



OPEN ACCESS

EDITED BY
Enrico Capobianco,
Jackson Laboratory, United States

REVIEWED BY
Marco Diego Dominietto,
Paul Scherrer Institut (PSI), Switzerland
Chengxi Zang,
Cornell University, United States

*CORRESPONDENCE
Alfonso Monaco
alfonso.monaco@ba.infn.it

SPECIALTY SECTION
This article was submitted to
Data Analytics for Social Impact,
a section of the journal
Frontiers in Big Data

RECEIVED 30 August 2022
ACCEPTED 23 November 2022
PUBLISHED 08 December 2022

CITATION
Bellantuono L, Monaco A, Amoroso N,
Lacalamita A, Pantaleo E, Tangaro S
and Bellotti R (2022) Worldwide impact
of lifestyle predictors of dementia
prevalence: An eXplainable Artificial
Intelligence analysis.
Front. Big Data 5:1027783.
doi: 10.3389/fdata.2022.1027783

COPYRIGHT
© 2022 Bellantuono, Monaco,
Amoroso, Lacalamita, Pantaleo,
Tangaro and Bellotti. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Worldwide impact of lifestyle predictors of dementia prevalence: An eXplainable Artificial Intelligence analysis

Loredana Bellantuono^{1,2}, Alfonso Monaco^{2,3*},
Nicola Amoroso^{2,4}, Antonio Lacalamita³, Ester Pantaleo^{2,3},
Sabina Tangaro^{2,5} and Roberto Bellotti^{2,3}

¹Dipartimento di Biomedicina Traslazionale e Neuroscienze (DiBrain), Università degli Studi di Bari Aldo Moro, Bari, Italy, ²Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Bari, Italy, ³Dipartimento Interateneo di Fisica, Università degli Studi di Bari Aldo Moro, Bari, Italy, ⁴Dipartimento di Farmacia-Scienze del Farmaco, Università degli Studi di Bari Aldo Moro, Bari, Italy, ⁵Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro, Bari, Italy

Introduction: Dementia is an umbrella term indicating a group of diseases that affect the cognitive sphere. Dementia is not a mere individual health issue, since its interference with the ability to carry out daily activities entails a series of collateral problems, comprising exclusion of patients from civil rights and welfare, unpaid caregiving work, mostly performed by women, and an additional burden on the public healthcare systems. Thus, gender and wealth inequalities (both among individuals and among countries) tend to amplify the social impact of such a disease. Since at present there is no cure for dementia but only drug treatments to slow down its progress and mitigate the symptoms, it is essential to work on prevention and early diagnosis, identifying the risk factors that increase the probability of its onset. The complex and multifactorial etiology of dementia, resulting from an interplay between genetics and environmental factors, can benefit from a multidisciplinary approach that follows the “One Health” guidelines of the World Health Organization.

Methods: In this work, we apply methods of Artificial Intelligence and complex systems physics to investigate the possibility to predict dementia prevalence throughout world countries from a set of variables concerning individual health, food consumption, substance use and abuse, healthcare system efficiency. The analysis uses publicly available indicator values at a country level, referred to a time window of 26 years.

Results: Employing methods based on eXplainable Artificial Intelligence (XAI) and complex networks, we identify a group of lifestyle factors, mostly concerning nutrition, that contribute the most to dementia incidence prediction.

Discussion: The proposed approach provides a methodological basis to develop quantitative tools for action patterns against such a disease, which involves issues deeply related with sustainable, such as good health and responsible food consumption.

KEYWORDS

dementia, eXplainable Artificial Intelligence, complex systems, One Health, sustainable development goals, data science for social good, computational social science, AI for social good

1. Introduction

The term *dementia* indicates a group of diseases that progress over time and affect cognitive abilities, memory, and behavior, ultimately interfering with the capability to carry out daily activities (Reitz et al., 2011; Arvanitakis et al., 2019). Dementia cases worldwide are constantly increasing and it is estimated that by 2050 they will exceed 139 millions (WHO, 2022a). Dementias show a greater impact on the world population than other diseases that affect the central nervous system, such as brain cancer, for which about 300,000 cases were diagnosed in 2020 (IARC, 2020) and an annual growth rate of 1.50% is expected in the U.S. until 2030 (GlobalData, 2021). Alzheimer's disease (AD) is the most common form of dementia with over 70% of cases (Zhong et al., 2015; Pistollato et al., 2016; Silva et al., 2019), followed by vascular dementia (15–30%) (Wolters and Ikram, 2019), dementia with Lewy bodies (3.2–7.1%) (Hogan et al., 2016a) and frontotemporal dementia (0.01–4.6%) (Hogan et al., 2016b), and generally it has mild initial symptoms, such as memory alterations, that worsen incrementally. The progression of the disease involves behavioral changes and, in the final stages, the patient becomes completely dependent. Currently, there is no cure for AD but rather drug treatments to mitigate symptoms, with a limited therapeutic and temporal range (Mosconi and McHugh, 2015; Porsteinsson et al., 2021). Overall dementia-related costs in 2015 were estimated at 818 billion USD, namely 1.1% of global GDP, including direct medical costs, social care costs, and caregiver income losses (WHO, 2015; El-Hayek et al., 2019). The proportion of such costs is widely varying according to the country development status: in high-income states, the dementia-related costs are essentially divided between informal care (45%) and social care (40%); in low- and middle-income countries, instead, social care costs (15%) are much smaller than informal care costs (Prince et al., 2015b). Such a discrepancy contributes to increase inequalities among different areas and populations, especially considering that nearly 60% of people with dementia currently lives in low- and middle- income states (Prince et al., 2015a).

Besides the individual health issues and the social costs, dementia entails a series of collateral problems, prominently including the fact that unpaid care is often a burden of family women, which exacerbates the gender gap in the labor markets (Anderson and Oderkirk, 2015), as well as the exclusion from basic social and human rights of the affected people (WHO, 2017a), especially in contexts where they are passive and dependent recipients of care (Young et al., 2019). Therefore, increasing knowledge on dementia and finding new solutions to tackle the problem must be considered a priority also in view of the Sustainable Development Goals (SDGs) defined in the United Nations 2030 Agenda (United Nations Department of Economic and Social Affairs, 2021). Actually, for the reasons

outlined above, the issue of dementia is connected not only to SDG 3 “Good Health and Well-Being,” which has the purpose to “ensure healthy lives and promote well-being for all at all ages,” but also to SDG 5 “Gender equality,” and to SDG 10 “Reduced Inequalities,” that specifically refers to the need of empowering inclusion of people with disabilities at all levels.

Although there are no certainties about the origin of the most relevant forms of dementia, including AD, several genetic and environmental factors that could influence its appearance and evolution are being examined (Fern and Ruiz-Gabarre, 2019; Bello-Corral et al., 2021). Despite age being the strongest risk factor, it is increasingly clear that dementia is not an inevitable consequence of biological aging, and can have a juvenile onset with symptoms appearing before the age of 65. Today the focus is on a number of additional lifestyle risk factors that include smoking (Batty et al., 2014; Zhong et al., 2015), alcohol abuse, unbalanced diets (Gardener and Rainey-Smith, 2018; Moore et al., 2018; van den Brink et al., 2019; Zhang et al., 2021), obesity (Singh-Manoux et al., 2018; Ma et al., 2020), physical inactivity, high blood sugar or cholesterol values, mid-life hypertension. Furthermore, low educational attainment, cognitive inactivity, social isolation and mid-life depression, as well as environmental conditions such as air pollution, are considered as potentially modifiable risk factors specific to dementia (WHO, 2022a).

The complex and multifactorial etiology of dementia, resulting from an interplay between genes and the environment, makes the study in this field particularly suited to the multidisciplinary “One Health” approach, aimed at designing and implementing programmes, policies, legislation and research in which multiple sectors, ranging from basic sciences and clinical studies to health-services policy analyses, cooperate together to achieve better public health outcomes (WHO, 2017b). Specifically, modern health research on dementia can benefit from the Big Data framework, due to the combined availability of *broad* and *deep data*: the former consist in massive amounts of routinely-collected population-based outcome and exposure data, while the latter provide clinical and biological information on individuals (Deetjen et al., 2015).

In the present research, we focus on broad data, applying methods of Artificial Intelligence (AI) and complex systems physics to predict the incidence of dementia from a set of variables, concerning individual health, food consumption, substance use and abuse, efficiency of the healthcare system. The analysis will be made at the level of countries, using publicly available indicators referred to a time window of 26 years, collected either directly from the Global Health Observatory data repository of the World Health Organization (WHO, 2022b) or from the Our World in Data repository (Global Change Data Lab, 2022), which includes data published by different international organizations.

Among the considered features, we will investigate the most influential ones in determining dementia prevalence in the population of each country, at different years. Specifically, we shall predict the dementia incidence through a Random Forest algorithm (Breiman, 2001), and evaluate the impact of the different lifestyle indicators on each prediction by means of eXplainable Artificial Intelligence (XAI) (Adadi and Berrada, 2018), an innovative approach that allows to improve the interpretability and transparency of Machine Learning models.

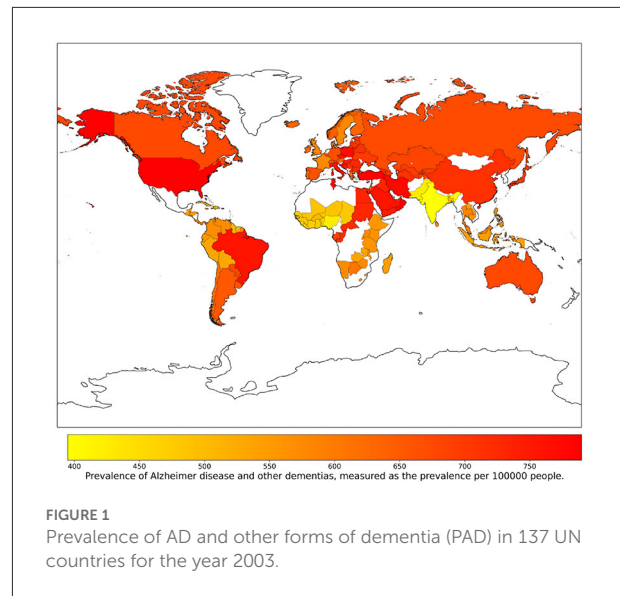
To increase the robustness of our analysis, we will employ the tool of complex networks (Newman, 2018), that is increasingly used, also in its recent formal developments (Bianconi, 2018; Amoroso et al., 2020, 2021b), along with or in replacement of traditional statistical techniques to unveil hidden information in many sectors of science and society, such as neuroscience (Sporns, 2011; Amoroso et al., 2018, 2019; Bellantuono et al., 2021), genetics (Monaco et al., 2019, 2020), economics (Hidalgo et al., 2007; Battiston et al., 2012; Tacchella et al., 2012; Hausmann et al., 2014; Bardoscia et al., 2017, 2021; Pugliese et al., 2019; Amoroso et al., 2021a), human mobility (Alessandretti et al., 2018), social and development-related issues (Bellantuono et al., 2020, 2022a,b).

The adoption of a framework combining XAI with complex system methods will lead us to first identifying predictors of dementia prevalence, and then verifying their occurrence in literature as acknowledged or suspected risk factors. The proposed workflow thus provides a possible methodological basis to develop quantitative tools for action patterns against such a problem, that involves issues deeply related with sustainable development, such as good health and, as we shall see, also responsible food consumption.

The article is organized as follows. In Section 2, we present details on data collection and processing, and on the machine learning, complex networks and XAI machinery employed in our research. The main results of this work are reported in Section 3. Finally, we discuss our findings in Section 4, pointing out their relation with previous literature and the insights they provide in terms of dementia prevention and its relation with sustainable development.

2. Materials and methods

The goal of our study was to explore whether social, economic, clinical and lifestyle factors could predict the Prevalence of AD and other forms of dementia (that we indicate as PAD) in 137 UN countries from 1993 to 2019. The choice of focusing on data referred to different geographical contexts and years is aimed at obtaining a broad and varied perspective. PAD is measured as the prevalence per 100,000 people and age-standardized, thus allowing for a fair comparison between countries and through time. Values of this index for the considered countries and the

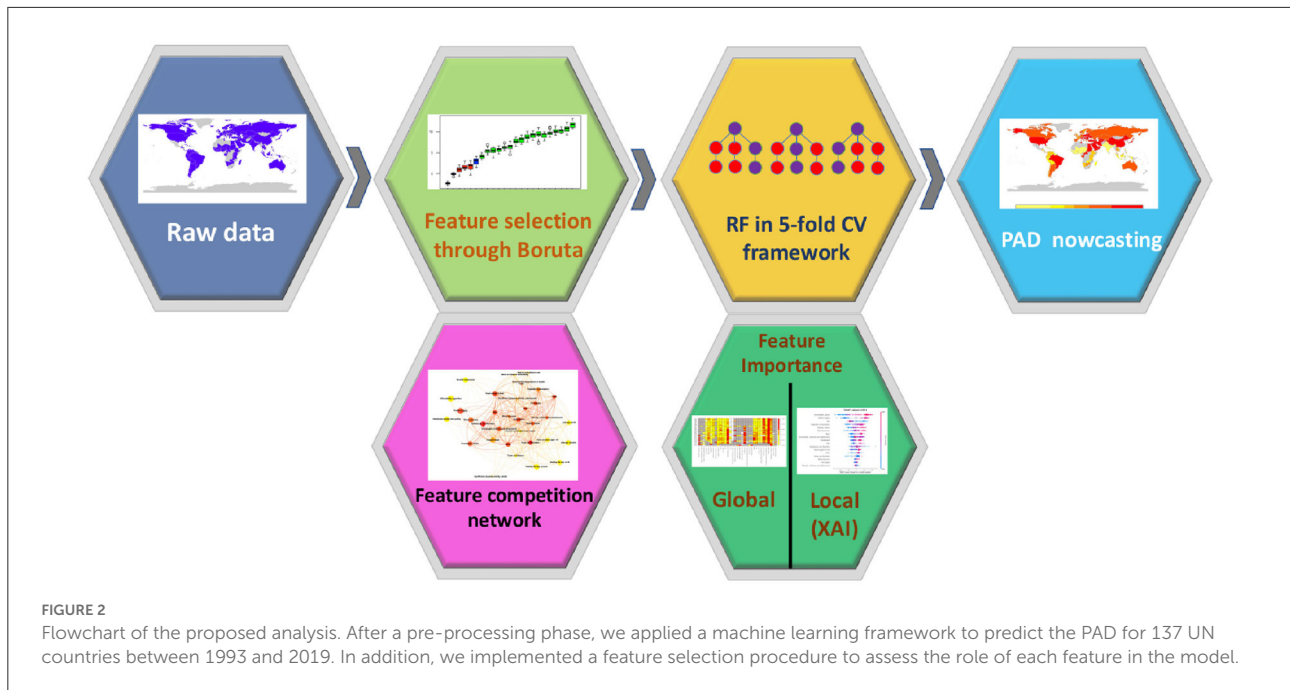


year 2003 are represented in Figure 1. This map can be regarded as representative of the whole period 1993–2019, since, in this time span, variations of PAD throughout the years are very limited: the median standard-deviation-to-mean ratio is 1.10%, with the largest value occurring for Japan (6.59%), and the smallest for Iceland (0.15%). For the prediction of the PAD, we use a machine learning approach, as summarized in Figure 2. After a preprocessing phase, we select the most informative features through the wrapper method Boruta. Based on Boruta's output, we build a feature competition network and feed a machine learning framework based on the Random Forest algorithm. Finally, we implement a feature importance procedure using two different approaches: i) global, with Random Forest internal functionalities; ii) local, based on the Shapley (SHAP) values method.

2.1. Data collection and preprocessing

In this study, we consider 34 intensive indicators (i.e., not proportional to either the population or the extension of a country) referred to the 137 United Nations (UN) Member States highlighted in Figure 1. We collect data for the years 1993–2019 from two public repositories, namely the Global Health Observatory data repository of the World Health Organization (WHO, 2022b) and the Our World in Data repository (Global Change Data Lab, 2022), which includes data published by different international organizations.

Both indicators and countries are included in this study considering data availability, which can fluctuate from 1 year



to another. In the [Supplementary material](#), we report all the indicators that we use to predict the PAD, together with information on the related sources and availability throughout the years. Since the dataset refers to a time span of 26 years, in which certain political borders or denominations have changed, we operate on input data to align country names to their current version (see [Supplementary material](#)).

In the dataset of 34 indicators for 137 countries, missing data occur throughout the availability period, amounting to about 2% of entries. We follow a preprocessing strategy to fill these gaps, summarized in these steps:

- For each pair (indicator and country) with at least 5 values in the availability period, and without any consecutive missing entries at its ends, we interpolate the indicator using a quadratic fit model with the year as independent variable.
- For other (indicator, country) pairs, missing values are replaced with the average of the available ones.

At the end of the aforementioned filling operation, we standardize over all states the distribution $F_{i,y} = \{f_{i,s}(y)\}_{1 \leq s \leq 137}$ of the values of feature f_i in the year y , by redefining entries as

$$\tilde{f}_{i,s}(y) = \frac{f_{i,s}(y) - \text{mean}(F_{i,y})}{\sqrt{\text{var}(F_{i,y})}}. \tag{1}$$

The standardized feature vector is used as input of the Boruta algorithm, that performs feature selection.

2.2. Feature selection

We implemented a feature selection wrapper method based on the Boruta framework ([Kursa and Rudnicki, 2010](#)) to reduce noise and redundant data, selecting only the uncorrelated features that improve the performance of a machine learning algorithm. Boruta is a robust and efficient feature selection tool that relies on a supervised learning Random Forest (RF) algorithm. In particular, Boruta exploits the founding concept of RF, according to which the negative effect of random fluctuations and correlation inherent in the learning model is mitigated by randomly perturbing the system and randomizing the training samples.

The method is based on attaching to the original set of features an equal number of *shadow* features, obtained by randomly shuffling the values of each original indicator. The extended dataset is used to feed a RF algorithm, which predicts a given quantity and evaluates the importance of both original and shadow features. After a series of independent random shuffling operations, Boruta selects the features that are more important, in a statistically significant way, than their respective shadow counterparts. In this way, such a feature selection method overcomes the limitations of classical techniques, where, since features compete with each other, an arbitrary importance threshold must be set to select relevant variables.

2.3. Feature competition network based on Boruta importance rankings

By comparing the Boruta feature importance rankings for different years, we construct a competition network (Criado et al., 2013; Fernández Tuesta et al., 2020), in which nodes represent features, and the strength of their connection is related with the tendency to switch their positions in the importance rankings associated to different years and countries. In such a picture, outstandingly relevant features, which tend to stay on top of the rankings for most years, are characterized by weak network connections. Competition networks provide a convenient tool to visualize and interpret the results of feature importance analysis, highlighting the competition patterns among different indicators and characterizing the stability of their hierarchies in the Boruta rankings across different years. Actually, encoding this information in a quantitative way is particularly useful when feature availability widely varies from 1 year to another. In this study, competition networks allow to combine information on availability and relevance of features in the rankings referred to a time span of 26 years. Features that are generally most relevant for prediction can be thus identified as the ones characterized at the same time by low network connectivity and high importance value.

For each pair of features $f_i(y)$ and $f_j(y)$ available to the Boruta algorithm for predicting the dementia incidence in a given year y , we consider the respective Boruta importances $I(f_i(y))$ and $I(f_j(y))$. These quantities provide the basis to construct the vector $V(f_i, f_j)$, whose entries correspond to the set of years $Y(f_i, f_j)$ in which both features $f_i(y)$ and $f_j(y)$ are available to the Boruta algorithm. Specifically, the entry $V_y(f_i, f_j)$ of $V(f_i, f_j)$ corresponding to a given year $y \in Y(f_i, f_j)$ reads

$$V_y(f_i, f_j) = \begin{cases} 1 & \text{if } I(f_i(y)) > I(f_j(y)) \\ 0 & \text{if } I(f_i(y)) \leq I(f_j(y)) \end{cases} \quad (2)$$

The variety of entries in $V(f_i, f_j)$ is quantified by its Shannon entropy $E(f_i, f_j)$. Hence, if the feature $f_i(y)$ has a higher Boruta importance than $f_j(y)$ in the prediction of dementia incidence for all years $y \in Y(f_i, f_j)$, then all the entries of $V(f_i, f_j)$ will be equal to zero, and $E(f_i, f_j) = 0$. The same occurs if the Boruta importance of $f_j(y)$ is always higher than the one of $f_i(y)$, with all the entries of $V(f_i, f_j)$ being equal to 1.

The competition network consists of nodes, coinciding with the features f_i which are available to the Boruta algorithm at least for 1 year, and weighted edges (f_i, f_j, w_{ij}) , with weights $w_{ij} = E(f_i, f_j)$. If $w_{ij} = 0$, there is no link between f_i and f_j , i.e. these features do not compete with each other.

2.4. Learning model

Based on the evaluated importance, the Boruta algorithm selects a subset of relevant features for each year. We use selected features to feed a Random Forest model. Random Forest (RF) (Breiman, 2001) is one of the most used algorithms in supervised machine learning applications due to its versatility, ease of tuning, and ability to model multimodal data. A RF configuration depends on only two parameters: (i) the number of trees of the forest, M ; (ii) the number of features to be chosen randomly at each split, f . RF comprises an ensemble of classification and regression trees (CART) that is grown in the training phase through a bootstrap process and a feature randomization procedure; this procedure makes RF robust against overfitting and only loosely correlated regressor trees. Another important advantage of RF is the possibility to internally evaluate the role of each feature for the model accuracy. In the present work we assess the feature importance using the mean decrease impurity and set a RF standard configuration with $M = 500$ trees and $f = F/3$, where F is the number of input features.

To further increase the robustness of our procedure we implement a 5-fold cross validation (CV) scheme repeated 100 times. The average of the 5×100 performance values obtained with this approach is a reliable indicator of the overall model performance. Similarly, RF feature importance is assessed through 100 CVs, and overall feature importance is computed by averaging. We measure performances in terms of coefficient of determination between predicted and actual values (R^2). In addition, we evaluate the root mean square error (RMSE), defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - F_i)^2}, \quad (3)$$

and the mean absolute percentage error (MAPE), defined as (de Myttenaere et al., 2016):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (4)$$

where A_i is the actual value and F_i is the forecast value. Data processing and statistical analyses have been performed in R 3.6.1 (R Core Team, 2018) and Python 3.7.

2.5. eXplainable Artificial Intelligence and Shapley values

The XAI framework meets the crucial need to increase transparency and interpretability of Machine Learning models, especially relevant in their real-life applications (Miller, 2019; Bussmann et al., 2020; Jiménez-Luna et al., 2020; Lombardi

et al., 2021). While, historically, a major role in developing AI models has been played by informativeness, quantified through performance metrics, and uncertainty estimation (Schaffer, 1993; Rao et al., 2008; Musil et al., 2019), an increasing attention is paid nowadays to the matters of generalization (reliability of predictions on previously unseen data) and transparency (ability to make the decision process as intelligible as possible) (Flach, 2019; Vollmer et al., 2020). XAI represents the common label of a framework of techniques that follow a unified view, which combines informativeness, uncertainty estimation, generalization, and transparency. In this work, we adopt the SHAP local explanation algorithm to detect importance of features for the PAD prediction in each country and for each year considered in the analysis.

The SHAP algorithm is based on the concept of the Shapley (SHAP) values, mutated from cooperative game theory (Lundberg and Lee, 2017; Lundberg et al., 2020), and consists in a local model-agnostic post-hoc explainer, that learns local interpretable linear models for the samples, focusing on the contributions of each feature on the prediction of each sample. The SHAP value for a feature j is evaluated as the difference between the prediction of the model output with and without that specific feature, considering all possible feature subsets. Hence, the model must be retrained on all feature subsets F of the entire set S of features ($F \subseteq S$). If $f_x(F)$ represents the prediction f of a model for the instance x , given a subset F that does not include the j -th feature, and $f_x(F \cup j)$ is the prediction of the same model when the j -th feature is added, the marginal contribution provided by the j -th feature can be computed as the difference $f_x(F \cup j) - f_x(F)$. The SHAP value of the j -th feature for the instance x is then assessed by considering the addition of the j -th feature to all possible subsets,

$$SHAP_j(x) = \sum_{F \subseteq S - \{j\}} \frac{|F|!(|S| - |F| - 1)!}{|S|!} [f_x(F \cup j) - f_x(F)], \quad (5)$$

where $|F|!$ represents the number of permutations of features positioned before the j -th feature, $(|S| - |F| - 1)!$ represents the number of permutations of feature values that appear after the j -th feature value and $|S|!$ is the total number of feature permutations (Lundberg and Lee, 2017).

In this work, the SHAP values are referred to importance of the features selected by Boruta in the PAD prediction, made by the RF algorithm. In particular, we consider the mean SHAP values computed after a 5-fold cross validation, repeated 100 times for each year. Then, in view of the geographical analysis, we operate for each year a further selection, by

- Evaluating the average impact of each feature as the mean absolute SHAP value on all countries,
- Computing the total mean absolute SHAP value, by summing on all features,

- Retaining only the features whose mean absolute SHAP value is larger than 25% of the total.

3. Results

3.1. Boruta competition network

Starting from the Boruta feature importance, we construct a competition network, based on the rankings of importance values for different years. Figure 3 shows in the upper panel the competition network structure, consisting of 34 nodes representing features, and 338 weighted edges that quantify the tendency of different features to switch their positions in the Boruta importance rankings. Nodes shown in Figure 3 are characterized by colors that, in a scale from yellow to red, indicate the availability of the corresponding feature, and sizes that increase with the related Boruta feature importance values.

The lower panel of Figure 3 contains a scatter plot in which points correspond to features and their coordinates are competition network degree and mean Boruta importance value. Here, points organize according to a “horseshoe” pattern, in which features with low degree tend to have either high or low importance, while features with high degree tend to have intermediate importance. Also in this figure, the color scale corresponds to the number of yearly rankings in which a feature appears. The competition network analysis corroborates and integrates the results of Boruta importance: features that appear in the top left part of the scatter plot in Figure 3 and occur with high frequency in rankings, such as Protein supply, are steadily among the most relevant. On the other hand, frequent features appearing in the bottom left part, such as Government expenditure on health, can be safely considered irrelevant for dementia incidence prediction.

3.2. Random forest predictions

We evaluate the effectiveness of predicting PAD through a RF regression algorithm, trained with the considered set of features. The performance of the regressor is quantified by the indicators described in Section 2.4. Figure 4 shows the relative p -value of the agreement (R^2) between the PAD actual values and RF predictions for each year. The distributions reported in the plot are obtained by performing 100 cycles of 5-fold cross-validation, providing a total of 500 predictions. In the Supplementary material, we display in Supplementary Figures S1–S3 the related plots of R^2 , $RMSE$, and $MAPE$, respectively.

The behavior of all the performance indicators, reported in Table 1, shows that the predictive capability of the RF model is good, with fluctuations in time, until the year 2016, where a sudden increase of the root-mean-square error indicates a

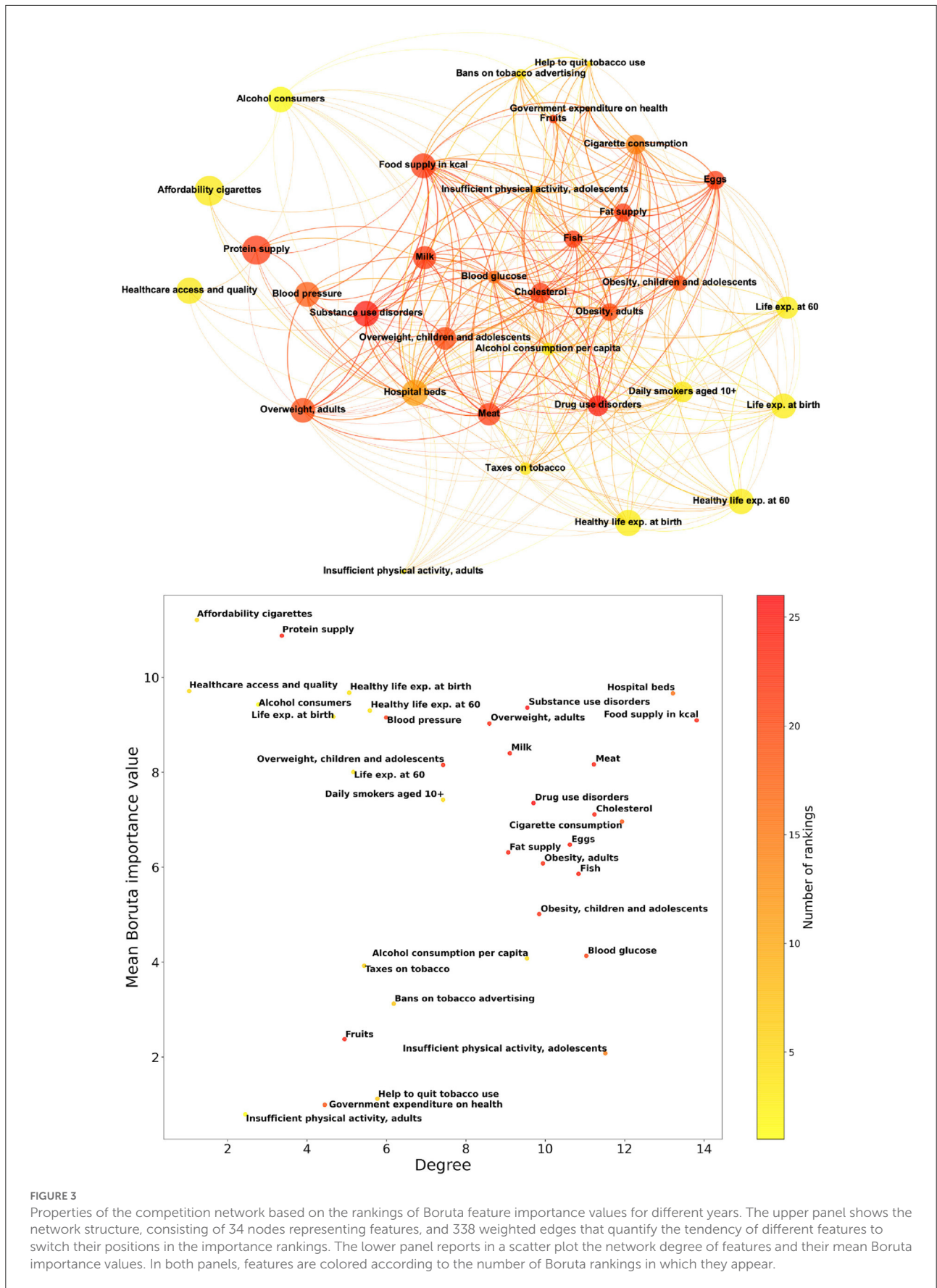
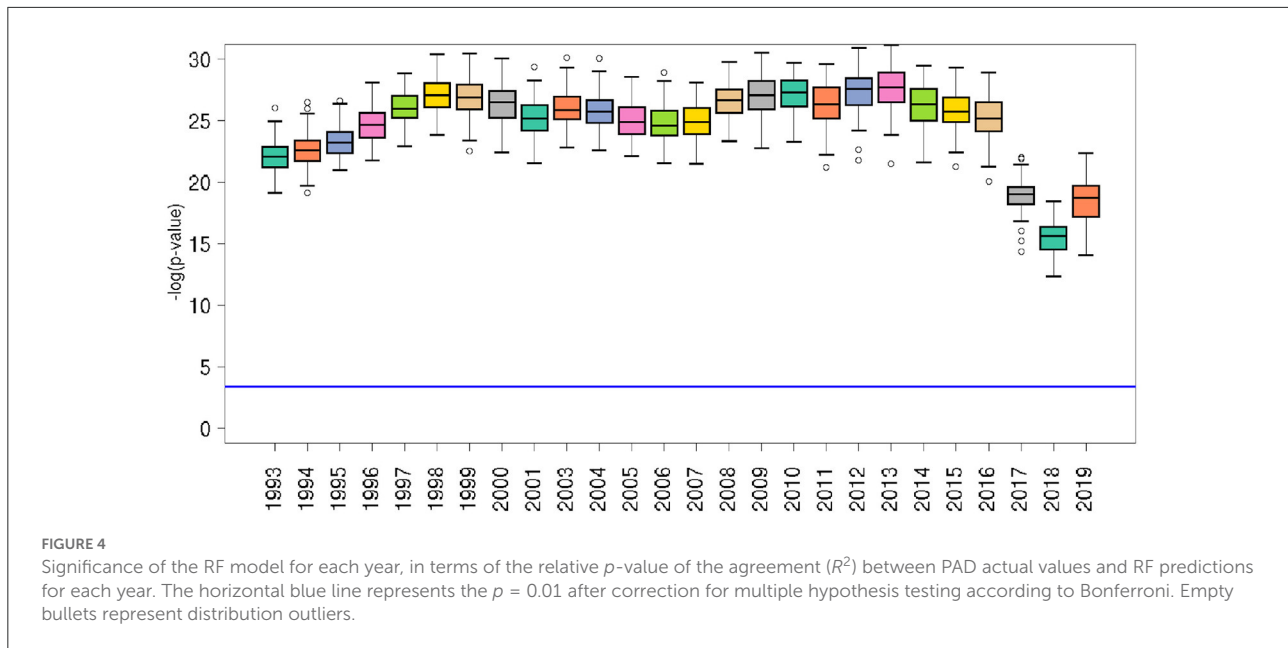


FIGURE 3

Properties of the competition network based on the rankings of Boruta feature importance values for different years. The upper panel shows the network structure, consisting of 34 nodes representing features, and 338 weighted edges that quantify the tendency of different features to switch their positions in the importance rankings. The lower panel reports in a scatter plot the network degree of features and their mean Boruta importance values. In both panels, features are colored according to the number of Boruta rankings in which they appear.



performance drop, followed in 2017 by all the other indicators. Such a change is due to the fact that the values of many relevant indicators are missing for the most recent years.

3.3. Feature importance

The results of the feature importance procedure performed through the RF algorithm are summarized in Figure 5, where the importance of features throughout the years is represented in a color scale from red (high) to yellow (low). White and gray boxes indicate that the corresponding features are either not selected by Boruta or missing for a given year, respectively. It is possible to detect that specific variables related to nutrition (Food supply in kcal, Overweight adults, Protein supply, and, especially for recent years, Meat) and to medical care (Hospital beds, Healthcare access and quality) tend to have a consistently high importance. In the last years, the number of unavailable indicators becomes larger, and RF tends to assign high importance to the remaining ones; however, as highlighted by the performance drop mentioned in Section 3.2, the predictive power of these indicators alone is low.

3.4. SHAP values

For each year and each feature selected by Boruta, we compute the SHAP values associated to the PAD prediction for each given country. The plots in Figures 6–8 show the distribution of SHAP values for each selected feature, for the years 1996, 2006, and 2016. In these plots, the relevant

features for the chosen year are ordered in terms of the mean absolute SHAP value, which indicates their overall impact on the dementia incidence prediction, independent of the way (positive or negative) in which they affect each outcome.

3.5. Geographical analysis

The SHAP values provide the basis for investigating the predictive power of the considered features from a geographical point of view, focused on single countries. As stated in Section 2.5, we retain only those features such that, in a given year, their mean absolute SHAP value is larger than 25% of the total absolute SHAP value. For the years 1993–2016, in which the RF model has a significant predictive power, this condition is satisfied in the following cases:

- Overweight, adults (1993–1999, and 2001);
- Protein supply (1993–2001, 2003–2009, 2011, 2013, and 2014);
- Food supply in kcal (2004, 2006–2009, 2011, 2013, and 2014);
- Meat (2013);
- Healthcare access and quality (2015);
- Hospital beds (2016).

In Figure 9, we show the distribution in a world map of the SHAP values related to each of the above features, for one selected year. Since the results are qualitatively stable throughout the years, we preferably choose, when available, the values for one of the years 1996, 2006, 2016, to which Figures 6–8

TABLE 1 Summary table of performance measures obtained for each year.

Year	R^2	MAPE	RMSE
1993	0.503 ± 0.020	0.085 ± 0.002	66.075 ± 1.340
1994	0.513 ± 0.020	0.083 ± 0.002	65.465 ± 1.332
1995	0.525 ± 0.018	0.082 ± 0.002	64.731 ± 1.262
1996	0.541 ± 0.020	0.082 ± 0.002	63.867 ± 1.310
1997	0.562 ± 0.017	0.080 ± 0.002	62.831 ± 1.214
1998	0.574 ± 0.017	0.080 ± 0.002	62.605 ± 1.228
1999	0.570 ± 0.018	0.082 ± 0.002	63.385 ± 1.349
2000	0.566 ± 0.019	0.081 ± 0.002	64.029 ± 1.420
2001	0.552 ± 0.020	0.084 ± 0.002	65.073 ± 1.423
2003	0.563 ± 0.018	0.083 ± 0.002	64.186 ± 1.320
2004	0.560 ± 0.018	0.083 ± 0.002	64.349 ± 1.323
2005	0.548 ± 0.019	0.083 ± 0.002	65.230 ± 1.390
2006	0.543 ± 0.020	0.084 ± 0.002	65.516 ± 1.472
2007	0.547 ± 0.020	0.083 ± 0.002	65.240 ± 1.448
2008	0.569 ± 0.018	0.081 ± 0.002	63.702 ± 1.334
2009	0.572 ± 0.020	0.082 ± 0.002	63.575 ± 1.482
2010	0.573 ± 0.018	0.082 ± 0.002	63.610 ± 1.319
2011	0.561 ± 0.022	0.082 ± 0.002	64.631 ± 1.570
2012	0.573 ± 0.020	0.082 ± 0.002	63.987 ± 1.451
2013	0.580 ± 0.020	0.080 ± 0.002	63.709 ± 1.543
2014	0.561 ± 0.021	0.082 ± 0.002	65.509 ± 1.552
2015	0.554 ± 0.019	0.082 ± 0.002	66.288 ± 1.425
2016	0.545 ± 0.022	0.086 ± 0.002	67.242 ± 1.622
2017	0.450 ± 0.022	0.093 ± 0.002	74.172 ± 1.505
2018	0.383 ± 0.024	0.102 ± 0.002	78.142 ± 1.512
2019	0.442 ± 0.030	0.094 ± 0.002	73.510 ± 1.939

R^2 , MAPE, and RMSE were averaged over 100 rounds of cross-validation and reported with the respective standard deviations.

are referred. The summary plots of SHAP values for the years 2013 and 2015, to which the maps related to Meat and Healthcare access and quality are referred, are reported in the [Supplementary Figures S4, S5](#). We elaborate on the implications of these results in Section 4.

3.6. Inclusion of indicators not related to lifestyle

The analysis presented in this work is oriented to a comparison inside the set of lifestyle indicators as predictors of PAD. However, in order to test the robustness of its outcomes, we extend the range of indicators to check whether the lifestyle ones are still among the most influential to predict PAD. Specifically, we add to the dataset indicators in the following categories:

- Prevalence of confounding health disorders (“Depressive disorders,” “Autism spectrum disorders,” “ADHD,” “Type 2 Diabetes Mellitus”);
- Air pollution (“Fine particulate matter”);
- Schooling (“Average years of schooling,” “School life expectancy”);
- Country wealth (“GDP per capita”).

For indicator availability reasons, we test the extended dataset for the PAD regression related to the year 2013. Details on the sources of the additional indicators are reported in the [Supplementary material](#). As expected, the RF performance slightly improves by extending the dataset, with $R^2 = 0.602 \pm 0.020$, $MAPE = 0.078 \pm 0.002$ and $RMSE = 62.071 \pm 1.574$ (see [Table 1](#) for comparison). The top-ranked SHAP values for the extended dataset are reported in [Figure 10](#), where it can be observed that lifestyle indicators are still among the most influential.

4. Discussion

We have investigated the factors that affect the prevalence of dementia in 137 UN countries, combining machine-learning based feature importance, the XAI approach, and the complex network formalism. First, we selected the most relevant variables by means of Boruta importance rankings, encompassing information concerning different data availability of the indicators through a competition network. From this part of the analysis, we found relevant features related to individual health parameters (overweight, blood pressure), life expectancy, which is trivially correlated with the onset of dementia, healthcare access and quality, substance use/abuse, and, most importantly, nutrition (protein supply, milk and meat consumption). On the other hand, among the least important features we found those related to government health expenditure, insufficient physical activity among adolescents, fruit consumption, and actions to discourage tobacco use. The importance of nutrition-related variables is emphasized also in the SHAP analysis, which highlights the impact of food supply in general, protein supply, and meat consumption, along with the percentage of overweight adults, on the prediction of dementia prevalence. From the plots in [Figures 6–8](#), one can also notice that low values of fish consumption are associated with a higher incidence prediction.

The results of the SHAP values geographical analysis provide interesting insights when compared to the PAD world map in [Figure 1](#). First, we must observe how the variables “Hospital beds” and “Healthcare access and quality” positively affect the prediction of PAD for countries in America, Europe, Australia, Central Asia, and some countries in Middle and Far East (including China and Japan). The obtained results indicate how the PAD can be underestimated in countries with a weak

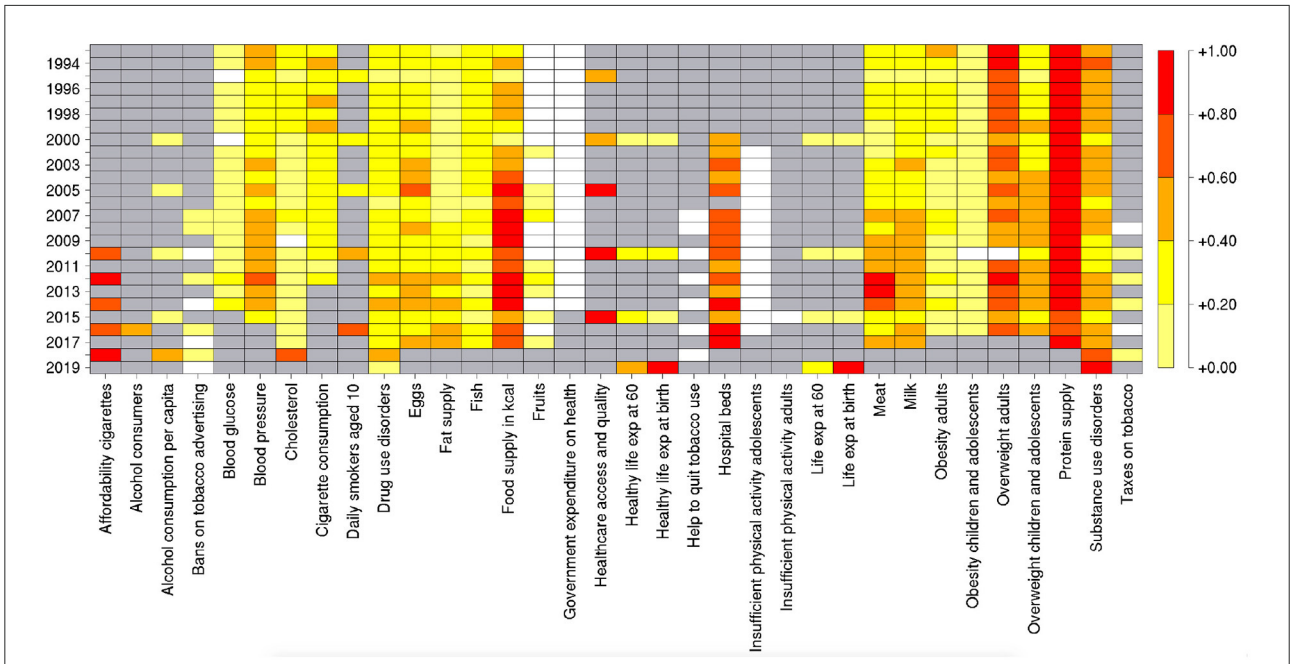


FIGURE 5
Average importance of the features used in the RF model over 100 repetitions of the 5-fold cross validation procedure, for 26 different years. The white boxes indicate that the corresponding feature was discarded in the preselection based on the Boruta algorithm. The gray boxes indicate that the corresponding indicators are missing for those years.

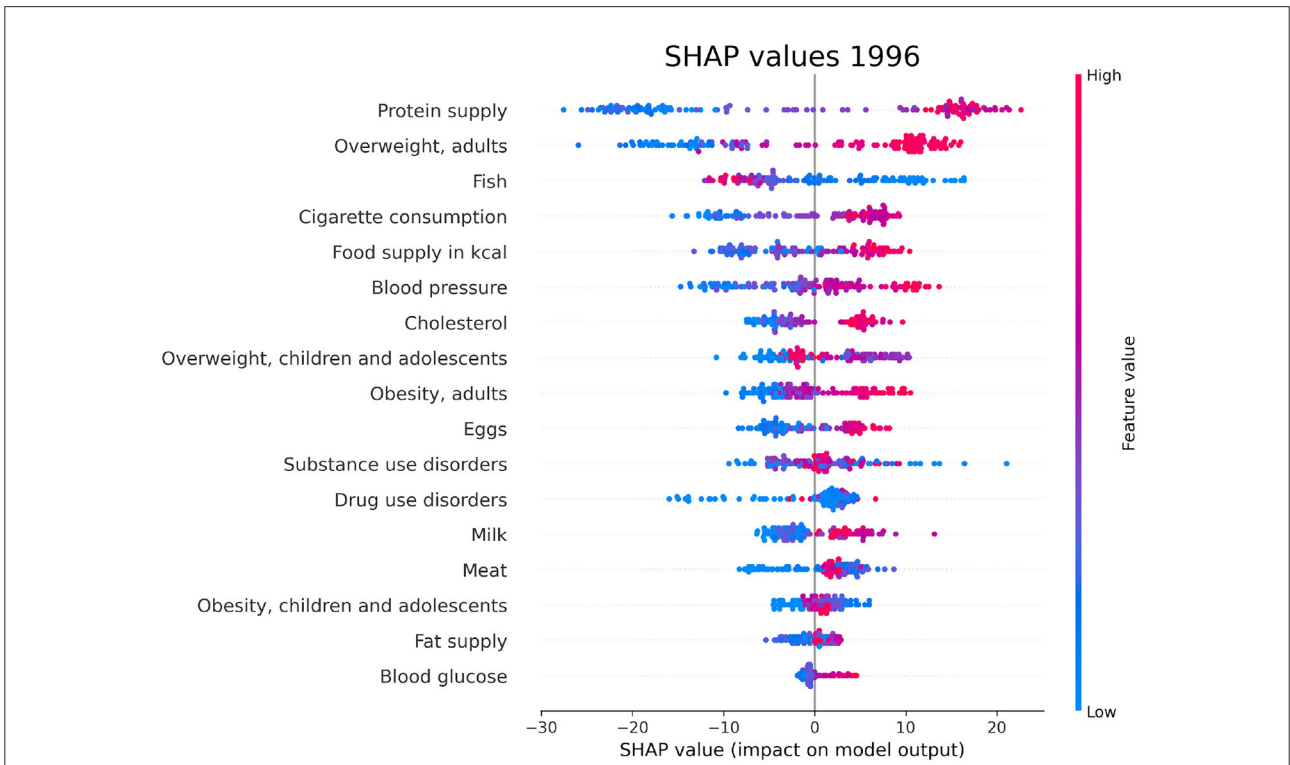
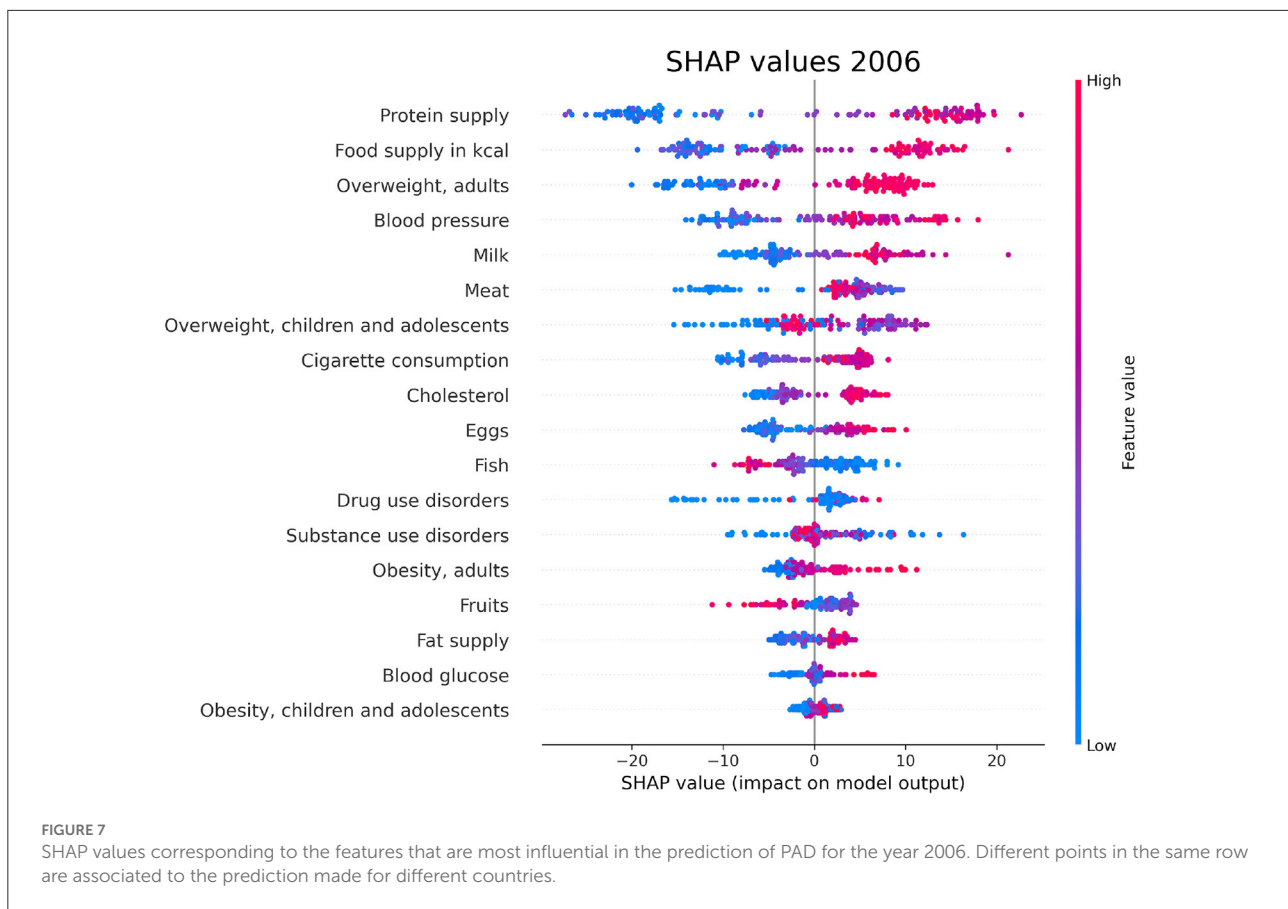


FIGURE 6
SHAP values corresponding to the features that are most influential in the prediction of PAD for the year 1996. Different points in the same row are associated to the prediction made for different countries.

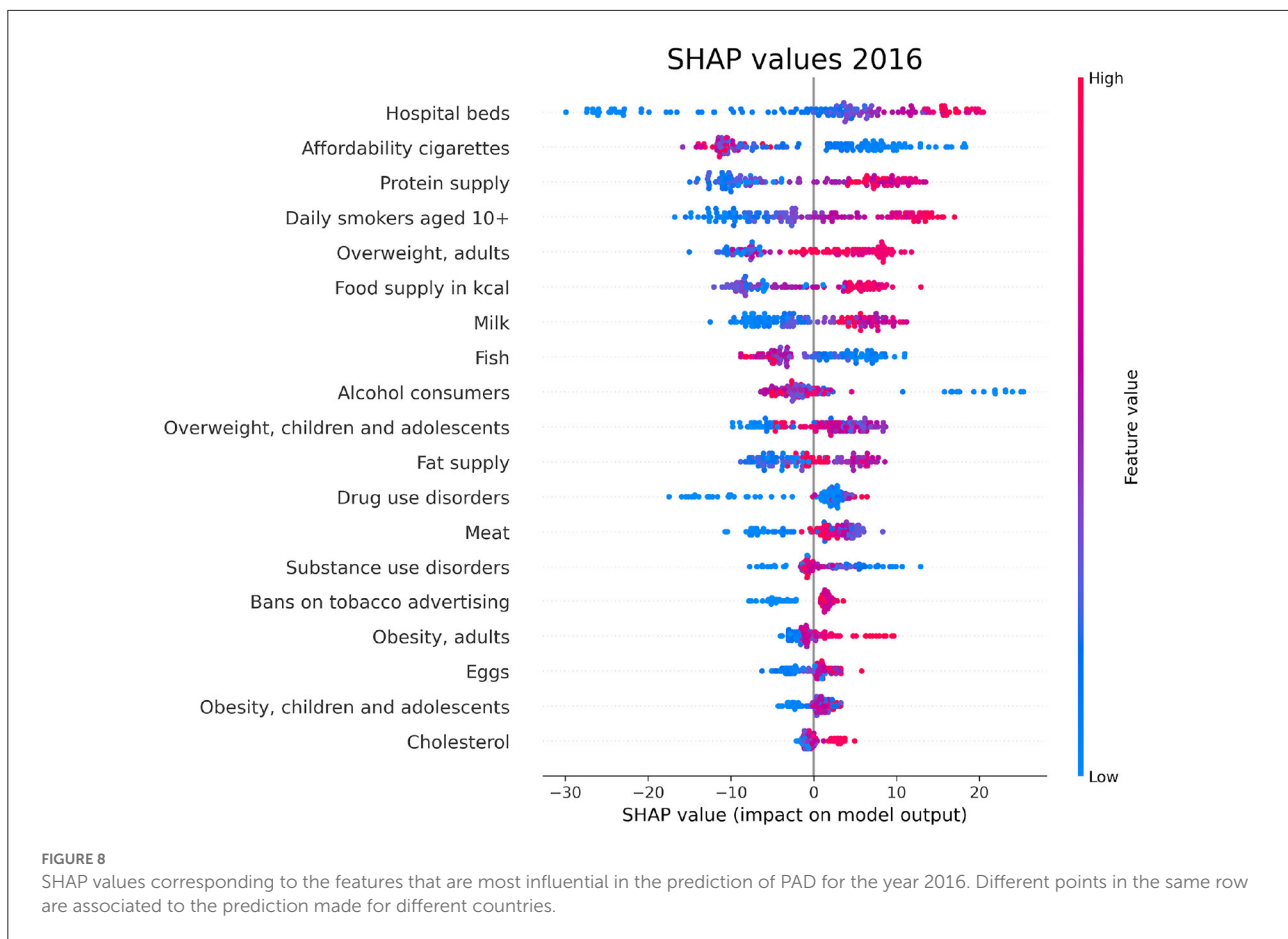


and poorly organized healthcare structure, where dementia diagnosis is more difficult. In the maps concerning overweight and nutrition, we can find interesting patterns that relate to specific features observed in Figure 1. We highlight in particular the following relevant cases:

- Higher SHAP values for the food supply and protein supply variables can explain the fact that Brazil, Argentina and Uruguay have larger PAD than surrounding countries in South America;
- The higher PAD for Thailand with respect to the surrounding countries is partly explained by meat consumption; however, one should also notice that a relevant role in the prediction for this country is played by the healthcare-related variables.
- The belt of African countries with high PAD in Figure 1, comprising Egypt, Sudan, Central African Republic, Republic of Congo, and Gabon, is characterized, as a whole, by high SHAP values for meat consumption.

The recurring relation between dementia incidence and dietary patterns finds many correspondences in literature, especially regarding AD. Previous works bring out a link

between nutrition and both dementia and depression, together with the benefits of fruit consumption on the central nervous system (Moore et al., 2018). Moreover, a healthy diet such as the Mediterranean one is known as a protective factor against neurodegenerative diseases, and AD in particular (Gardener and Rainey-Smith, 2018; van den Brink et al., 2019), as it preserves neuronal synapses and delays cognitive impairment through the intake of nutrients with antioxidant, anti-inflammatory and free radical contrast properties (Dominguez and Barbagallo, 2016). On the other hand, processed meat consumption has been identified as a potential risk factor for dementia (Zhang et al., 2021), although there is no unanimous consensus on this result, since other studies relate very low meat consumption with a long-term risk of dementia and AD (Ngabirano et al., 2019). The central role of nutrition in the PAD prediction, emerging in our XAI framework, corroborates the necessity to implement prevention strategies based on the One Health approach (WHO, 2017b). This model proposes to take on the needs of the most vulnerable subjects, including people with disabilities, considering the deep relation between the well-being of individuals, animals and the ecosystem. The case of dementia prevention is emblematic in this regard, as it highlights connections between individual health and



sustainability of food policies, especially concerning meat production and consumption.

The role of health parameters closely associated to nutrition, such as overweight and obesity, has already been recognized in previous literature. Specifically, a research by Singh-Manoux et al. (2018) showed that obesity at age 50 can increase the risk of dementia. This finding was corroborated by another study (Ma et al., 2020) associating increased body weight and abdominal obesity with increased dementia incidence. Another emerging factor in our analysis, namely tobacco use, has been investigated in studies which found that mortality associated with dementia is higher in smokers than in never smokers (Batty et al., 2014), and that the increased risk of dementia for smokers decreases after smoking cessation (Zhong et al., 2015). We finally observe that, as one can observe in Figure 10, the relevance of the considered lifestyle indicators in determining PAD prediction persists even after extending the dataset to the categories described in Section 3.6. Remarkably, the first occurrence of an additional variable, ranking seventh in terms of mean absolute SHAP value, is constituted by “Fine particulate matter,” which has been already identified as

a predictor of dementia in previous literature (Peters et al., 2019).

5. Conclusions and outlook

In this research work, we focused on the prediction of PAD, an aggregated value representing prevalence of dementia in the population, available for a time period of 26 years, starting from a set of lifestyle indicators, referred to the same year of the prediction. As we previously observed, dementia starts developing with mild symptoms, and usually remains latent until their worsening. The delay between dementia onset and diagnosis depends on various factors, related to both the individual and the features of the healthcare system, such as diagnostic promptness and technological advance of medical instruments. Even the progression speed of the disease can be influenced by individual and environmental factors. Therefore, it is reasonable to think that dementia prevalence, quantified by PAD, takes into account people who became ill in different times, and are thus undergoing

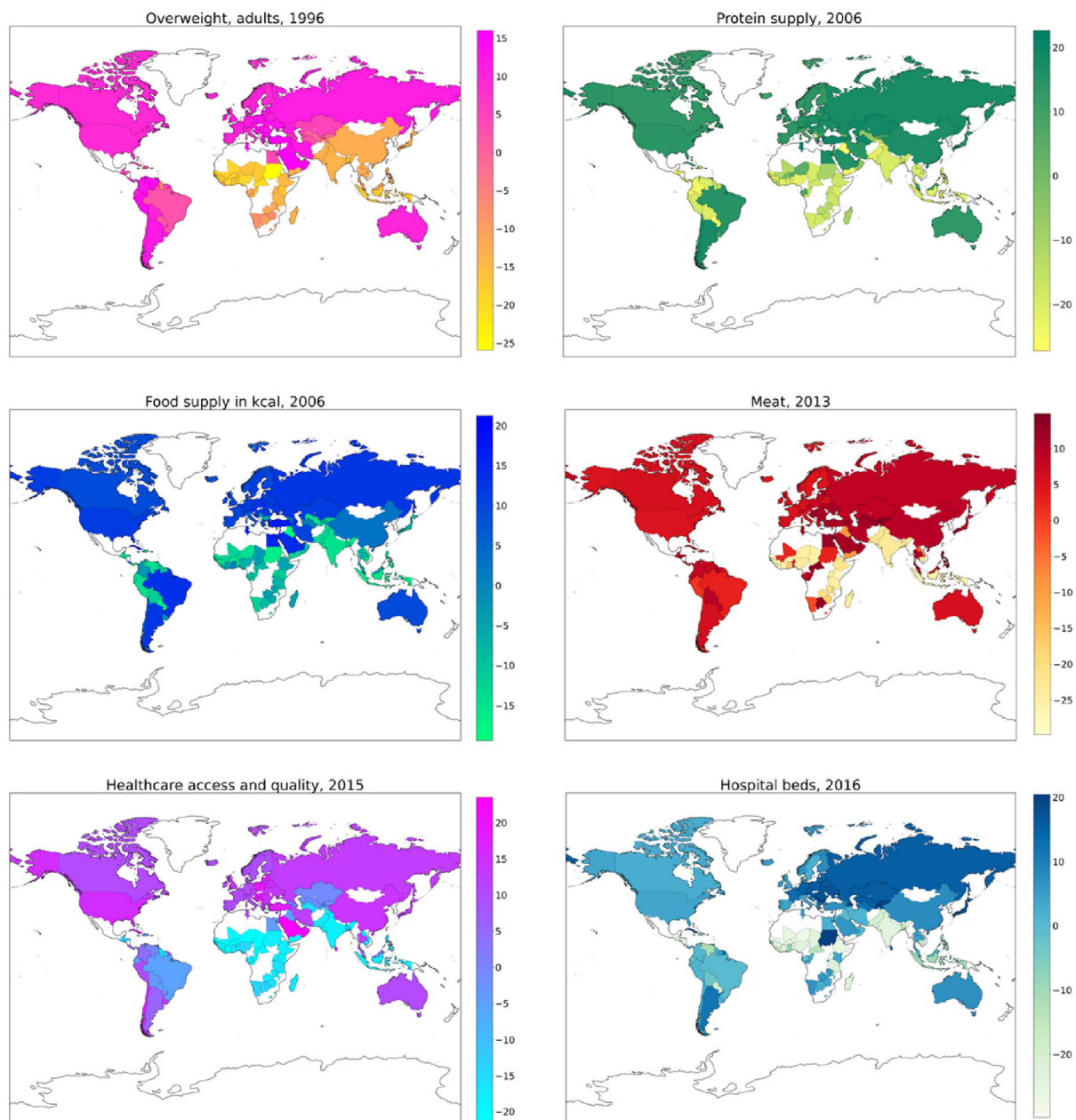
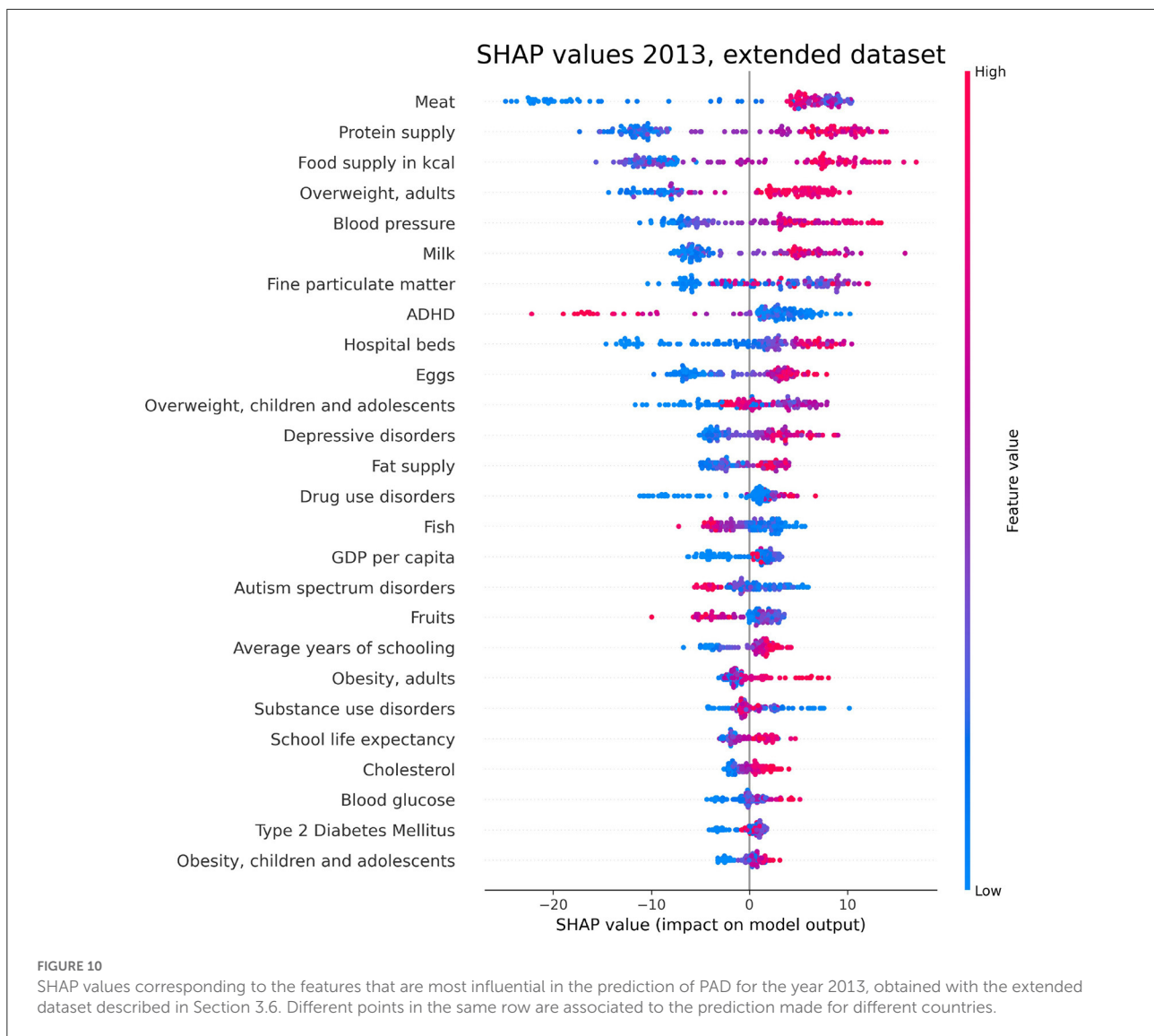


FIGURE 9
World maps of SHAP values associated with PAD prediction, related to the feature and year indicated in the each map label. Color bars are reported for numerical reference.

different stages of the disease. Moreover, lifestyle risk and protective factors can act on very different time scales, depending on their specific nature, e.g. associated with nutrition, on personal health parameters, or habits. Such an intrinsic variability in setting up the problem led us to the choice of training and testing the model for each fixed year. On the other hand, an extension perspective of

this work concerns the possibility to predict PAD values referred to a specific advancement stage of the disease, using lifestyle indicators of previous years and optimizing the reference timescales.

The analysis performed in this work, based on the use of broad data in the aggregated form, can be extended with innovative information acquisition strategies, that can make



the data collection procedure multimodal and more pervasive. A crucial contribution to mapping predictors of dementia onset can come from integrating deep data with routine health data collected by online patient platforms, but also from retailers and mobile phone providers, that allow to gain insights on individual social habits and lifestyle (Deetjen et al., 2015). This approach, already followed in other applications of Artificial Intelligence for Social Good to predict and monitor vulnerabilities related to poverty (Steele and et al., 2017), urban segregation (Lamanna et al., 2018), food insecurity (Martini et al., 2021), and gender inequalities (Garcia et al., 2018) represents a promising frontier to develop dementia prevention strategies, as well as decisional support and personalized medicine tools.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LB and AM: conceptualization, methodology, software, validation, investigation, data curation, writing–original draft preparation, and visualization. LB, AM, NA, AL, EP, ST, and RB: writing–review and editing. RB: supervision. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adadi, A., and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160. doi: 10.1109/ACCESS.2018.2870052
- Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., and Baronchelli, A. (2018). Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* 2, 485–491. doi: 10.1038/s41562-018-0364-x
- Amoroso, N., Bellantuono, L., Monaco, A., De Nicolò, F., Somma, E., and Bellotti, R. (2021a). Economic interplay forecasting business success. *Complexity* 2021, 8861267. doi: 10.1155/2021/8861267
- Amoroso, N., Bellantuono, L., Pascazio, S., Lombardi, A., Monaco, A., Tangaro, S., et al. (2020). Potential energy of complex networks: a quantum mechanical perspective. *Sci. Rep.* 10, 18387. doi: 10.1038/s41598-020-75147-w
- Amoroso, N., Bellantuono, L., Pascazio, S., Monaco, A., and Bellotti, R. (2021b). Characterization of real-world networks through quantum potentials. *PLoS ONE* 16, e0254384. doi: 10.1371/journal.pone.0254384
- Amoroso, N., La Rocca, M., Bellantuono, L., Diacono, D., Fanizzi, A., Lella, E., et al. (2019). Deep learning and multiplex networks for accurate modeling of brain age. *Front. Aging Neurosci.* 11, 115. doi: 10.3389/fnagi.2019.00115
- Amoroso, N., La Rocca, M., Bruno, S., Maggipinto, T., Monaco, A., Bellotti, R., et al. (2018). Multiplex networks for early diagnosis of Alzheimer's disease. *Front. Aging Neurosci.* 10, 365. doi: 10.3389/fnagi.2018.00365
- Anderson, G., and Odekerk, J. (2015). *Dementia Research and Care: Can Big Data Help?* Paris: OECD Publishing.
- Arvanitakis, Z., Shah, R., and Bennett, D. (2019). Diagnosis and management of dementia: review. *JAMA* 322, 1589–1599. doi: 10.1001/jama.2019.4782
- Bardoscia, M., Barucca, P., Battiston, S., Caccioli, F., Cimini, G., Garlaschelli, D., et al. (2021). The physics of financial networks. *Nat. Rev. Phys.* 3, 490–507. doi: 10.1038/s42254-021-00322-5
- Bardoscia, M., Battiston, S., Caccioli, F., and Caldarelli, G. (2017). Pathways towards instability in financial networks. *Nat. Commun.* 8, 14416. doi: 10.1038/ncomms14416
- Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. (2012). DebtRank: too central to fail? Financial networks, the FED and systemic risk. *Sci. Rep.* 2, 541. doi: 10.1038/srep00541
- Batty, G. D., Russ, T. C., Starr, J. M., Stamatakis, E., and Kivimäki, M. (2014). Modifiable cardiovascular disease risk factors as predictors of dementia death: pooling of ten general population-based cohort studies. *J. Negat. Results BioMed* 13, 8. doi: 10.1186/1477-5751-13-8
- Bellantuono, L., Marzano, L., La Rocca, M., Duncan, D., Lombardi, A., Maggipinto, T., et al. (2021). Predicting brain age with complex networks: from adolescence to adulthood. *Neuroimage* 225, 117458. doi: 10.1016/j.neuroimage.2020.117458
- Bellantuono, L., Monaco, A., Amoroso, N., Aquaro, V., Bardoscia, M., Demarinis Loiotile, A., et al. (2022a). Territorial bias in university rankings: a complex network approach. *Sci. Rep.* 12, 4995. doi: 10.1038/s41598-022-08859-w
- Bellantuono, L., Monaco, A., Amoroso, N., Aquaro, V., Lombardi, A., Tangaro, S., et al. (2022b). Sustainable development goals: conceptualization, communication and achievement synergies in a complex network framework. *Appl. Netw. Sci.* 7, 14. doi: 10.1007/s41109-022-00455-1
- Bellantuono, L., Monaco, A., Tangaro, S., Amoroso, N., Aquaro, V., and Bellotti, R. (2020). An equity-oriented rethink of global rankings with complex networks mapping development. *Sci. Rep.* 10, 18046. doi: 10.1038/s41598-020-74964-3
- Bello-Corral, L., Sánchez-Valdeón, L., Casado-Verdejo, I., Ángel Seco-Calvo, J., Fernández-Fernández, J. A., and Fernández-Martínez, M. N. (2021). The influence of nutrition in Alzheimer's disease: neuroinflammation and the microbiome vs. transmissible prion. *Front. Neurosci.* 15, 677777. doi: 10.3389/fnins.2021.677777
- Bianconi, G. (2018). *Multilayer Networks-Structure and Function*. New York, NY: Oxford University Press.
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 32–45. doi: 10.1023/A:1010933404324
- Bussmann, N., Giudici, P., Marinelli, D., and Papenbrock, J. (2020). Explainable AI in fintech risk management. *Front. Artif. Intell.* 3, 26. doi: 10.3389/frai.2020.00026
- Criado, R., García, E., Pedroche, F., and Romance, M. (2013). A new method for comparing rankings through complex networks: model and analysis of competitiveness of major European soccer leagues. *Chaos* 23, 043114. doi: 10.1063/1.4826446
- de Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing* 192, 38–48. doi: 10.1016/j.neucom.2015.12.114
- Deetjen, U. E., Meyer, R., and Schroeder, R. (2015). Big data for advancing dementia research: an evaluation of data sharing practices in research on age-related neurodegenerative diseases. *OECD Digit. Econ. Pap.* 246, 76–84. doi: 10.1787/5js4sdbdf7jk-en
- Dominguez, L. J., and Barbagallo, M. (2016). The biology of the metabolic syndrome and aging. *Curr. Opin. Clin. Nutr. Metab. Care* 19, 5–11. doi: 10.1097/MCO.0000000000000243
- El-Hayek, Y. H., Wiley, R. E., Khoury, C. P., Daya, R. P., Ballard, C., Evans, A. R., et al. (2019). Tip of the iceberg: assessing the global socioeconomic costs of Alzheimer's disease and related dementias and strategic implications for stakeholders. *J. Alzheimers Dis.* 70, 323–341. doi: 10.3233/JAD-190426
- Fern, P., and Ruiz-Gabarré, D. (2019). Modulating effect of diet on Alzheimer's disease. *Diseases* 7, 1–32. doi: 10.3390/diseases7010012
- Fernández Tuesta, E., Bola nos Pizarro, M., Pimentel Neves, M. D., Fernández, G., and Axel-Berg, J. (2020). Complex networks for

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdata.2022.1027783/full#supplementary-material>

Supplementary data

The Supplementary Material file reports a list of the indicators used for the analysis in the main text, with indication of their sources and their availability ranges, and the values of the additional performance indicators, that corroborate the evaluation of the PAD prediction made by the RF algorithm.

- benchmarking in global universities rankings. *Scientometrics*. 125, 405–425. doi: 10.1007/s11192-020-03637-9
- Flach, P. (2019). Performance evaluation in machine learning: the good, the bad, the ugly, and the way forward. *Proc. AAAI Conf. Artif. Intell.* 33, 9808–9814. doi: 10.1609/aaai.v33i01.33019808
- Garcia, D., Mitike Kassa, Y., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., et al. (2018). Analyzing gender inequality through large-scale Facebook advertising data. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6958–6963. doi: 10.1073/pnas.1717781115
- Gardener, S. L., and Rainey-Smith, S. R. (2018). The role of nutrition in cognitive function and brain ageing in the elderly. *Curr. Nutr. Rep.* 7, 139–149. doi: 10.1007/s13668-018-0229-y
- Global Change Data Lab (2022). *Our World in Data*. Available online at: <https://ourworldindata.org>
- GlobalData (2021). *Brain Cancer-Epidemiology Forecast to 2030*. Available online at: <https://www.globaldata.com/store/report/brain-cancer-epidemiology-analysis/>
- Hausmann, R., Hidalgo, C., Bustos, S., Coscia, M., Simoes, A., and Yildirim, M. (2014). *The Atlas of Economic Complexity*. Cambridge, MA: MIT Press.
- Hidalgo, C., Klinger, B., Barabasi, A.-L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science* 317, 482–487. doi: 10.1126/science.1144581
- Hogan, D., Fiest, K., Roberts, J., Maxwell, C., Dykeman, J., Pringsheim, T., et al. (2016a). The prevalence and incidence of dementia with lewy bodies: a systematic review. *Can. J. Neurol. Sci.* 43, S83–S95. doi: 10.1017/cjn.2016.2
- Hogan, D., Jetté, N., Fiest, K., Roberts, J., Pearson, D., Smith, E., et al. (2016b). The prevalence and incidence of frontotemporal dementia: a systematic review. *Can. J. Neurol. Sci.* 43, S96–S109. doi: 10.1017/cjn.2016.25
- IARC (2020). *Globocan 2020*. Available online at: <https://gco.iarc.fr/today/data/factsheets/cancers/31-Brain-central-nervous-system-fact-sheet.pdf>
- Jiménez-Luna, J., Grisoni, F., and Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* 2, 573–584. doi: 10.1038/s42256-020-00236-4
- Kursa, M., and Rudnicki, W. (2010). Feature selection with the Boruta package. *J. Stat. Software* 36, 1–13. doi: 10.18637/jss.v036.i11
- Lamanna, F., Lenormand, M., Salas-Olmedo, M. H., Romanillos, G., Gonçalves, B., and Ramasco, J. J. (2018). Immigrant community integration in world cities. *PLoS ONE* 13, e0191612. doi: 10.1371/journal.pone.0191612
- Lombardi, A., Diacono, D., Amoroso, N., Monaco, A., Tavares, J., Bellotti, R., et al. (2021). Explainable deep learning for personalized age prediction with brain morphology. *Front. Neurosci.* 15, 674055. doi: 10.3389/fnins.2021.674055
- Lundberg, S., Erion, G., Chen, H., DeGrave, A., Prutkin, J., Nair, B., et al. (2020). From local explanations to global understanding with explainable ai for trees. *Nat. Mach. Intell.* 2, 56–67. doi: 10.1038/s42256-019-0138-9
- Lundberg, S., and Lee, S. (2017). “A unified approach to interpreting model predictions,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (New York, NY), 44768–44777.
- Ma, Y., Ajnakina, O., Steptoe, A., and Cadar, D. (2020). Higher risk of dementia in english older individuals who are overweight or obese. *Int. J. Epidemiol.* 49, 1353–1365. doi: 10.1093/ije/dyaa099
- Martini, G., Bracci, A., Jaiswal, S., Corea, M., Riches, L., Rivers, J., et al. (2021). Nowcasting food insecurity on a global scale. *medRxiv* 2021.06.23.21259419. doi: 10.1101/2021.06.23.21259419
- Miller, T. (2019). Explanation in artificial intelligence: insights from the social sciences. *Artif. Intell.* 267, 1–38. doi: 10.1016/j.artint.2018.07.007
- Monaco, A., Amoroso, N., Bellantuono, L., Lella, E., Lombardi, A., Monda, A., et al. (2019). Shannon entropy approach reveals relevant genes in Alzheimer's disease. *PLoS ONE* 14, e0226190. doi: 10.1371/journal.pone.0226190
- Monaco, A., Pantaleo, E., Amoroso, N., Bellantuono, L., Lombardi, A., Tateo, A., et al. (2020). Identifying potential gene biomarkers for Parkinson's disease through an information entropy based approach. *Phys. Biol.* 18, 016003. doi: 10.1088/1478-3975/abc09a
- Moore, K., Hughes, C., Ward, M., Hoey, L., and McNulty, H. (2018). Diet, nutrition and the ageing brain: current evidence and new directions. *Proc. Nutr. Soc.* 77, 152–163. doi: 10.1017/S0029665117004177
- Mosconi, L., and McHugh, P. F. (2015). Let food be thy medicine: diet, nutrition, and biomarkers of Alzheimer's disease. *Curr. Nutr. Rep.* 4, 126–135. doi: 10.1007/s13668-014-0111-5
- Musil, F., Willatt, M. J., Langovoy, M. A., and Ceriotti, M. (2019). Fast and accurate uncertainty estimation in chemical machine learning. *J. Chem. Theory Comput.* 15, 906–915. doi: 10.1021/acs.jctc.8b00959
- Newman, M. (2018). *Networks, 2nd Edn*. New York, NY: Oxford University Press Inc.
- Ngabirano, L., Samieri, C., Feart, C., Gabelle, A., Artero, S., Duflos, C., et al. (2019). Intake of meat, fish, fruits, and vegetables and long-term risk of dementia and Alzheimer's disease. *J. Alzheimers Dis.* 68, 711–722. doi: 10.3233/JAD-180919
- Peters, R., Ee, N., Peters, J., Booth, A., Mudway, I., K.J., et al. (2019). Air pollution and dementia: a systematic review. *J. Alzheimers Dis.* 70:S145–S163. doi: 10.3233/JAD-180631
- Pistolato, F., Ohayon, E. L., Lam, A., Langley, G. R., Novak, T. J., Pamiés, D., et al. (2016). Alzheimer disease research in the 21st century: past and current failures, new perspectives and funding priorities. *Oncotarget* 7, 38999–39016. doi: 10.18632/oncotarget.9175
- Porsteinson, A. P., Isaacson, R. S., Knox, S., Sabbagh, M. N., and Rubino, I. (2021). Diagnosis of early Alzheimer's disease: clinical practice in 2021. *J. Prev. Alzheimers Dis.* 8, 371–386. doi: 10.14283/jpad.2021.23
- Prince, M., Guerchet, M., and Prina, M. (2015a). *The Epidemiology and Impact of Dementia-Current State and Future Trends*. WHO Thematic Briefing. Geneva: World Health Organization.
- Prince, M., Wimo, A., Guerchet, M., Ali, G., Wu, Y., and Prina, M. (2015b). *World Alzheimer Report 2015. The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends*. London: Alzheimer's Disease International.
- Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., and Gabrielli, A. (2019). Unfolding the innovation system for the development of countries: co-evolution of Science, technology and production. *Sci. Rep.* 9, 16440. doi: 10.1038/s41598-019-52767-5
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.r-project.org/>
- Rao, R. B., Fung, G., and Rosales, R. (2008). “On the dangers of cross-validation. an experimental evaluation,” in *Proceedings of the 2008 SIAM International Conference on Data Mining* (Philadelphia, PA: Society for Industrial and Applied Mathematics), 588–596.
- Reitz, C., Brayne, C., and Mayeux, R. (2011). Epidemiology of Alzheimer disease. *Nat. Rev. Neurol.* 7, 137–152. doi: 10.1038/nrneurol.2011.2
- Schaffer, C. (1993). Selecting a classification method by cross-validation. *Mach. Learn.* 13, 135–143. doi: 10.1007/BF00993106
- Silva, M., Loures, C., Alves, L., de Souza, L., Borges, K., and Carvalho, M. (2019). Alzheimer's disease: risk factors and potentially protective measures. *J. Biomed. Sci.* 26, 33. doi: 10.1186/s12929-019-0524-y
- Singh-Manoux, A., Dugravot, A., Shipley, M., Brunner, E. J., Elbaz, A., Sabia, S., et al. (2018). Obesity trajectories and risk of dementia: 28 years of follow-up in the whitehall ii study. *Alzheimers Dement* 14, 178–186. doi: 10.1016/j.jalz.2017.06.2637
- Sorns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125. doi: 10.1111/j.1749-6632.2010.05888.x
- Steele, J. E., and et al. (2017). Mapping poverty using mobile phone and satellite data. *J. R. Soc. Interface* 14, 20160690. doi: 10.1098/rsif.2016.0690
- Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries fitness and products complexity. *Sci. Rep.* 2, 723. doi: 10.1038/srep00723
- United Nations Department of Economic and Social Affairs (2021). *Transforming our world: The 2030 Agenda for Sustainable Development*. Available online at: <https://sdgs.un.org/2030agenda>
- van den Brink, A. C., Brouwer-Brolsma, E. M., Berendsen, A. A. M., and van de Rest, O. (2019). The mediterranean, dietary approaches to stop hypertension (DASH), and Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) diets are associated with less cognitive decline and a lower risk of Alzheimer's disease—a review. *Adv. Nutr.* 10, 1040–1065. doi: 10.1093/advances/nmz054
- Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., et al. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* 368, l6927. doi: 10.1136/bmj.l6927
- WHO (2015). *The Epidemiology and Impact of Dementia: Current State and Future Trends*. Geneva: World Health Organization.
- WHO (2017a). *Global Action Plan on the Public Health Response to Dementia 2017–2025*. Geneva: World Health Organization.
- WHO (2017b). *One Health*. Available online at: <https://www.who.int/news-room/questions-and-answers/item/one-health>
- WHO (2022a). *Dementia*. Available online at: <https://www.who.int/news-room/fact-sheets/detail/dementia>

WHO (2022b). *Global Health Observatory Data Repository*. Available online at: <https://apps.who.int/gho/data/node/home>

Wolters, F., and Ikram, M. (2019). Epidemiology of vascular dementia. *Arterioscler. Thromb. Vasc. Biol.* 39, 1542–1549. doi: 10.1161/ATVBAHA.119.311908

Young, J. A., Lind, C., Orange, J., and Savundranayagam, M. Y. (2019). Expanding current understandings of epistemic injustice and dementia: learning from stigma theory. *J. Aging Stud.* 48, 76–84. doi: 10.1016/j.jaging.2019.01.003

Zhang, H., Greenwood, D. C., Risch, H. A., Bunce, D., Hardie, L. J., and Cade, J. E. (2021). Meat consumption and risk of incident dementia: cohort study of 493,888 uk biobank participants. *Am. J. Clin. Nutr.* 114, 175–184. doi: 10.1093/ajcn/nqab028

Zhong, G., Wang, Y., Zhang, Y., Guo, J. J., and Zhao, Y. (2015). Smoking is associated with an increased risk of dementia: a meta-analysis of prospective cohort studies with investigation of potential effect modifiers. *PLoS ONE* 10, e0118333. doi: 10.1371/journal.pone.0118333