire, E.A., 2013. Representations of recent and remote autobiographical memories in hippocampal subfields. Hippocampus 23 (10), 849–854. https://doi.org/10.1002/HIPO.22155.

Carlesimo, G.A., Cherubini, A., Caltagirone, C., Spalletta, G., 2010. Hippocampal mean diffusivity and memory in healthy elderly individuals. Neurology 74 (3), 194–200. https://doi.org/10.1212/WNL.0B013E3181CB3E39.

Chersi, F., Burgess, N., 2015. The cognitive architecture of spatial navigation: hippocampal and striatal contributions. Neuron 88 (1), 64–77. https://doi.org/10.1016/J.NEURON.2015.09.021.

Chun, M.M., Turk-Browne, N.B., 2007. Interactions between attention and memory. Curr. Opin. Neurobiol. 17 (2), 177–184. https://doi.org/10.1016/J.CONB.2007.03.005.

Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. In: Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, third ed. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, pp. 1–704. https://doi.org/10.4324/9780203774441/APPLIED-MULTIPLE-REGRESSION-CORRELATION-ANALYSIS-BEHAVIORAL-SCIENCES-JACOB-COHEN-PATRICIA-COHEN-STEPHEN-WEST-LEONA-AIKEN. Third Edition. Published online January 1.

Collins, R., 2012. What makes UK Biobank special? Lancet 379 (9822), 1173–1174. https://doi.org/10.1016/S0140-6736(12)60404-8.

Dadvand, P., Pujol, J., Maciá, D., et al., 2018. The association between lifelong greenspace exposure and 3-dimensional brain magnetic resonance imaging in barcelona schoolchildren. Environ. Health Perspect. 126 (2). https://doi.org/10.1289/EHP1876.

de Hoogh, K., Gulliver, J., Donkelaar, A van, et al., 2016. Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10. https://doi.org/10.1016/J.ENVRES.2016.07.005.

Dockery, D.W., Pope, C.A., Xu, X., et al., 1993. An association between air pollution and mortality in six U.S. cities. NEJM 329 (24), 1753–1759. https://doi.org/10.1056/nejm199312093292401.

Dominski, F.H., Lorenzetti Branco, J.H., Buonanno, G., Stabile, L., Gameiro da Silva, M., Andrade, A., 2021. Effects of air pollution on health: a mapping review of systematic reviews and meta-analyses. Environ. Res. 201. https://doi.org/10.1016/j.envres.2021.111487.

Dugger, B.N., Dickson, D.W., 2017. Pathology of neurodegenerative diseases. Cold Spring Harb Perspect Biol 9 (7), a028035. https://doi.org/10.1101/CSHPERSPECT.A028035.

Eeftens, M., Beelen, R., De Hoogh, K., et al., 2012. Development of Land Use Regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46 (20), 11195–11205. https://doi.org/10.1021/ES301948K.

Ehsanifar, M., Yavari, Z., Rafati, M., 2022. Exposure to urban air pollution particulate matter: neurobehavioral alteration and hippocampal inflammation. Environ. Sci. Pollut. Res. Int. 29 (33), 50856–50866. https://doi.org/10.1007/s11356-022-19367-9.

Eichenbaum, H., 2000. A cortical–hippocampal system for declarative memory. Nat. Rev. Neurosci. 1 (1), 41–50. https://doi.org/10.1038/35036213.

Erickson, K.I., Voss, M.W., Prakash, R.S., et al., 2011a. Exercise training increases size of hippocampus and improves memory. Proc Natl Acad Sci U S A 108 (7), 3017–3022. https://doi.org/10.1073/PNAS.1015950108/SUPPL_FILE/PNAS.201015950SI.PDF.

Erickson, K.I., Voss, M.W., Prakash, R.S., et al., 2011b. Exercise training increases size of hippocampus and improves memory. Proc Natl Acad Sci U S A 108 (7), 3017–3022. https://doi.org/10.1073/PNAS.1015950108.

Ewers, M., Walsh, C., Trojanowski, J.Q., et al., 2012. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol. Aging 33 (7), 1203–1214.e2. https://doi.org/10.1016/J.NEUROBIOLAGING.2010.10.019.

Finlay, B.L., Darlington, R.B., Nicastro, N., 2001. Developmental structure in brain evolution. Behavioral and Brain Sciences 24 (2), 263–278. https://doi.org/10.1017/S0140525X01003958.

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B., 2013. Brain changes in older adults at very low risk for alzheimer's disease. J. Neurosci. 33 (19), 8237. https://doi.org/10.1523/JNEUROSCI.5506-12.2013.

Fossati, P., 2012. Neural correlates of emotion processing: from emotional to social brain. Eur. Neuropsychopharmacol 22 (Suppl. 3). https://doi.org/10.1016/J.EURONEURO.2012.07.008. SUPPL3.

Fry, A., Littlejohns, T.J., Sudlow, C., et al., 2017. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am. J. Epidemiol. 186 (9), 1026–1034. https://doi.org/10.1093/AJE/KWX246.

Kaczkurkin, A.N., Raznahan, A., Satterthwaite, T.D., 2019. Sex differences in the developing brain: insights from multimodal neuroimaging. Neuropsychopharmacology 44 (1), 71–85. https://doi.org/10.1038/S41386-018-0111-Z.

Kaneko, H., 2023. Interpretation of machine learning models for data sets with many features using feature importance. ACS Omega 8 (25), 23218–23225. https://doi.org/10.1021/ACSOMEGA.3C03722/ASSET/IMAGES/LARGE/AO3C03722_0007.JPEG.

Kharabian, Masouleh S., Plachti, A., Hoffstaedter, F., Eickhoff, S., Genon, S., 2020. Characterizing the gradients of structural covariance in the human hippocampus. Neuroimage 218. https://doi.org/10.1016/J.NEUROIMAGE.2020.116972.
Kobayashi, K.S., Matsuo, N., 2023. Persistent representation of the environment in the hippocampus. Cell Rep. 42 (1). https://doi.org/10.1016/J.CELREP.2022.111989.
Kühn, S., Düzel, S., Eibich, P., et al., 2017. In search of features that constitute an "enriched environment" in humans: associations between geographical properties and brain structure. Sci. Rep. 7 (1), 1–8. https://doi.org/10.1038/s41598-017-12046-7, 2017 7:1.
Kühn, S., Düzel, S., Mascherek, A., et al., 2021. Urban Green Is More than the Absence of City: Structural and Functional Neural Basis of Urbanicity and Green Space in the Neighbourhood of Older Adults. https://doi.org/10.1016/j.landurbplan.2021.104196. Published online.
Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the boruta package. J Stat Softw 36 (11), 1–13. https://doi.org/10.18637/JSS.V036.I11.
Lagali, P.S., Corcoran, C.P., Picketts, D.J., 2010. Hippocampus development and function: role of epigenetic factors and implications for cognitive disease. Clin. Genet. 78 (4), 321–333. https://doi.org/10.1111/J.1399-0004.2010.01503.X.
Li, C., Gao, D., Cai, Y.S., et al., 2023. Relationships of residential distance to major traffic roads with dementia incidence and brain structure measures: mediation role of air pollution. Health Data Science 3. https://doi.org/10.34133/HDS.0091/SUPPL_FILE/HDS.0091.F1.DOCX.
Lieberman, J.A., Girgis, R.R., Brucato, G., et al., 2018. Hippocampal dysfunction in the pathophysiology of schizophrenia: a selective review and hypothesis for early detection and intervention. Mol. Psychiatr. 23 (8), 1764–1772. https://doi.org/10.1038/mp.2017.249, 2018 23:8.
MacQueen, G., Frodl, T., 2010a. The hippocampus in major depression: evidence for the convergence of the bench and bedside in psychiatric research? Mol. Psychiatr. 16 (3), 252–264. https://doi.org/10.1038/mp.2010.80, 2011 16:3.
MacQueen, G., Frodl, T., 2010b. The hippocampus in major depression: evidence for the convergence of the bench and bedside in psychiatric research? Mol. Psychiatr. 16 (3), 252–264. https://doi.org/10.1038/mp.2010.80, 2011 16:3.
Mamouei, M., Zhu, Y., Nazarzadeh, M., et al., 2022. Investigating the association of environmental exposures and all-cause mortality in the UK Biobank using sparse principal component analysis. Sci. Rep. 12 (1). https://doi.org/10.1038/S41598-022-13362-3.
Mascherek, A., Weber, S., Riebandt, K., et al., 2022. On the relation between a green and bright window view and length of hospital stay in affective disorders. Eur Psychiatry 65 (1). https://doi.org/10.1192/J.EURPSY.2022.9.
McDonald, R.I., Marcotullio, P.J., Güneralp, B., 2013. Urbanization and global trends in biodiversity and ecosystem services. Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities: A Global Assessment 31–52. https://doi.org/10.1007/978-94-007-7088-1_3/FIGURES/9. Published online January 1.
McEwen, B.S., Gianaros, P.J., 2011. Stress- and Allostasis-Induced Brain Plasticity 62, 431–445. https://doi.org/10.1146/ANNUREV-MED-052209-100430.
Morley, D.W., De Hoogh, K., Fecht, D., et al., 2015. International scale implementation of the CNOSSOS-EU road traffic noise prediction model for epidemiological studies. Environmental Pollution 206, 332–341. https://doi.org/10.1016/J.ENVPOL.2015.07.031.
Narvacan, K., Treit, S., Camicioli, R., Martin, W., Beaulieu, C., 2017. Evolution of deep gray matter volume across the human lifespan. Hum. Brain Mapp. 38 (8), 3771–3790. https://doi.org/10.1002/HBM.23604.
Nations, U., of Economic, D., Affairs, S., Division, P., 2018. World Urbanization Prospects The 2018 Revision. Published online.
Nicole, W., 2018. Nature and neurodevelopment: differences in brain volume by residential exposure to greenness. Environ. Health Perspect. 126 (6). https://doi.org/10.1289/EHP3465.
Nobis, L., Manohar, S.G., Smith, S.M., et al., 2019. Hippocampal volume across age: nomograms derived from over 19,700 people in UK Biobank. Neuroimage Clin 23. https://doi.org/10.1016/J.NICL.2019.101904.
Norman, P., Boyle, P., Rees, P., 2005. Selective migration, health and deprivation: a longitudinal analysis. Soc. Sci. Med. 60 (12), 2755–2771. https://doi.org/10.1016/J.SOCSCIMED.2004.11.008.
Nußbaum, R., Lucht, S., Jockwitz, C., et al., 2020. Associations of air pollution and noise with local brain structure in a cohort of older adults. Environ. Health Perspect. 128 (6), 1–11. https://doi.org/10.1289/EHP5859.
Peeples, L., 2020. How air pollution threatens brain health. Proc Natl Acad Sci U S A. 117 (25), 13856–13860. https://doi.org/10.1073/PNAS.2008940117/ASSET/5C65EED7-A045-47BF-940F-BC939CE6E1DB/ASSETS/PNAS.2008940117.FP.PNG.
Pujol, J., Martínez-Vilavella, G., Macià, D., et al., 2016. Traffic pollution exposure is associated with altered brain connectivity in school children. Neuroimage 129, 175–184. https://doi.org/10.1016/j.neuroimage.2016.01.036.
R Core Team, 2016. Radiokhimiya: A Language and Environment for Statistical Computing. Published online.
R Development Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Published online.
Ritchie, S.J., Cox, S.R., Shen, X., et al., 2018. Sex differences in the adult human brain: evidence from 5216 UK biobank participants. Cerebr. Cortex 28 (8), 2959–2975. https://doi.org/10.1093/CERCOR/BHY109.
Rolls, E.T., 2020. Spatial coordinate transforms linking the allocentric hippocampal and egocentric parietal primate brain systems for memory, action in space, and navigation. Hippocampus 30 (4), 332–353. https://doi.org/10.1002/hipo.23171.
Rudnicki, W.R., Wrzesień, M., Paja, W., 2015a. All relevant feature selection methods and applications. Studies in Computational Intelligence 584, 11–28. https://doi.org/10.1007/978-3-662-45620-0_2/TABLES/6.
Rudnicki, W.R., Wrzesień, M., Paja, W., 2015b. All relevant feature selection methods and applications. Studies in Computational Intelligence 584, 11–28. https://doi.org/10.1007/978-3-662-45620-0_2/TABLES/6.
Ruigrok, A.N.V., Salimi-Khorshidi, G., Lai, M.C., et al., 2014. A meta-analysis of sex differences in human brain structure. Neurosci. Biobehav. Rev. 39 (100), 34–50. https://doi.org/10.1016/J.NEUBIOREV.2013.12.004.
Sarkar, C., Webster, C., Gallacher, J., 2018a. Residential greenness and prevalence of major depressive disorders: a cross-sectional, observational, associational study of 94 879 adult UK Biobank participants. Lancet Planet. Health 2 (4), e162–e173. https://doi.org/10.1016/S2542-5196(18)30051-2.
Sarkar, C., Webster, C., Gallacher, J., 2018b. Neighbourhood walkability and incidence of hypertension: findings from the study of 429,334 UK Biobank participants. Int. J. Hyg Environ. Health 221 (3), 458–468. https://doi.org/10.1016/J.IJHEH.2018.01.009.
Stella, F., Treves, A., 2011. Associative memory storage and retrieval: involvement of theta oscillations in hippocampal information processing. Neural Plast. 2011. https://doi.org/10.1155/2011/683961.
Sudlow, C., Gallacher, J., Allen, N., et al., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12 (3). https://doi.org/10.1371/JOURNAL.PMED.1001779.
Tartt, A.N., Mariani, M.B., Hen, R., Mann, J.J., Boldrini, M., 2022. Dysregulation of adult hippocampal neuroplasticity in major depression: pathogenesis and therapeutic implications. Mol Psychiatry 27 (6), 2689–2699. https://doi.org/10.1038/S41380-022-01520-Y.
Thomas, R.M., Hotsenpiller, G., Peterson, D.A., 2007. Acute psychosocial stress reduces cell survival in adult hippocampal neurogenesis without altering proliferation. J. Neurosci. 27 (11), 2734–2743. https://doi.org/10.1523/jneurosci.3849-06.2007.
Tost, H., Reichert, M., Braun, U., et al., 2019. Neural correlates of individual differences in affective benefit of real-life urban green space exposure. Nat. Neurosci. 22 (9), 1389–1393. https://doi.org/10.1038/S41593-019-0451-Y.
UK Biobank-Environmental Exposures-Metadata. Accessed November 27, 2023. http://www.escapeproject.eu/.
Vienneau, D., De Hoogh, K., Bechle, M.J., et al., 2013. Western european land use regression incorporating satellite- and ground-based measurements of NO2 and PM10. Environ. Sci. Technol. 47 (23), 13555–13564. https://doi.org/10.1021/ES403089Q/SUPPL_FILE/ES403089Q_SI_001.PDF.
Vlahov, D., Galea, S., 2002. Urbanization, urbanicity, and health. J. Urban Health 79 (1), S1–S12. https://doi.org/10.1093/jurban/79.suppl_1.s1.
Williams, C.M., Peyre, H., Toro, R., Ramus, F., 2021a. Sex differences in the brain are not reduced to differences in body size. Neurosci. Biobehav. Rev. 130, 509–511. https://doi.org/10.1016/J.NEUBIOREV.2021.09.015.
Williams, C.M., Peyre, H., Toro, R., Ramus, F., 2021b. Neuroanatomical norms in the UK Biobank: the impact of allometric scaling, sex, and age. Hum. Brain Mapp. 42 (14), 4623–4642. https://doi.org/10.1002/HBM.25572.
Xu, J., Liu, X., Li, Q., et al., 2021. Global urbanicity is associated with brain and behaviour in young people. Nat. Human Behav. 6 (2), 279–293. https://doi.org/10.1038/s41562-021-01204-7, 2021 6:2.
Zundel, C.G., Ryan, P., Brokamp, C., et al., 2022. Air pollution, depressive and anxiety disorders, and brain effects: a systematic review. Neurotoxicology 93, 272–300. https://doi.org/10.1016/j.neuro.2022.10.011.